OBESITY EXPERTISE

A first-principles curriculum in 196 books

BOOK 12 13th of 196 in the series

Probability and statistical inference · Rung 1

Probability, random variables, and the difference between a parameter and an estimate

WHY A POSITIVE TEST USUALLY MEANS LESS THAN IT SEEMS, AND WHAT A REPORTED NUMBER IS A NUMBER OF

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

drharshmaheshwari.com · version 1.0

OBESITY EXPERTISE · BOOK 12

Probability and statistical inference · Rung 1

Probability, random variables, and the difference between a parameter and an estimate

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

contact@drharshmaheshwari.com · drharshmaheshwari.com · ORCID 0000-0001-8848-1918

SERIES	Obesity Expertise: A first-principles curriculum in 196 books
BOOK	Book 12, Probability and statistical inference · Rung 1 (13th of 196 in the series)
VERSION	1.0
LAST UPDATED	13 October 2026
AUTHOR	Dr. Harsh Maheshwari, MD Community Medicine Resident, AIIMS Raipur
CONTACT	contact@drharshmaheshwari.com
ORCID	https://orcid.org/0000-0001-8848-1918
NEWEST VERSION	drharshmaheshwari.com

© 2026 Harsh Maheshwari.

This book is shared free, for anyone to read, under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International licence (CC BY-NC-SA 4.0). You may copy, share and adapt it for non-commercial purposes, if you credit the author and share what you make under the same licence.

It quotes and adapts material from OpenStax textbooks (Rice University), which are licensed CC BY-NC-SA 4.0; that is why this book carries the same licence. It also quotes short passages from laws, reports and other works, each credited where it is used; those passages remain under their owners' terms. Fonts: Source Serif 4, Inter and JetBrains Mono, SIL Open Font License 1.1.

HOW THIS BOOK WAS MADE

Written for the author's own study, then shared. Drafted and checked with AI assistance (Claude, by Anthropic) under the author's direction. Every factual claim was checked against a source the checker had open, every quotation against its source text, and every worked answer recomputed. Mistakes can remain; the feedback page says how to report one.

NOT ADVICE

This book is for learning. It is not medical, legal or financial advice, and it does not replace a qualified professional who knows your situation.

About the author



Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

Harsh Maheshwari completed his MBBS at King George's Medical University (KGMU), Lucknow, and is a resident in MD Community Medicine at the All India Institute of Medical Sciences (AIIMS), Raipur. His MD thesis, on chrononutrition and obesity, was selected by the Indian Council of Medical Research (ICMR) for funding in 2026.

He learns best from first principles: rebuilding an idea from what came before it until the formula feels like the only reasonable answer. This series grew out of that habit.

His long-term aims are to reduce obesity in India and to help make India healthy.

Email	contact@drharshmaheshwari.com
Website	drharshmaheshwari.com
ORCID	0000-0001-8848-1918

Introduction

Why this book exists

A positive test result usually feels like an answer. Often it is not, and most doctors asked to work out what it means give a number far too high. This book teaches the rules of probability that settle the question, ending at Bayes' theorem, and the handful of patterns of chance that keep turning up in health data. Then it separates three things a reported number can be: a fact about everyone, an estimate from a sample, or a prediction. Every later book on statistics leans on these ideas, and so does every conversation with a patient about a test.

How to read it

Read Book 0 and Book 2 first, or at least the sections this book names as it goes. Then read this one in order, with a pen and a calculator, and do the practice problems in each section before you read on. The book asks you to build one thing: a one-page handout explaining a positive result for a test used in your own clinic. Choose the test and find your clinic's register before you start, because one section asks you to count how common the condition is among your own patients.

Every section has the same shape. **Definition** states the idea exactly. **In plain terms** says it again in everyday words. **Illustration** works through one real case, usually with a source you can open yourself. **Where this picture breaks** says where the illustration stops being true. **Must-know points** are what should change what you do. Then come **exercises** and, in most sections, a graded **practice set**.

Do the exercises before you look anything up. The worked answers are in the appendix at the back, under the same numbers. The references for each Part follow that Part. The glossary lists every term of art in the plain words it was first given. The last pages list all 196 books in the series and show where this one sits.

If you are new to this

Book 0, in particular percentages (A4), ratios and proportions (A5), powers and roots (A6), and all of Part D: what a probability is and odds (D1), independence (D2), the two-way table with sensitivity, specificity and the positive predictive value (D3), mean and standard deviation (D5) and parameter against statistic (D6). Book 2, sections 1, 4 and 13 to 15: the symbols and the Σ sign (1), e and natural logarithms (4), and the random variable, expectation and variance (13 to 15). The mathematics is arithmetic with a calculator, fractions, powers, the sum sign, and one-line algebra; no calculus is needed beyond reading an area under a curve as a probability, which Book 2 taught.

On a first reading, read each section's definition, plain-terms explanation, illustration and must-know points, and do the exercises. Worked derivations can be skimmed and revisited once the idea feels familiar. Keep a calculator beside you: every worked answer can be followed with pen, paper and a calculator. Anything beyond the stated prerequisites is explained where it first

appears, or marked as beyond this book. The **Symbols used in this book** page lists every sign and Greek letter with how to read it and where it is first used.

How to send feedback

If something is wrong, unclear, or could be better, please write to **contact@drharshmaheshwari.com**. Put the book number, the version and the section in the subject line (for example: *Book 12, v1.0, section C5*). Say what the text says, what you think it should say, and, if you can, the source.

Corrections are made in new versions. The version number and the date of the last update are on the copyright page of every book, and the newest version of every book is at **drharshmaheshwari.com**.

> **WHERE THIS BOOK SITS**
> Book 12 of the series' 196 books (Book 0 and 195 rung books). Before it: Book 11, Adjacent specialist domains: recognise and route · Rung 1. After it: Book 13, Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 1. The full list is at the back.

Symbols used in this book

Every sign and Greek letter this book prints, how to read it aloud, what it means in this series, and the section where it first appears. Each is also explained where it is first used.

Symbol	Read as	Meaning	First in
SIGNS			
–	minus	Subtraction, or a negative number	1 · Sample space, events and the three rules every probability obeys
×	times	Multiplication	1 · Sample space, events and the three rules every probability obeys
≈	approximately equal to	The two sides are close but not exactly equal	1 · Sample space, events and the three rules every probability obeys
§	section	A numbered section of a document, as in a statute's §5 or a book's §7.3	1 · Sample space, events and the three rules every probability obeys
|	given	In a probability, P(A | B) is the probability of A given that B has happened	2 · Conditional probability as a definition, and the multiplication rule
÷	divided by	Division	2 · Conditional probability as a definition, and the multiplication rule
≥	greater than or equal to	An inequality that includes its boundary	10 · Where the prior comes from: pre-test probability in your own clinic
√	square root of	The number that, multiplied by itself, gives the value: √9 = 3	12 · The Bernoulli and binomial distributions: counting successes in independent trials
≤	less than or equal to	An inequality that includes its boundary	13 · The geometric and hypergeometric distributions: waiting for the first success, and drawing without replacement
±	plus or minus	A range either side of a value: 11.8 ± 0.6 means 11.2 to 12.4	15 · The normal distribution: what generates it, and areas under it
Σ	sigma, the sum of	Add up everything that follows: $\Sigma x = x_1 + x_2 + \ldots + x_n$	17 · Covariance and correlation of two random variables
MARKS ON LETTERS			

Symbol	Read as	Meaning	First in
x^2, 10^{-3}	superscript	Small raised characters are powers: $x^2 = x \times x$, and $10^{-3} = 0.001$	12 · The Bernoulli and binomial distributions: counting successes in independent trials
$\bar{x}$	bar	A bar over a letter marks a mean: $\bar{x}$ is the mean of x	16 · Skewed quantities and the log scale: when effects multiply
$\hat{p}$	hat	A hat over a letter marks an estimate or a prediction: $\hat{p}$	19 · Parameter, statistic, estimator, estimate
GREEK LETTERS			
σ	sigma	A population standard deviation (σ^2 its variance)	12 · The Bernoulli and binomial distributions: counting successes in independent trials
λ	lambda	A rate (events per unit time), or a wavelength	14 · The Poisson distribution: counts of events at a steady rate
μ	mu	In statistics a population mean; in units, micro- (one millionth), as in µg	15 · The normal distribution: what generates it, and areas under it
π	pi	The constant 3.14159...	15 · The normal distribution: what generates it, and areas under it
ρ	rho	A correlation in a population, or a density	17 · Covariance and correlation of two random variables

Contents

Before you start 10

Rung 1 — Introductory 11

1 · Sample space, events and the three rules every probability obeys 11
2 · Conditional probability as a definition, and the multiplication rule 18
3 · Independence given a condition, and where it fails 25
4 · The law of total probability: a probability as a weighted average over groups 32
5 · Bayes' theorem, from the multiplication rule, on a diagnostic test 40
6 · The base-rate case that defeats most clinicians 47
7 · Natural frequencies: Bayes by counting people 54
8 · The odds form of Bayes' theorem and the likelihood ratio 61
9 · The prosecutor's fallacy: P(evidence | innocent) is not P(innocent | evidence) 68
10 · Where the prior comes from: pre-test probability in your own clinic 77
11 · Counting arrangements: factorials and "n choose k" 85
12 · The Bernoulli and binomial distributions: counting successes in independent trials 90
13 · The geometric and hypergeometric distributions: waiting for the first success, and drawing without replacement 97
14 · The Poisson distribution: counts of events at a steady rate 105
15 · The normal distribution: what generates it, and areas under it 114
16 · Skewed quantities and the log scale: when effects multiply 124
17 · Covariance and correlation of two random variables 133
18 · Sums and averages of many independent draws, simulated by hand 141
19 · Parameter, statistic, estimator, estimate 150
20 · How good is an estimator: bias and spread, found by listing every sample 156
21 · The likelihood: the same formula read as a function of the parameter 164
22 · Parameter, estimate or prediction: what a reported number is a number of 170
23 · Worked journey: One clinic test, from the question to a handout a patient can read 177
References *186*

Appendix · Worked answers 199

Glossary 265

The series: all 196 books 269

Before you start

This book uses the following Book 0 sections. Each is summarised where it is first needed; work through Book 0 itself if a summary is not enough.

- Book 0, A2 · Fractions
- Book 0, A4 · Percentages, and the percentage of a percentage
- Book 0, A5 · Ratios, rates and proportions
- Book 0, A6 · Powers, roots and scientific notation
- Book 0, A7 · Logarithms — what they are for before how they work
- Book 0, B5 · Concentration units: mg/dL, mmol/L, and converting between them
- Book 0, C1 · Variables and algebra: what a letter stands for
- Book 0, C2 · Rearranging an equation
- Book 0, D1 · What a probability is
- Book 0, D2 · Counting outcomes; independence
- Book 0, D3 · Conditional probability through the two-way table
- Book 0, D4 · Variation: what differs and by how much
- Book 0, D5 · Average and spread, and why the average is not the person
- Book 0, D6 · Sampling: how a part can tell you about a whole
- Book 0, D7 · Random error and systematic error as different things
- Book 0, F1 · Reading a table
- Book 0, F2 · Reading a figure, and the ways a figure misleads
- Book 0, F4 · Argument: premise, inference, conclusion

Rung 1 — Introductory

1 · Sample space, events and the three rules every probability obeys

DEFINITION. A chance process is anything whose result is not known until it runs: a roll of a die, a man picked at random from a list. An outcome is one result it can give. The sample space is the complete list of outcomes, the term of Book 0, D1. An event is any set of outcomes taken from the sample space.

From two events A and B, three more are built.

- "A and B" is the set of outcomes that are in both A and B.
- "A or B" is the set of outcomes in at least one of them. Outcomes in both are included.
- "Not A", the complement of A, is every outcome in the sample space that is outside A.

A Venn diagram draws these. A box stands for the sample space and a circle for each event. The overlap of two circles is "A and B". Everything inside either circle is "A or B".

P(A), read "the probability of A", is a number given to the event A. Every probability obeys three rules.

1. No probability is negative: P(A) is 0 or more, for every event A.
2. The whole sample space has probability 1.
3. If A and B are mutually exclusive, so that no outcome is in both, then P(A or B) = P(A) + P(B).

First, the complement. A and not A are mutually exclusive, and together they make up the whole sample space.

```
P(A) + P(not A) = P(A or not A)    by rule 3
P(A or not A) = 1    by rule 2
so P(not A) = 1 - P(A)
```

Second, the general addition rule, for any two events, mutually exclusive or not. Cut "A or B" into three pieces that share no outcome: "A and not B", "A and B", and "B and not A". Rule 3, used twice, adds them. For any number of pieces that share no outcome, add them.

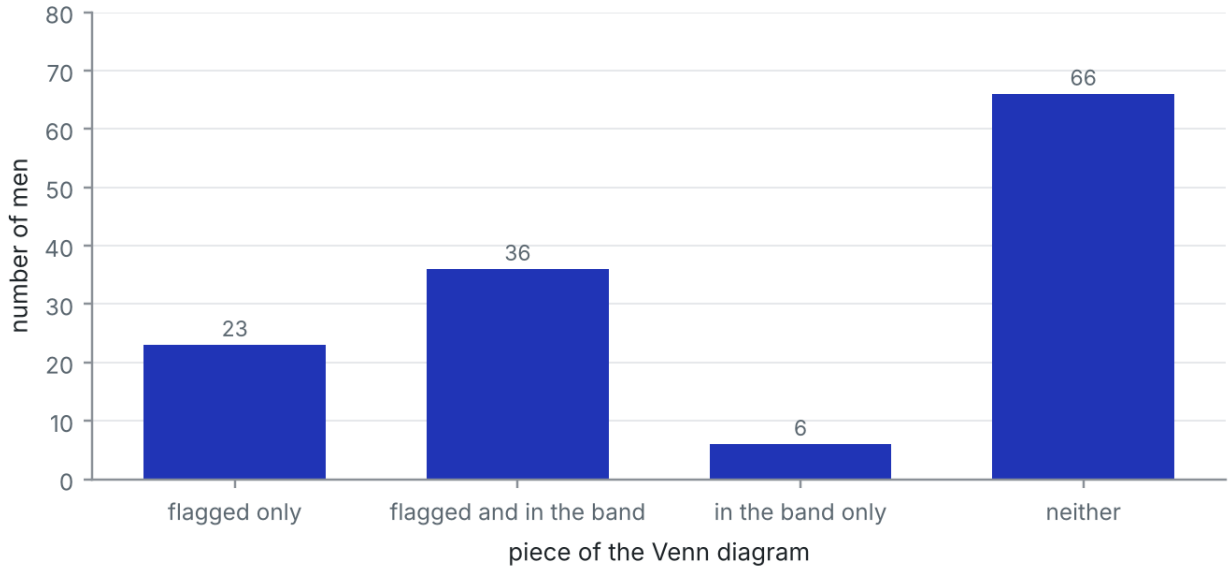
```
P(A or B) = P(A and not B) + P(A and B) + P(B and not A)
P(A) = P(A and not B) + P(A and B)
P(B) = P(B and not A) + P(A and B)
P(A) + P(B) = P(A and not B) + P(B and not A) + 2 × P(A and B)
P(A) + P(B) = P(A or B) + P(A and B)
P(A or B) = P(A) + P(B) - P(A and B)
```

When A and B are mutually exclusive, P(A and B) is 0, and rule 3 comes back.

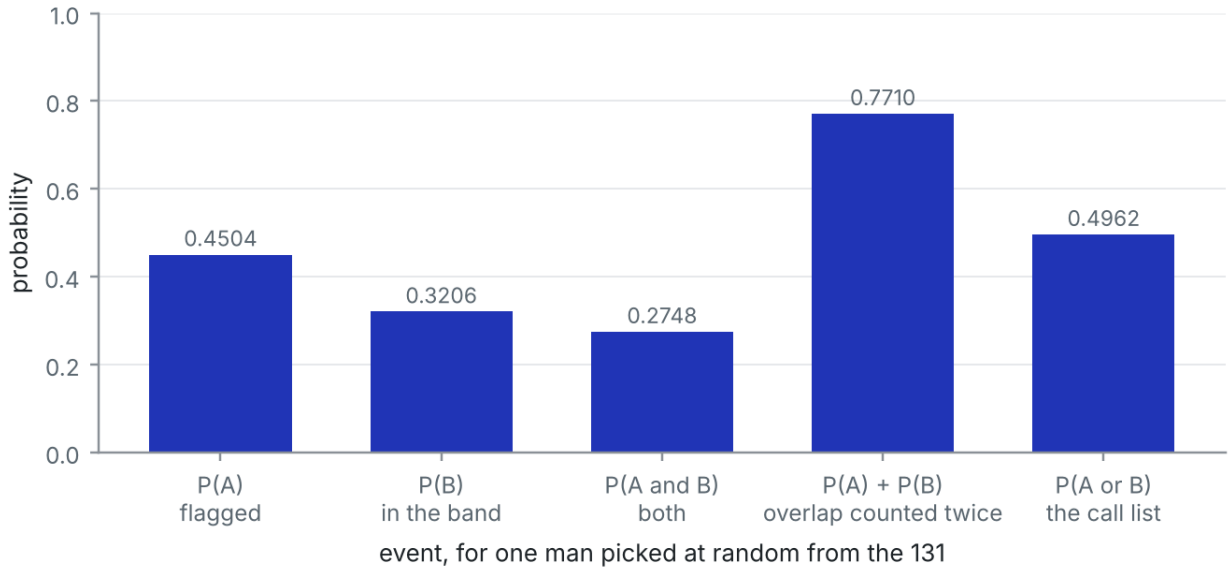
The overlap P(A and B) is at least 0. It is at most the smaller of P(A) and P(B). Put each limit into the rule. With no overlap, P(A or B) is P(A) + P(B), its largest. When one event sits inside the other, P(A or B) is the larger of the two, its smallest. It can never pass 1. [1, 2, 3, 4, 5, 6, 7, 8, 9]

IN PLAIN TERMS. Write down everything that could happen. That list is the sample space. Any group of items on it is an event.

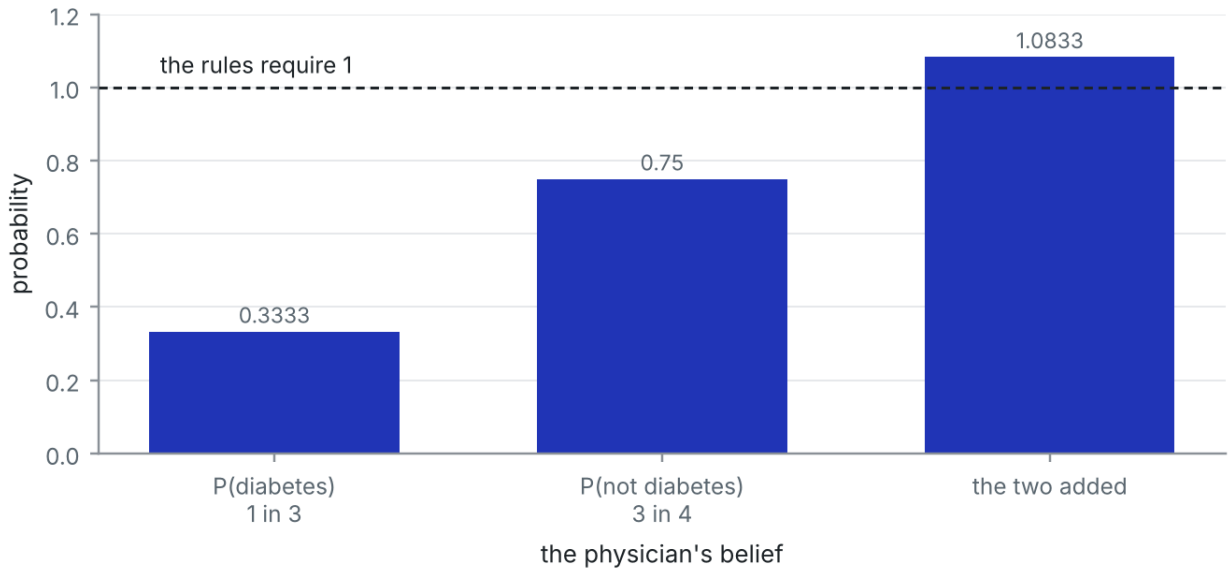
Two events that can both happen need one more step. Adding them counts the overlap twice, so you take the overlap away once.



The 131 men of Kiran's rebuilt table (Illustration 1), cut into the four pieces a Venn diagram of "flagged by the arm tape" and "in the 25-29.9 band" makes. No circles are drawn here: each bar is one piece, and its height is the number of men in it: 23 flagged only, 36 both, 6 in the band only, 66 neither. "Flagged or in the band" is the first three pieces, 65 men. Adding the flagged 59 and the band's 42 gives 101, because the 36 men in both are counted twice.



Pick one of the 131 men at random; A is “flagged” and B is “in the band”. Adding P(A), 0.4504, and P(B), 0.3206, gives 0.7710, which counts the 36 men in both twice. Taking P(A and B), 0.2748, away once gives P(A or B), 0.4962: the 65 men of the call list divided by 131.



The physician’s two made-up beliefs about one man: one in three that he has diabetes, three in four that he has not. The two events are mutually exclusive and cover every possibility, so the rules require them to add to 1. They add to 1.0833, so both cannot be right.

Illustration 1. Book 0 rebuilt a table from Kiran and colleagues’ study of medical students in Mangaluru (Book 0, D3). It covers 131 men. The tape flags a man when his upper arm measures 31.3 centimetres or more. “In the band” means a body-mass index from 25 to 29.9. The table is rebuilt from the paper’s printed percentages, not printed by the paper in this form.

	in the band	not in the band	total
arm 31.3 cm or more	36	23	59
arm under 31.3 cm	6	66	72

	in the band	not in the band	total
total	42	89	131

A camp organiser wants to call back every man the tape flags, and every man in the band. How many men is that? The move you reach for first is to add the row and the column.

```
59 plus 42 = 101
```

That is 101 men, and it is wrong. Look at the 36 men who are flagged and in the band. They sit in the flagged row and in the band column. Adding the row and the column counted each of them twice.

Picture a Venn diagram. The box is the 131 men. One circle is “flagged”, the other is “in the band”. The circles cut the box into four pieces that share no man. The first figure draws the four pieces as bars.

```
59 minus 36 = 23
42 minus 36 = 6
23 plus 36 plus 6 = 65
131 minus 65 = 66
```

So 23 men are flagged only, 36 are both, 6 are in the band only, and 66 are neither. The call list is the first three pieces: 65 men.

Now do it with probabilities. Pick one of the 131 men at random. Each man has the same chance, so the probability of an event is its count divided by 131. Call “flagged” A and “in the band” B. Use the general addition rule. These divisions do not come out exactly, so each is rounded to four decimal places, and the sign ≈, read “is about”, marks a rounded result.

```
59 divided by 131 ≈ 0.4504
42 divided by 131 ≈ 0.3206
36 divided by 131 ≈ 0.2748
0.4504 plus 0.3206 minus 0.2748 = 0.4962
65 divided by 131 ≈ 0.4962
```

P(A or B) is 0.4962, and counting the 65 men directly gives the same. The 101 men of the first try would have given 0.7710, which is 101 divided by 131.

The complement rule finds the men left off the list.

```
1 minus 0.4962 = 0.5038
66 divided by 131 ≈ 0.5038
```

About half the men, 66 of 131, are neither flagged nor in the band.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The counts are the rebuilt table of Book 0, D3, worked back from the paper’s rounded percentages. They hold for these 131 men, students at one college. “In the

band" leaves out the 10 men with an index of 30 or more, so "not in the band" is not the same as "a normal index".

Turning counts into probabilities by dividing by 131 needs every man to have the same chance of being picked. The rules themselves do not need that. They hold however the probabilities were set.

Illustration 2. The same rules bind a probability that is a belief rather than a count. Suppose a physician looks at a man in clinic, before any test. She says: "I'd put his chance of diabetes at one in three. And I'd put his chance of not having it at three in four." Check her two numbers against the rules. Being diabetic and not being diabetic are mutually exclusive, and together they cover every possibility. So the two must add to 1.

```
1 divided by 3 ≈ 0.3333
3 divided by 4 = 0.75
0.3333 plus 0.75 = 1.0833
```

They add to more than 1. The rules cannot say which number is wrong. They say both cannot be right. If she stands by one in three, the other must be two in three.

```
1 minus 0.3333 = 0.6667
```

Grinstead and Snell describe this kind of probability in *Introduction to Probability* (https://math.dartmouth.edu/~prob/prob/prob.pdf), §1.2: "you can determine your own personal probability by seeing what kind of a bet you would be willing to make." Price her beliefs as bets. At one in three, a ticket that pays 1 rupee if he has diabetes is worth a third of a rupee to her. At three in four, a ticket that pays 1 rupee if he has not is worth three-quarters of a rupee. Buying both costs her 1.0833 rupees. Whatever is true, exactly one ticket pays, and it pays 1 rupee.

```
0.3333 plus 0.75 minus 1 = 0.0833
```

She loses 0.0833 rupees for certain. That is what breaking the rules costs.

WHERE THIS PICTURE BREAKS. The physician and her numbers are made up. The check shows only that two beliefs contradict each other. A belief that passes every rule can still be badly wrong about this man. The rules test how numbers fit together, never whether any one of them matches the world.

MUST KNOW POINTS FOR YOU.

- Before you add, ask whether one person can be in both groups. If yes, subtract the overlap once.
- In this book "A or B" means at least one, both included.
- When two percentages are reported and their overlap is not, P(A or B) is known only as a range. It is at least the larger of the two. It is at most their sum, and never above 1. The same holds for three or more groups: at least the largest, at most the sum. Quote the range, not the sum.

- The shares of categories that cannot overlap must add to 100 per cent.
- The three rules check that probabilities fit together. They cannot check that any one probability is true.
- Before you subtract an overlap, check that both percentages share one denominator.

COMMON MISREADING. "P(A or B) is P(A) plus P(B)." That is rule 3, and it holds only when A and B cannot both happen. For events that can overlap it counts the overlap twice. Kiran's flagged row and band column add to 101 men when only 65 are on the call list: the 36 men in both were counted twice.

EXERCISE 1 (retrieval). From memory: state the three rules every probability obeys. Then write the general addition rule, and say in one sentence why the last term is subtracted.

EXERCISE 2 (interpretation). A made-up survey of 400 clinic attenders asks "Which of these do you have?" and lets each person tick as many as apply. The report prints:

condition	share ticking it (%)
diabetes	30
hypertension	45
overweight	40

A committee member adds the three and says "115 per cent: so everyone has at least one, and some have two." Say what the table does and does not show.

PRACTICE. Eight problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A fair die is rolled once. Its sample space is 1, 2, 3, 4, 5 and 6, each equally likely. Event A is "an even number". Event B is "4 or more".

List the outcomes in A, in B, in "A and B", in "A or B" and in "not A".

2. P(A) is 0.3, P(B) is 0.5 and P(A and B) is 0.1. Work out P(A or B), and the probability that neither A nor B happens.

3. P(A) is 0.45, P(B) is 0.30 and P(A or B) is 0.60. Work out P(A and B). Are A and B mutually exclusive?

4. Kiran and colleagues' Table 2 sorts the 131 men by body-mass index (BMI, weight in kilograms divided by the square of height in metres). It gives 5 men below 18.5, 74 from 18.5 to 24.9, 42 from 25 to 29.9 and 10 at 30 or more.

Pick one of the 131 men at random. Work out the probability that his BMI is 25 or more, and the probability that it is 18.5 or more. Say which rule you used each time.

5. Take the same 131 men. Kiran's Table 2 gives 52 with a BMI of 25 or more. It gives 41 with a waist of 90 cm or more. The paper does not say how many men are in both groups.

Pick one man at random. Between what limits must the probability lie that he has a BMI of 25 or more or a waist of 90 cm or more? Say when each limit would be reached.

6. Here is a worked answer. Find the step that broke.

```
Kiran's rebuilt table: 131 men, 59 flagged by the arm tape, 42 in the 25-29.9 band
P(flagged or in the band) = P(flagged) + P(in the band)
= 59 divided by 131 plus 42 divided by 131
= 0.4504 plus 0.3206 = 0.7710
so about 77 per cent of the men are flagged or in the band
```

7. Here is a worked answer. Find the step that broke.

```
A and B cannot both happen. P(A) = 0.2 and P(B) = 0.3.
P(A and B) = P(A) times P(B) = 0.2 times 0.3 = 0.06
P(A or B) = 0.2 plus 0.3 minus 0.06 = 0.44
```

8. The NFHS-5 India fact sheet (2019-21) gives two figures for women aged 15-49. In all, 24.0% have overweight or obesity (indicator 88). And 57.0% are anaemic (indicator 95). A district presentation says: "So 81 per cent of women have one problem or the other."

Decide what to compute, compute it, and say what the presentation's number does not establish.

2 · Conditional probability as a definition, and the multiplication rule

DEFINITION. Let A and B be events, with P(B) above 0. The conditional probability of A given B, written P(A | B), is the probability of A once you know that B has happened. The vertical bar is read "given". It is defined as the probability of both, divided by the probability of the condition. The sign ÷ means "divided by".

```
P(A | B) = P(A and B) ÷ P(B)
```

Knowing B removes every outcome outside B. What is left is a smaller sample space, B itself, and the division rescales it so that it has probability 1 again. Inside B the three rules of section 1 hold again, so P(not A | B) = 1 − P(A | B).

In a two-way table, the probability in one cell, such as P(A and B), is a joint probability: a probability about two events together. A probability about one event alone, read from a row or column total, is a marginal probability. A conditional probability is a joint probability divided by a marginal one.

Multiply both sides of the definition by P(B). This gives the general multiplication rule, which holds for any two events, independent or not.

```
P(A and B) = P(B) × P(A | B)
```

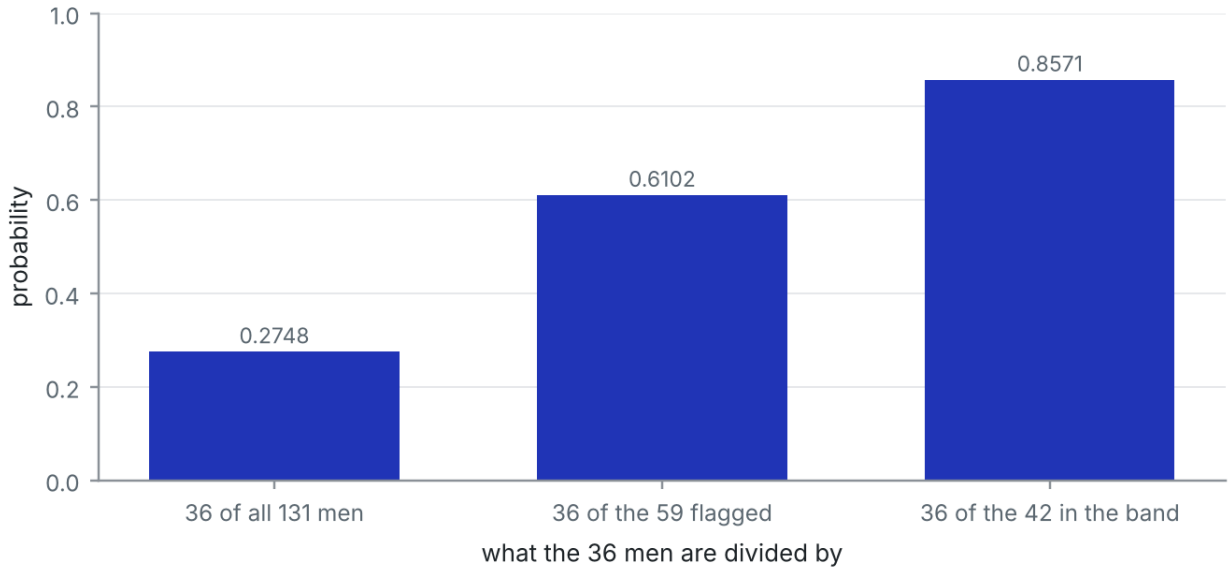
The same joint probability can be built in the other order, starting from A.

```
P(A and B) = P(A) × P(B | A)
```

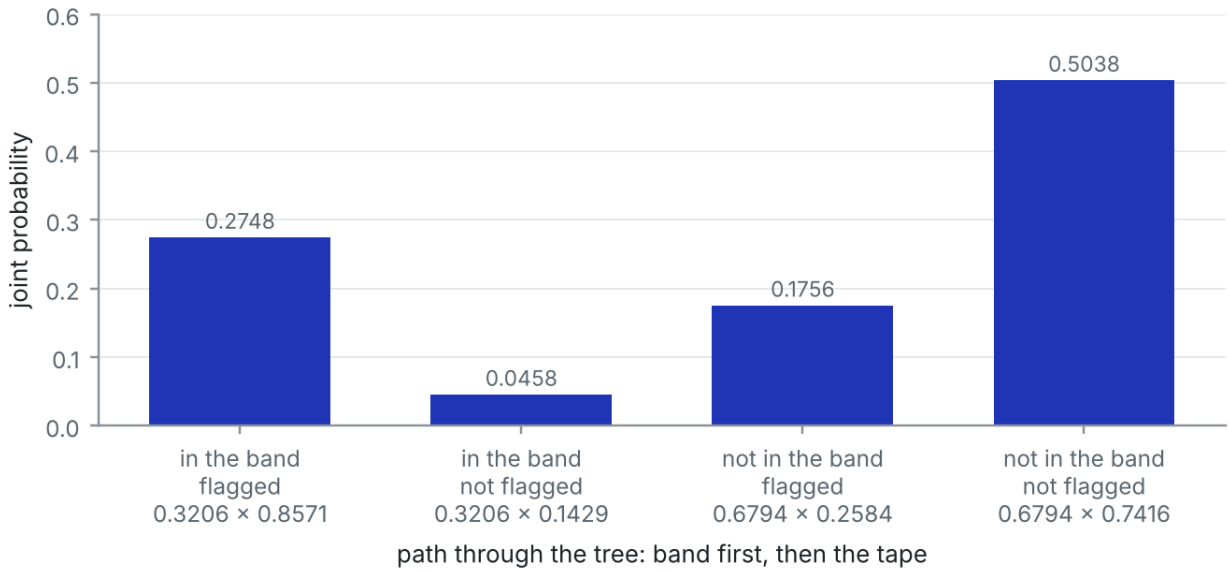
A tree diagram draws the multiplication rule in stages. The first split gives the marginal probabilities of one event. Each second split gives the conditional probabilities of the other event, given the first. Multiplying along a path from left to right gives a joint probability. [11, 12, 13, 14, 15, 16, 17]

IN PLAIN TERMS. "Given B" means: forget everyone who is not B. Among the people left, what share are A? That share is P(A | B).

The order matters.



In Kiran's rebuilt table of 131 men, 36 are both flagged by the arm tape and in the 25-29.9 band. Divided by all 131 men, that is the joint probability 0.2748. Divided by the 59 flagged men, it is P(in the band | flagged), 0.6102. Divided by the 42 men in the band, it is P(flagged | in the band), 0.8571. One top, three bottoms, three different numbers.



The tree for Kiran's 131 men, split first by the band (0.3206 in it, 0.6794 not) and then by the tape. Multiplying along each path gives a joint probability: 0.2748, 0.0458, 0.1756 and 0.5038, the four cells of the probability table. The four ends add to 1.

Illustration 1. Start from the table of Book 0, D3, rebuilt from Kiran and colleagues' study of 131 men at a medical college in Mangaluru. The tape flags a man whose upper arm measures 31.3 centimetres or more. "In the band" means a body-mass index from 25 to 29.9.

	in the band	not in the band	total
arm 31.3 cm or more	36	23	59
arm under 31.3 cm	6	66	72

	in the band	not in the band	total
total	42	89	131

Pick one of the 131 men at random. Divide every count by 131 to turn the table into probabilities.

```
36 divided by 131 ≈ 0.2748
23 divided by 131 ≈ 0.1756
6 divided by 131 ≈ 0.0458
66 divided by 131 ≈ 0.5038
42 divided by 131 ≈ 0.3206
59 divided by 131 ≈ 0.4504
```

	in the band	not in the band	total
arm 31.3 cm or more	0.2748	0.1756	0.4504
arm under 31.3 cm	0.0458	0.5038	0.5496
total	0.3206	0.6794	1

The four inner cells are joint probabilities: P(flagged and in the band) is 0.2748. The row and column totals are marginal probabilities: P(in the band) is 0.3206.

Now ask the clinic's question. A man is flagged. What is the chance he is in the band? Use the definition: the joint probability divided by the marginal probability of the condition.

```
0.2748 divided by 0.4504 = 0.6101
36 divided by 59 ≈ 0.6102
```

About 61 per cent, the positive predictive value of Book 0, D3. The two lines differ in the fourth decimal only because 0.2748 and 0.4504 were rounded. Dividing by 131 and then dividing again cancels the 131, so the formula and the count give the same answer.

Now ask the other question. A man is in the band. What is the chance he is flagged?

```
0.2748 divided by 0.3206 = 0.8571
36 divided by 42 ≈ 0.8571
```

About 86 per cent, the sensitivity. P(in the band | flagged) is 0.61. P(flagged | in the band) is 0.86. Same top, 0.2748. Different bottoms, 0.4504 and 0.3206.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The table is rebuilt from the paper's rounded percentages, not printed by the paper. It holds for these 131 men, students at one college. "Not in the band" includes the 10 men with an index of 30 or more, so "flagged and not in the band" is not the same as "flagged wrongly" for the question "is his index 25 or more?".

The 61 per cent belongs to this group, in which about 32 per cent of the men are in the band. In a group where fewer men are in the band, the same tape gives a lower figure, as Book 0, D3 showed.

Illustration 2. Often you have no table, only a marginal probability and a conditional one. Then the multiplication rule builds the joint probability, and a tree diagram keeps the stages in order.

Split the same 131 men first by the band. On the first split, P(in the band) is 0.3206 and P(not in the band) is 0.6794. On the second split, read each conditional probability from its own column.

```
36 divided by 42 ≈ 0.8571
6 divided by 42 ≈ 0.1429
23 divided by 89 ≈ 0.2584
66 divided by 89 ≈ 0.7416
```

Multiply along each of the four paths, left to right.

```
0.3206 times 0.8571 = 0.2748
0.3206 times 0.1429 = 0.0458
0.6794 times 0.2584 = 0.1756
0.6794 times 0.7416 = 0.5038
0.2748 plus 0.0458 plus 0.1756 plus 0.5038 = 1.0000
```

The four ends are the four cells of the probability table, and they add to 1. The diagram and the table hold the same information.

Here is the move you will use in clinic, from published figures alone. The paper's Table 2 prints the share of men in the band as 32.1 per cent. Its abstract prints the sensitivity as 86 per cent. Multiply.

```
0.321 times 0.86 = 0.2761
0.2761 times 131 = 36.2
```

About 36 men are both in the band and flagged, as the rebuilt table says. The small gap comes from the two rounded percentages.

You could equally split the men first by the tape: P(flagged) is 0.4504, then P(in the band | flagged) is 0.6102.

```
0.4504 times 0.6102 = 0.2748
```

The same joint probability, built in the other order. That is the fact Bayes' theorem will turn into a formula, three sections on, in section 5.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. Multiplying a prevalence by a sensitivity gives the joint probability only for the group both numbers describe. Here both describe the same 131 men. A preval-

ence from one place and a sensitivity from a study elsewhere give a joint probability that is only as good as the assumption that the test behaves the same in both places.

MUST KNOW POINTS FOR YOU.

- "P(A | B) and P(B | A) are the same number, said two ways" is the error. They share a top and divide it by different bottoms. Before you quote a conditional, say aloud what it is conditioned on.
- The denominator of a conditional is the condition. Dividing a cell by the grand total gives a joint probability, a different and usually smaller number.
- You can build the chance of "both" from a marginal and a conditional: P(A and B) = P(B) × P(A | B). Use it when a paper gives a prevalence and a sensitivity but no table. Then rebuild the table as counts and check that the counts are whole people.
- The multiplication rule needs no independence. What independence changes is only that P(A | B) may be replaced by P(A).
- Before you quote a conditional for your clinic, ask whose table it came from.
- When you read "x per cent of those who...", the group after "of those who" is the condition. Rewrite the sentence as P(... | those who ...) before you repeat it.

COMMON MISREADING. "86 per cent of the men in the band were flagged, so a flagged man has an 86 per cent chance of being in the band." The first is P(flagged | in the band). The second asks for P(in the band | flagged). They share the top, 36 men, and divide it by different bottoms: 42 men in the band, 59 men flagged. In Kiran's men the second is about 61 per cent.

REPORTING IT. Kiran and colleagues measured 131 men. An arm circumference of 31.3 cm or more flagged 36 of the 42 men with a body-mass index from 25 to 29.9 (86 per cent). Of the 59 men it flagged, 36 were in that band (61 per cent). The second figure comes from a table rebuilt from the paper's percentages.

EXERCISE 1 (retrieval). From memory: write the definition of P(A | B) and the general multiplication rule. Then say in one sentence why P(A | B) and P(B | A) are usually different numbers.

EXERCISE 2 (interpretation). A district report on a made-up screening camp says: "Of the 400 adults screened, 120 had a raised blood sugar. Of these, 90 had overweight. In all, 160 adults had overweight."

Write each number as a probability for an adult picked at random from the 400: say which are marginal, which joint and which conditional. Then work out the chance that an adult with overweight had a raised blood sugar, and say why it is not 90 divided by 120.

PRACTICE. Thirteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. In a group of 100 people, 20 are both A and B, 30 are A but not B, 10 are B but not A, and 40 are neither. Work out P(A | B) and P(B | A).

2. P(A and B) is 0.12 and P(B) is 0.4. Work out P(A | B).

3. P(B) is 0.25 and P(A | B) is 0.6. Work out P(A and B).

4. P(A and B) is 0.18 and P(A | B) is 0.45. Work out P(B).

5. P(B) is 0.3. P(A | B) is 0.5 and P(A | not B) is 0.2. Draw a tree diagram that splits first by B, and work out the four joint probabilities at its ends. Check that they add to 1.

6. In Kiran's rebuilt table of 131 men, 36 are flagged by the arm tape and in the 25-29.9 band, and 42 are in the band. Work out P(flagged and in the band) and P(in the band) as probabilities for a man picked at random. Then use the definition to work out P(flagged | in the band), and say what it is called.

7. For women, Kiran's Table 2 gives 151 women, 41 of them with a body-mass index from 25 to 29.9. The paper's cut-off for women is an arm circumference of 28.5 cm, with sensitivity 88 per cent and specificity 83 per cent. As for the men, these fit the 41 women in the band against the other women.

Rebuild the four cells of the table as whole numbers of women. Then work out the chance that a woman the tape flags is in the band.

8. Kiran's Table 2 prints the share of men in the 25-29.9 band as 32.1 per cent. The abstract prints the arm tape's specificity as 74 per cent. Use the multiplication rule to work out the probability that a man picked at random is outside the band and wrongly flagged. Then turn it into a number of the 131 men, and compare it with the rebuilt table's 23.

9. Here is a worked answer. Find the step that broke.

```
Kiran's rebuilt table: 36 men flagged and in the band, 42 in the band, 59 flagged
question: a man is flagged; what is the chance he is in the band?
P(in the band | flagged) = 36 divided by 42 = 0.857
so about 86 per cent of flagged men are in the band
```

10. Here is a worked answer. Find the step that broke.

```
131 men: P(in the band) = 0.3206, so P(not in the band) = 0.6794
sensitivity 0.86, specificity 0.74
P(flagged and not in the band) = 0.6794 times 0.86 = 0.5843
so about 77 of the 131 men are flagged though outside the band
```

11. Here is a worked answer. Find the step that broke.

```
a camp screens 500 adults; 100 have a raised sugar, and 60 of those have overweight
P(overweight | raised sugar) = 60 divided by 500 = 0.12
so 12 per cent of adults with a raised sugar have overweight
```

12. A news item reports Kiran and colleagues' study: "The arm-tape test was 86 per cent accurate: 86 per cent of the young men it flagged had overweight."

Decide what to compute, compute it, and say what the corrected number does not establish.

13. A speaker at a meeting says: "Two in three of our patients with diabetes have overweight. So a patient with overweight has a two-in-three chance of having diabetes."

Suppose, in a made-up clinic, 10 per cent of patients have diabetes and 30 per cent are overweight. Decide what to compute, compute it, and say what your answer does not establish.

3 · Independence given a condition, and where it fails

DEFINITION. Two events A and B, each with probability above 0, are independent when knowing that one happened leaves the probability of the other unchanged.

```
P(A | B) = P(A)
```

Put this into the multiplication rule of the last section, P(A and B) = P(B) × P(A | B), and it becomes the check of Book 0, D2: P(A and B) = P(A) × P(B). Going back: divide P(A and B) = P(A) × P(B) by P(B), and the definition of section 2 gives P(A | B) = P(A). So checking one form is enough. Events that are not independent are dependent.

Independent is not mutually exclusive. If A and B are mutually exclusive and each has probability above 0, then knowing B happened makes A impossible: P(A | B) is 0, not P(A). So such events are always dependent.

Independence can hold inside groups and fail across them. A and B are conditionally independent given C when the definition holds among the outcomes in C:

```
P(A | B and C) = P(A | C)
```

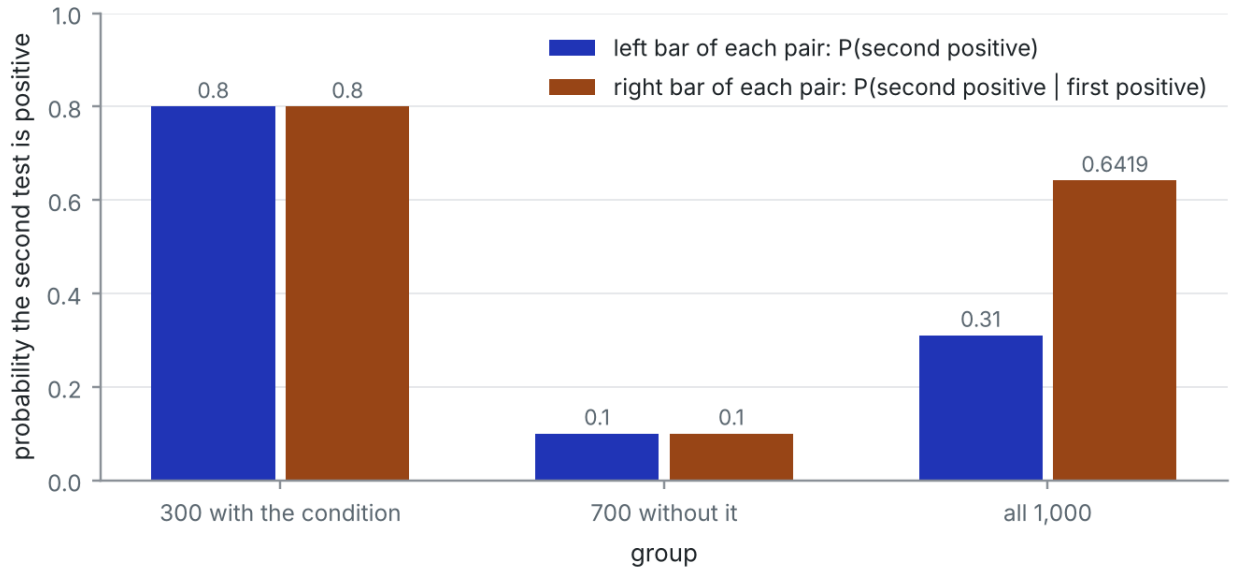
The commonest case is two tests on one person. Among people who have the condition, one test's result may tell you nothing more about the other's. The same may hold among people who do not. Across everyone together, the two results are still dependent, because a positive first result makes it more likely the person has the condition, and so more likely the second test is positive. [18, 19, 20, 21, 22, 23, 24, 25, 26]

IN PLAIN TERMS. Two things are independent if learning one tells you nothing about the other.

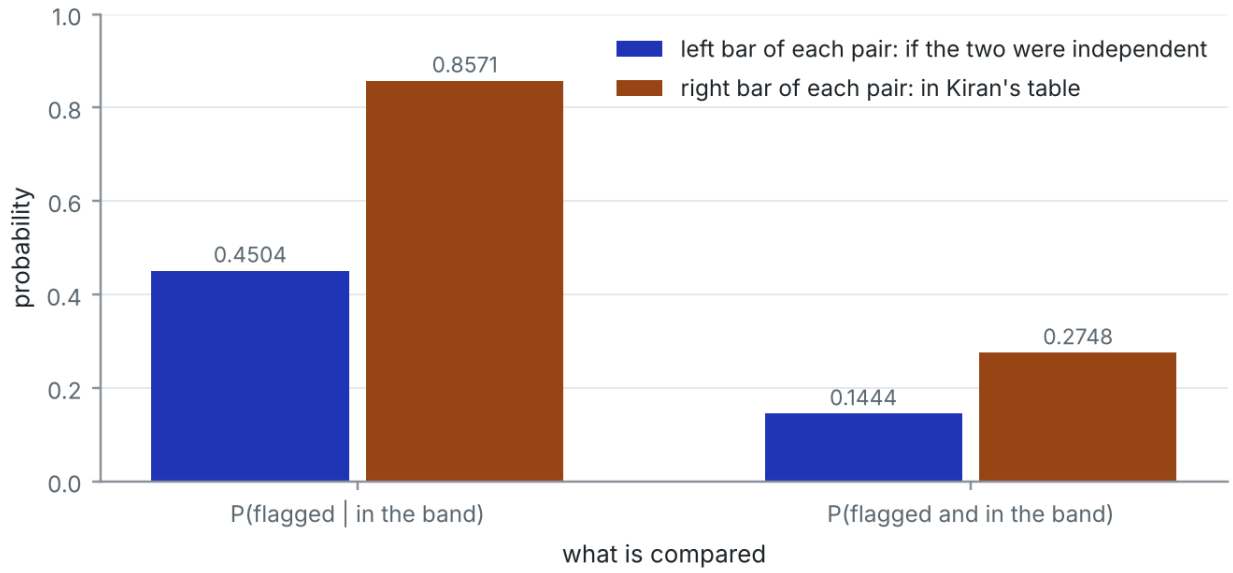
"Cannot both happen" is a different idea, and it forces dependence. If two things cannot both happen, learning one tells you a great deal about the other: it did not happen.

Two tests on the same patient are a trap. Among the sick they may behave independently, and among the well too. Put everyone together and a positive first test makes a positive second test more likely, because it points to a sick person.

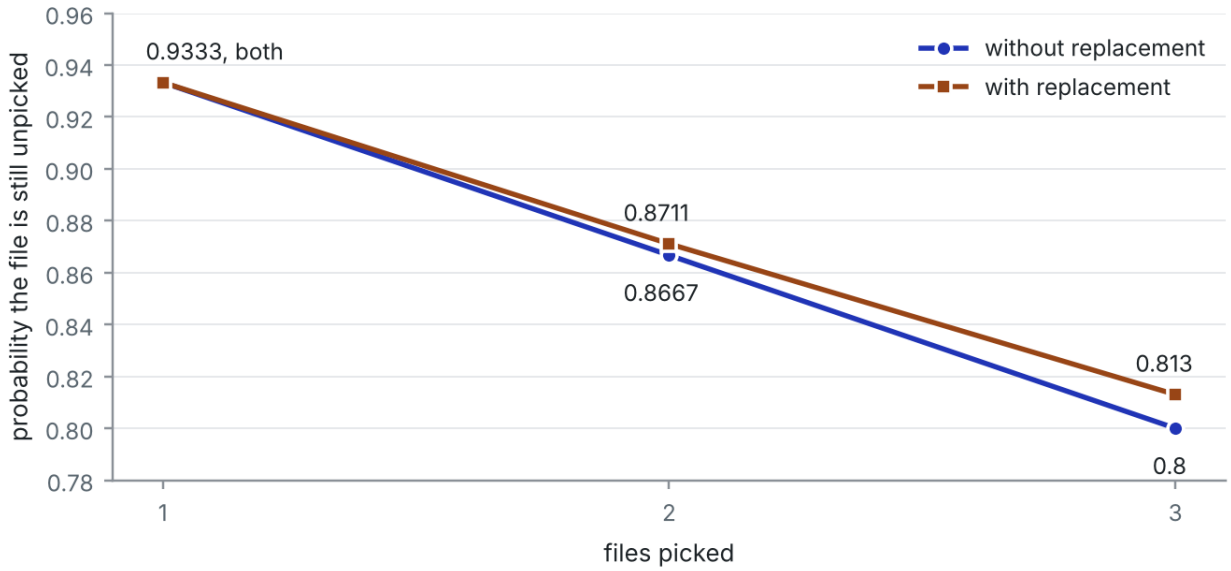
People who share a household, a clinic or a measurer are tied together. So are people drawn one after another from a small group. Before you multiply two chances, ask what ties them.



Two made-up tests, each catching 0.8 of people with a condition and wrongly flagging 0.1 of people without it, in 1,000 people of whom 300 have it. Within each group a positive first test leaves the second test's chance unchanged: 0.8 and 0.8 among the 300, 0.1 and 0.1 among the 700. Across all 1,000 it does not: 0.31 overall, 0.6419 given a positive first test.



Kiran's 131 men. If being flagged were independent of being in the band, the chance of being flagged would be 0.4504 among men in the band too, and 0.4504 times 0.3206, about 0.1444, of the men would be both. In the table it is 0.8571 among men in the band, and 0.2748 of the men are both. The two events are dependent.



The chance that one patient's file is still unpicked as the auditor draws from the ward's 15 files. Without replacement the chain 14 in 15, 13 in 14, 12 in 13 ends at 0.8 after 3 picks. With replacement every pick misses it with 14 in 15, and the chance ends at 0.813. Three files of 15 is 20 per cent of the ward, so the gap shows.

Illustration 1. Start with a check you can do on Kiran's table, rebuilt in Book 0, D3 for 131 men. Is "flagged by the arm tape" independent of "in the 25-29.9 band"? Compare the chance of being flagged overall with the chance among men in the band.

```
59 divided by 131 ≈ 0.4504
36 divided by 42 ≈ 0.8571
```

0.45 overall, 0.86 in the band. Knowing a man is in the band changes the chance he is flagged, so the two are dependent. That is what you want from a test. Check the product form too.

```
0.4504 times 0.3206 = 0.1444
36 divided by 131 ≈ 0.2748
```

If they were independent, about 0.14 of the men would be both. In fact 0.27 are.

Now a trap that looks like the opposite. Suppose two quick tests for the same condition are used one after the other on each person. The numbers are made up. In 1,000 people, 300 have the condition. Each test catches 80 per cent of people with it, and wrongly flags 10 per cent of people without it. Among people with the condition, the second test's result does not depend on the first's, and the same among people without it. That is conditional independence.

Count the people the first test flags, then the second.

```
1000 minus 300 = 700
0.8 times 300 = 240
0.8 times 240 = 192
0.1 times 700 = 70
0.1 times 70 = 7
240 plus 70 = 310
192 plus 7 = 199
```

Among the 300 with the condition, 240 are flagged by the first test and 192 by both. Among the 700 without, 70 by the first and 7 by both. In all, 310 people are flagged by the first test and 199 by both. The second test flags 310 too, by the same counting, so P(second positive) is 0.31. Now take the people the first test flagged.

```
310 divided by 1000 = 0.31
199 divided by 310 ≈ 0.6419
```

Overall, P(second positive) is 0.31. Given a positive first test, it is 0.64. Inside each group the two tests were independent, 0.8 and 0.8 among the sick, 0.1 and 0.1 among the well. Across everyone, they are strongly dependent. The first positive mostly picks out people with the condition, and among them the second test is positive far more often.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The two tests and their numbers are made up. Real tests on one person are often not independent even within the sick, because they respond to the same thing: two tape measures of body size both miss the person whose fat is not where they measure. Then the count above overstates what the second test adds, and a later section, on likelihood ratios, needs that warning.

Kiran's table is rebuilt from the paper's rounded percentages, for 131 men at one college.

Illustration 2. Dependence also comes from how people are picked. A ward has 15 patients, and an auditor picks 3 files, one after another, never the same file twice. This is drawing without replacement. What is the chance that one particular patient's file is not picked?

The first pick misses it with probability 14 in 15. If it was missed, 14 files remain, so the second pick misses it with probability 13 in 14. Then 12 in 13. The multiplication rule chains these conditionals.

```
P(A and B and C) = P(A) × P(B | A) × P(C | A and B)
```

This is the rule of section 2 used twice: first for A and B, then for that pair and C.

```
14 divided by 15 ≈ 0.9333
13 divided by 14 ≈ 0.9286
12 divided by 13 ≈ 0.9231
0.9333 times 0.9286 times 0.9231 = 0.8
```

Now suppose the auditor puts each file back, so the same file can be picked again. This is drawing with replacement, and each pick is independent.

```
0.93333 times 0.93333 times 0.93333 = 0.8130
```

0.800 against 0.813. OpenIntro Statistics works the same case with a professor and 15 students, in §3.3 (https://www.openintro.org/book/os/). Its rule of thumb: when the sample is under 10 per cent of the population, draws without replacement are nearly independent. Three of 15 is 20 per cent, so here the difference shows.

Shared circumstances tie observations the same way. Kiran and colleagues report, in their strengths and limitations, that "The strength of the study was single observer measurements". One observer means less scatter between measurers. It also means any habit of that one person, pulling the tape a little tight say, is in every one of the 131 men's arm measurements. Their errors share a cause, so they are not independent of one another. A second study with another measurer would not carry the same shift.

Figures quoted in this illustration are from [10, 27].

WHERE THIS PICTURE BREAKS. The ward and the auditor are made up, and the arithmetic is OpenIntro's example in another setting. The 10 per cent rule of thumb is OpenIntro's cut-off for when the difference starts to matter, not a law: how much it matters depends on what you then do with the answer.

Kiran's paper does not say its observer had any habit. The point is only that a shared measurer is a shared cause, the systematic error of Book 0, D7.

MUST KNOW POINTS FOR YOU.

- Two events that cannot both happen, each possible on its own, are always dependent: knowing one happened tells you the other did not. Before you multiply, ask "does knowing one change the chance of the other?", not "can they both happen?".
- Never compute "the chance both tests are positive" by multiplying the two tests' overall positive rates. Work inside each group, then add the groups.
- Two tests that respond to the same thing, such as two tape measures of body size, are probably not independent even among the sick. A second positive from such a test adds less than its numbers suggest.
- People in one household, one clinic, one school or measured by one person share causes. Before you multiply their probabilities, or count them as separate pieces of evidence, ask what ties them, and say so in your methods. Count the joint outcome from your own data rather than multiplying, and treat the household or clinic, not the person, as the unit when you work out how sure you can be.
- Drawing without replacement from a small group makes each draw depend on the ones before.
- A very small product, such as "one in many millions", should send you looking for what ties the events together before anyone quotes it.

COMMON MISREADING. "The two tests are independent, so the chance both are positive is the product of their positive rates." In the made-up example each test flags 0.31 of the 1,000 people,

and the product is about 0.096. In fact 199 of the 1,000, 0.199, are flagged by both. The tests were independent only within the sick and within the well. Across everyone, a positive first result picks out the sick, and the second result follows.

EXERCISE 1 (retrieval). From memory: give the definition of independence as a conditional probability, and the check that follows from it. Then explain in two sentences why two mutually exclusive events, each possible, cannot be independent.

EXERCISE 2 (critique). A protocol for a made-up screening study says: "Two family members were enrolled from each household. Since each person has a 25 per cent chance of having overweight, the chance that both members of a household have overweight is 0.25 × 0.25, about 6 per cent." Find the flaw, and say what the protocol should do instead.

PRACTICE. Nine problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. P(A) is 0.4, P(B) is 0.5 and P(A and B) is 0.2. Are A and B independent? Work out P(A | B).

2. P(A) is 0.3 and P(B) is 0.2, and A and B are mutually exclusive. Work out P(A and B), P(A) × P(B) and P(A | B). Are A and B independent?

3. A box holds 5 slips, 2 of them marked. Two slips are drawn. Work out the probability that both are marked when the first is not put back, and when it is put back before the second draw.

4. In Kiran's rebuilt table of 131 men, 42 are in the 25-29.9 band, 59 are flagged by the arm tape, and 36 are both. Test, by both forms of the definition, whether "flagged" and "in the band" are independent.

5. A made-up ward has 12 patients, 3 of them with diabetes. An auditor picks 2 patients' files for review, without picking the same file twice. Work out the probability that both files belong to patients with diabetes. Then work it out as if the two picks were independent, and say why the two answers differ.

6. In a made-up group of 1,000 people, 200 have a condition. Two tests are used on everyone. Each catches 90 per cent of people with the condition and wrongly flags 20 per cent of people without it. Within each group, the two tests' results are independent.

Work out how many people the first test flags, how many both tests flag, and P(second test positive | first test positive). Compare it with P(second test positive).

7. Here is a worked answer. Find the step that broke.

```
the arm tape misses 14 per cent of men in the 25-29.9 band (sensitivity 86 per cent)
a man in the band is measured twice, on two days, by the same observer
P(missed both times) = 0.14 times 0.14 = 0.0196
so measuring twice cuts the chance of a miss to about 2 per cent
```

8. Here is a worked answer. Find the step that broke.

```
A and B are independent, with P(A) = 0.3 and P(B) = 0.4
independent events do not overlap, so P(A and B) = 0
P(A or B) = 0.3 plus 0.4 = 0.7
```

9. A district note says: “NFHS-6 (2023-24) finds 30.7% of women and 27.3% of men aged 15-49 overweight or obese. So in about 8.4 per cent of married couples, both partners have overweight or obesity.”

Decide what to compute, compute it, and say what the note’s number does not establish.

4 · The law of total probability: a probability as a weighted average over groups

DEFINITION. Split a population into groups that do not overlap and that between them take in everyone. Call the groups group 1, group 2, and so on up to group m. Let B be any event.

The law of total probability says that the probability of B is a sum with one term for each group. Each term is the probability of B given that group, times the probability of that group.

```
P(B) = P(B | group 1) × P(group 1) + P(B | group 2) × P(group 2) + … + P(B | group m) × P(group m)
```

It follows from two rules already in hand. First, B splits into “B and group 1”, “B and group 2”, and so on. No two of these pieces overlap, because no two groups do. Together they make up all of B, because the groups take in everyone. Pieces that cannot happen together add, however many there are. That is the third rule of section 1.

```
P(B) = P(B and group 1) + P(B and group 2) + … + P(B and group m)
```

Second, the general multiplication rule of section 2 writes each piece as a conditional probability times a group’s probability: P(B and group 1) = P(B | group 1) × P(group 1). Put each of these into the sum, and the law appears.

The group probabilities P(group 1) to P(group m) are never negative and add to 1. So P(B) is a weighted average of the groups’ own probabilities of B, each weighted by the group’s share of the population. With every group the same size, it is their plain average. Otherwise it generally is not.

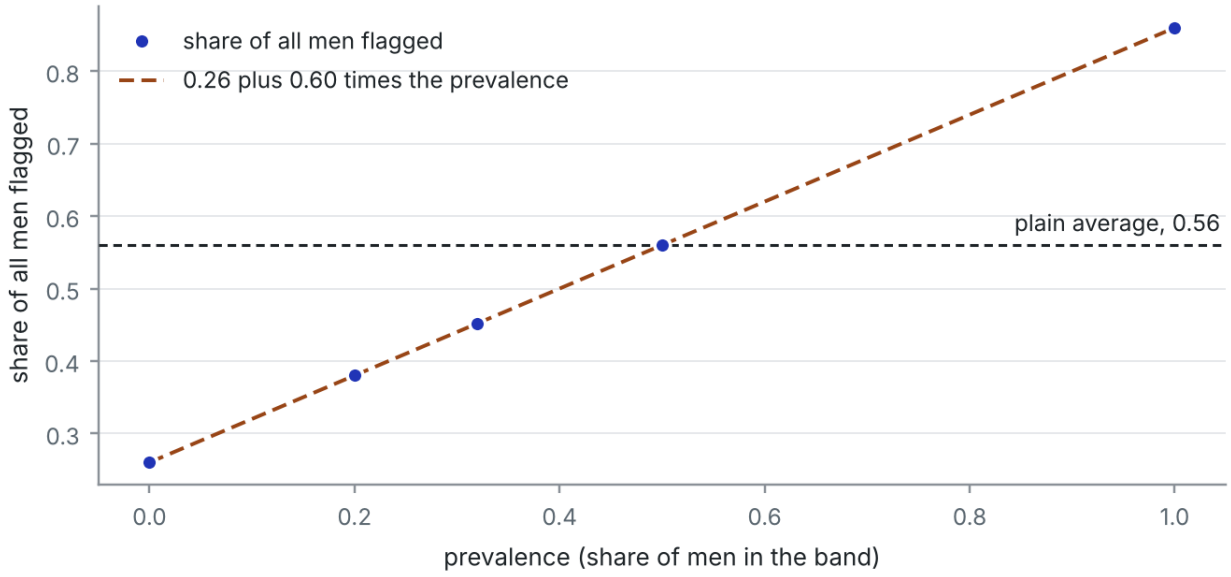
For a test and a condition the groups are two: the people with the condition and the people without it. Their shares are the prevalence and 1 minus the prevalence. The probability of a positive result given the condition is the sensitivity of Book 0, D3. The probability of a positive result given no condition is 1 minus the specificity, as in section 2. It is called the false-positive rate: the share of people without the condition whom the test wrongly flags.

```
P(positive) = sensitivity × prevalence + (1 - specificity) × (1 - prevalence)
```

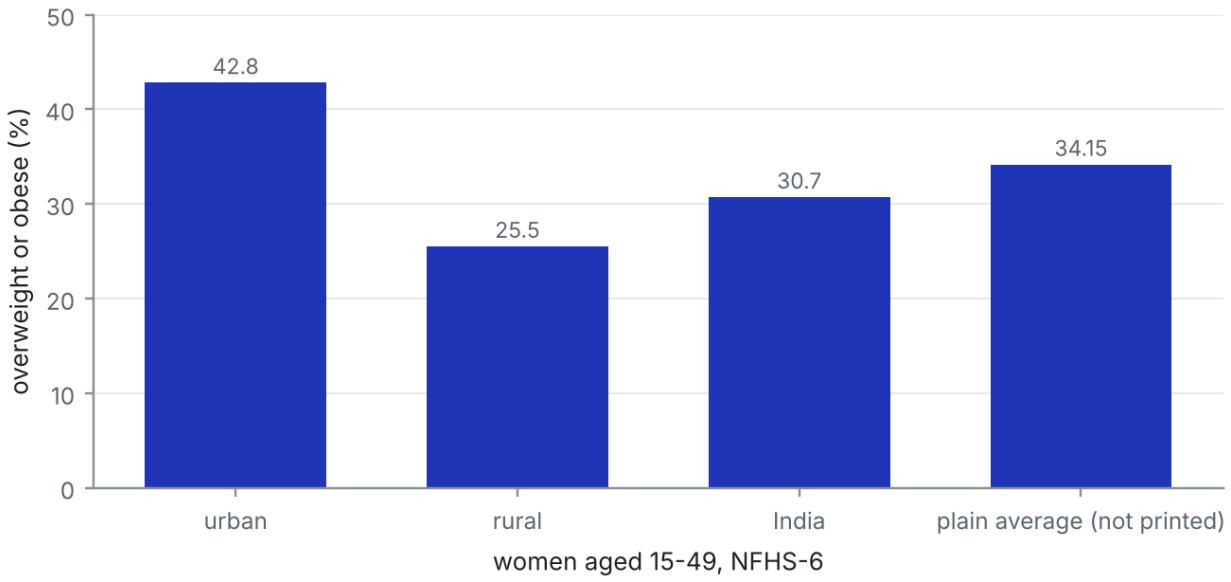
[28, 29, 30, 31]

IN PLAIN TERMS. Some people are in one group and the rest are in another. Each group has its own chance of something, such as a positive test. The chance for everyone together is a mix of the two.

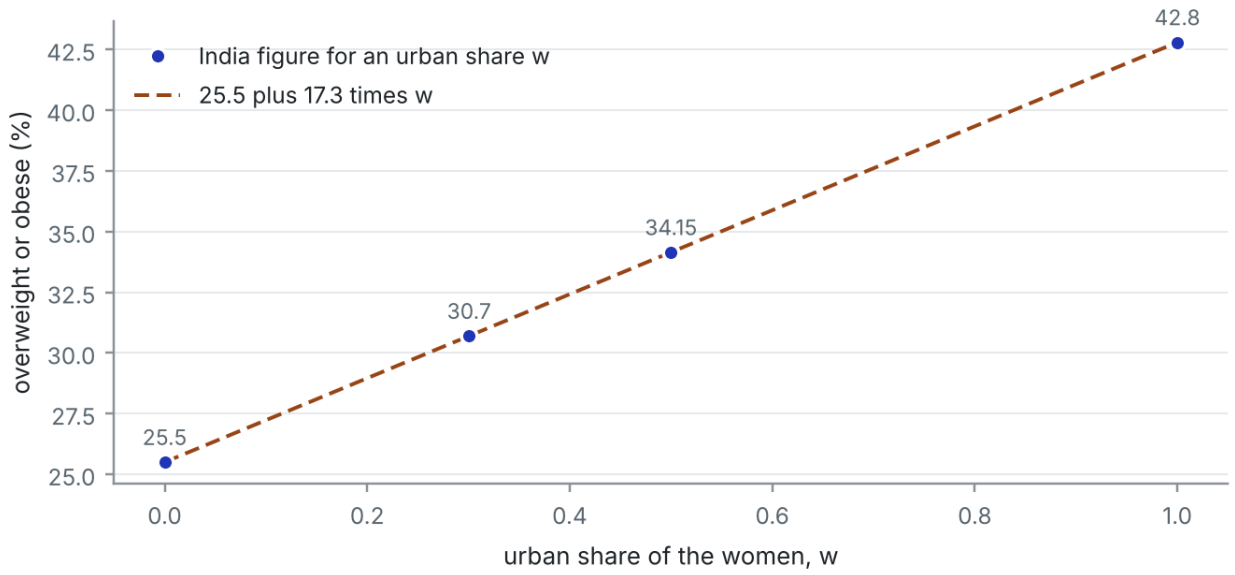
It is not the halfway point. It sits nearer the group that has more people in it.



The share of all men the arm tape flags, against the share of men in the 25-to-29.9 band. With sensitivity 0.86 and a false-positive rate of 0.26 fixed, the share flagged is 0.26 plus 0.60 times the prevalence: 0.38 at one man in five, 0.452 at about 0.32 as among the 131 students. The plain average of the two rates, 0.56, is right only at a prevalence of 0.5.



Women aged 15-49 who have overweight or obesity, NFHS-6 (2023-24): 42.8% urban, 25.5% rural, 30.7% for India. The India figure is a weighted average of the other two and sits nearer the rural one. The plain average, 34.15% (right-hand bar), is not what the survey prints. The three figures imply that about 0.30 of the women were urban.



Running the rule backwards. If w is the urban share of the women (NFHS-6), the India figure is 25.5 plus 17.3 times w: a straight line from the rural 25.5 per cent at w = 0 to the urban 42.8 per cent at w = 1. The printed 30.7 sits on it at w = 0.3006, about 30 per cent. The plain average, 34.15, would need w = 0.5.

Illustration 1. Book 0 took the arm-circumference test of Kiran and colleagues (Heliyon 2022, https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/) to a stated group of 1,000 young men, one in five of them with a body-mass index from 25 to 29.9. A man whose arm measures 31.3 cm or more is flagged. The paper gives sensitivity 86 per cent and specificity 74 per cent. Book 0 counted 380 of the 1,000 flagged. Now find that share without the table.

Before you work it, here is the move most people reach for. Take the two rates the test has, the 86 per cent of men in the band it flags and the share of men outside the band it flags, and average them. The second rate is the false-positive rate.

```
1 minus 0.74 = 0.26
0.86 plus 0.26 = 1.12
1.12 divided by 2 = 0.56
```

That says 56 per cent of men are flagged. It is wrong, and Book 0's table says 38 per cent. The plain average treats the two groups as the same size. They are not: 200 men are in the band and 800 are not.

Weight each rate by its group's share instead.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
```

The share flagged is 0.38, which is 380 of 1,000, as the table had it. The 0.38 sits much nearer 0.26 than 0.86, because four men in five are outside the band.

Now let the share in the band change and keep the test the same. The weights move, and the share flagged moves with them. Among the 131 medical students themselves, 42 were in the band. That is a prevalence of about 0.32, so about 0.68 were outside it.

```
42 divided by 131 = 0.3206…
0.86 times 0.32 = 0.2752
0.26 times 0.68 = 0.1768
0.2752 plus 0.1768 = 0.452
0.452 times 131 = 59.21
```

About 59 of the 131 men flagged. Book 0's rebuilt table for these men has 59 in its flagged row: 36 in the band and 23 outside it.

Do it for a few prevalences, and a pattern shows. Each 0.1 added to the prevalence moves one tenth of the men from the 0.26 group to the 0.86 group. So the share flagged rises by 0.1 times 0.60 each time, where 0.60 is 0.86 minus 0.26.

prevalence	share of all men flagged
0	0.26
0.2	0.38
0.32	0.452
0.5	0.56
1	0.86

The share flagged is 0.26 plus 0.60 times the prevalence: a straight line from the false-positive rate to the sensitivity. The plain average, 0.56, is right at one prevalence only, a half.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The sensitivity and specificity were measured on 131 medical students in one college, and they fit the 42 men in the 25-to-29.9 band, not every man at 25 or more. Carrying them to 1,000 other young men is an assumption, as Book 0 said. The straight line holds only while the two rates stay fixed as the prevalence changes. A test whose sensitivity is higher in a group with more severe disease breaks it.

Illustration 2. The same rule works on a survey table. The National Family Health Survey 2023-24 (NFHS-6) India fact sheet (https://www.nfhsiips.in), indicator 76, gives the share of women aged 15 to 49 with a body-mass index of 25 or more, which the survey calls overweight or obese. It gives three figures.

women aged 15-49	overweight or obese (%)
urban	42.8
rural	25.5
India	30.7

The groups are urban and rural women. They do not overlap, and between them they take in every woman the indicator covers. So the India figure must be a weighted average of the other two. Check that it is not the plain average.

```
42.8 plus 25.5 = 68.3
68.3 divided by 2 = 34.15
```

The plain average is 34.15 per cent, and the survey prints 30.7. The India figure sits nearer the rural one, because more of the women the indicator covers are rural.

Now run the rule backwards and find the weight. Call the urban share of the women w. Then the rural share is 1 minus w, and the law of total probability says this.

```
30.7 = 42.8 × w + 25.5 × (1 - w)
30.7 = 42.8 × w + 25.5 - 25.5 × w
```

The two terms in w combine: 42.8 × w minus 25.5 × w is (42.8 − 25.5) × w.

```
30.7 = 25.5 + (42.8 - 25.5) × w
```

Take 25.5 from both sides, then divide by the gap between the two groups.

```
30.7 minus 25.5 = 5.2
42.8 minus 25.5 = 17.3
5.2 divided by 17.3 = 0.3006
```

The three printed figures imply that about 0.30 of the women, 30 per cent, were urban. Check it by putting the weight back.

```
42.8 times 0.3006 = 12.8657
25.5 times 0.6994 = 17.8347
12.8657 plus 17.8347 = 30.7004
```

That rounds to 30.7. The weight is only as good as the rounding, though. Each printed figure could be up to 0.05 away from its true value. The weight is (India minus rural) divided by (urban minus rural). So it is largest with the India figure at its highest and the other two at their lowest, and smallest the other way round. The figures 30.75, 25.45 and 42.75 give 5.3 divided by 17.3, which is 0.306. The figures 30.65, 25.55 and 42.85 give 5.1 divided by 17.3, which is 0.295. So say “about 30 per cent”, not 30.06 per cent.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The weight found is the urban share that the three printed percentages imply among the women the indicator covers. It is not a count of India’s urban women, and the indicator leaves out pregnant women and women who gave birth in the two months be-

fore. The backwards step inherits the printed figures' rounding: a gap of 17.3 points in the bottom line turns a rounding of 0.05 in each figure into about 0.006 in the weight.

MUST KNOW POINTS FOR YOU.

- "The positive rate is somewhere between the sensitivity and the false-positive rate, so call it the middle" is the error. Where it sits depends on the prevalence. When most people are well, most positive results come from the false-positive rate, however good the sensitivity.
- The groups you split by must not overlap and must take in everyone. Write the groups down and check both conditions before you multiply anything.
- When a national figure is quoted beside its urban and rural figures, the national one is a weighted average and must lie between them. If it does not, one of the three is wrong or they cover different people.
- A national rate can rise when no group's rate rises, because the weights moved: more people in the group with the higher rate.
- Working backwards, from a total and two group rates to the weight, divides by the gap between the group rates. When that gap is small, a rounding of the last digit can move the answer a great deal.
- The positive rate of a test in a camp tells you the workload of confirming the positives. It is not the prevalence.

COMMON MISREADING. "The test flags 86 per cent of those with the condition and 26 per cent of those without, so it flags about 56 per cent of everyone." That averages the two rates as if the two groups were the same size. Weight each rate by its group's share. At one man in five in the band, the tape flags 38 per cent, not 56.

EXERCISE 1 (retrieval). From memory: state the law of total probability in words, for groups that do not overlap and take in everyone. Then write the share of all tests that come back positive in terms of the sensitivity, the specificity and the prevalence, and say what makes it a weighted average.

EXERCISE 2 (interpretation). A district report gives the share of women aged 15 to 49 who have overweight or obesity as 31 per cent in urban areas, 18 per cent in rural areas, and 33 per cent for the whole district. These figures are made up. Say what is wrong, and what you would ask before quoting any of them.

PRACTICE. Nine problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A population is split into two groups that do not overlap and take in everyone. Group 1 holds 0.3 of the population and group 2 the rest. P(B | group 1) is 0.9 and P(B | group 2) is 0.2.

Work out P(B).

2. Three groups that do not overlap take in everyone. Their shares are 0.5, 0.3 and 0.2. The probability of B in each is 0.1, 0.4 and 0.7.

Work out P(B), and compare it with the plain average of 0.1, 0.4 and 0.7.

3. Two groups that do not overlap take in everyone. P(B | group 1) is 0.8 and P(B | group 2) is 0.1. P(B) for everyone is 0.24.

Work out the share of the population in group 1.

4. In Kiran and colleagues' study, 42 of the 131 men had a body-mass index from 25 to 29.9. The 31.3 cm arm cut-off has sensitivity 86 per cent and specificity 74 per cent against that band.

Use the law of total probability to work out how many of the 131 men you expect the cut-off to flag. Do not build the table.

5. The NFHS-6 India fact sheet, indicator 77, gives the share of men aged 15 to 49 who have overweight or obesity: 36.3 per cent urban, 23.0 per cent rural and 27.3 per cent for India.

Work out the urban share of the men that these three figures imply. Then say how far the rounding of the printed figures lets that share move.

6. Here is a worked answer. Find the step that broke.

```
a test has sensitivity 0.86 and specificity 0.74
in a group of 1,000 men, 200 have the condition
false-positive rate = 1 minus 0.74 = 0.26
share flagged = (0.86 plus 0.26) divided by 2 = 0.56
so about 560 of the 1,000 men will be flagged
```

7. Here is a worked answer. Find the step that broke.

```
a test has sensitivity 0.86 and specificity 0.74
prevalence 0.2
P(positive) = 0.86 times 0.2 plus 0.74 times 0.8
= 0.172 plus 0.592 = 0.764
so about three in four people tested will be positive
```

8. A screening-camp summary says this.

> *Of the 500 young men screened with the arm tape, 190 were positive. So 38 per cent of young men in this area are overweight.*

The tape is the 31.3 cm cut-off, with sensitivity 86 per cent and specificity 74 per cent against a body-mass index of 25 to 29.9. The camp's counts are made up. Decide what to compute, compute it, and say what your answer does not establish.

9. A press line says this.

> *The share of women who are overweight rose in the district between the two surveys. So it must have risen in the towns, or in the villages, or in both.*

Take a made-up district where the share is 30 per cent in towns and 15 per cent in villages in both surveys, and towns hold 0.3 of the women in the first survey and 0.4 in the second. Decide what to compute, compute it, and say what your answer does not establish.

5 · Bayes' theorem, from the multiplication rule, on a diagnostic test

DEFINITION. Let A and B be events, with P(B) above 0. The general multiplication rule of section 2 builds the joint probability P(A and B) in two orders.

```
P(A and B) = P(B) × P(A | B)
P(A and B) = P(A) × P(B | A)
```

The two right-hand sides are the same number, so they are equal. Divide both by P(B).

```
P(A | B) = P(B | A) × P(A) ÷ P(B)
```

This is Bayes' theorem. It turns a conditional probability round.

P(A) is the prior probability of A: its probability before you learn whether B happened. P(A | B) is the posterior probability: the same probability, revised once you know that B happened.

For a test, let A be "has the condition" and B be "tests positive". The prior is the prevalence. P(B | A) is the sensitivity. P(B | not A) is the false-positive rate, 1 minus the specificity. The posterior is the positive predictive value, PPV, the share of positive results that are right (Book 0, D3). Put these into Bayes' theorem.

```
PPV = sensitivity × prevalence ÷ [sensitivity × prevalence + (1 - specificity) × (1
- prevalence)]
```

The square brackets group everything inside them, as round brackets do.

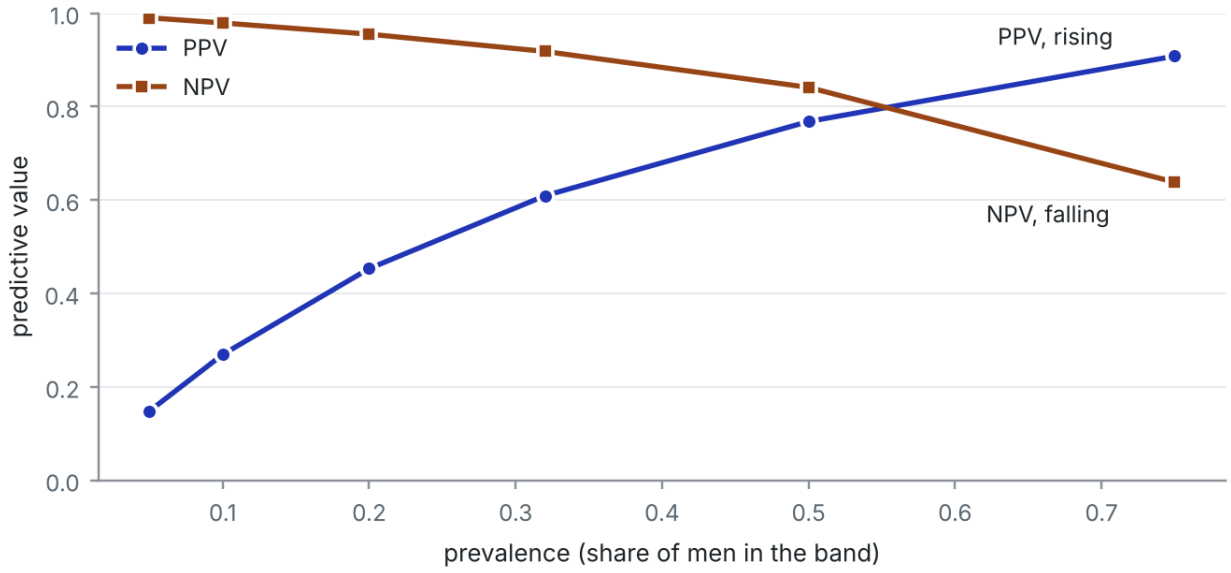
The same steps, with "does not have the condition" in place of "has the condition" and "tests negative" in place of "tests positive", give the negative predictive value, NPV, the share of negative results that are right.

```
NPV = specificity × (1 - prevalence) ÷ [(1 - sensitivity) × prevalence + specificity
× (1 - prevalence)]
```

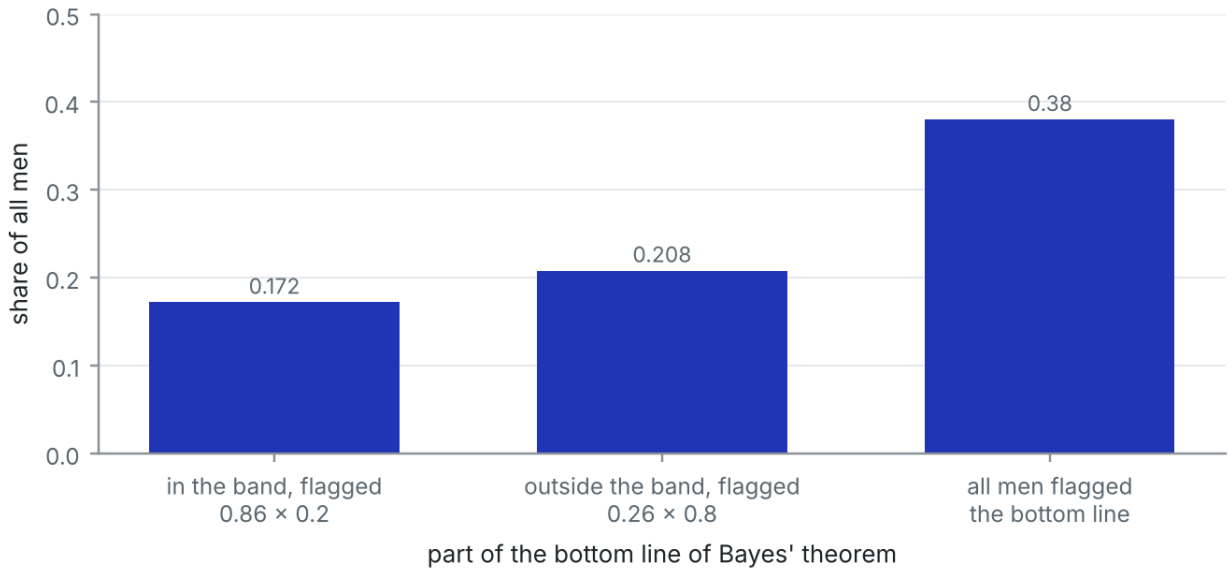
[33, 34, 35, 36, 37, 38]

IN PLAIN TERMS. A test result does not tell you the chance a person has the condition. It changes the chance they already had.

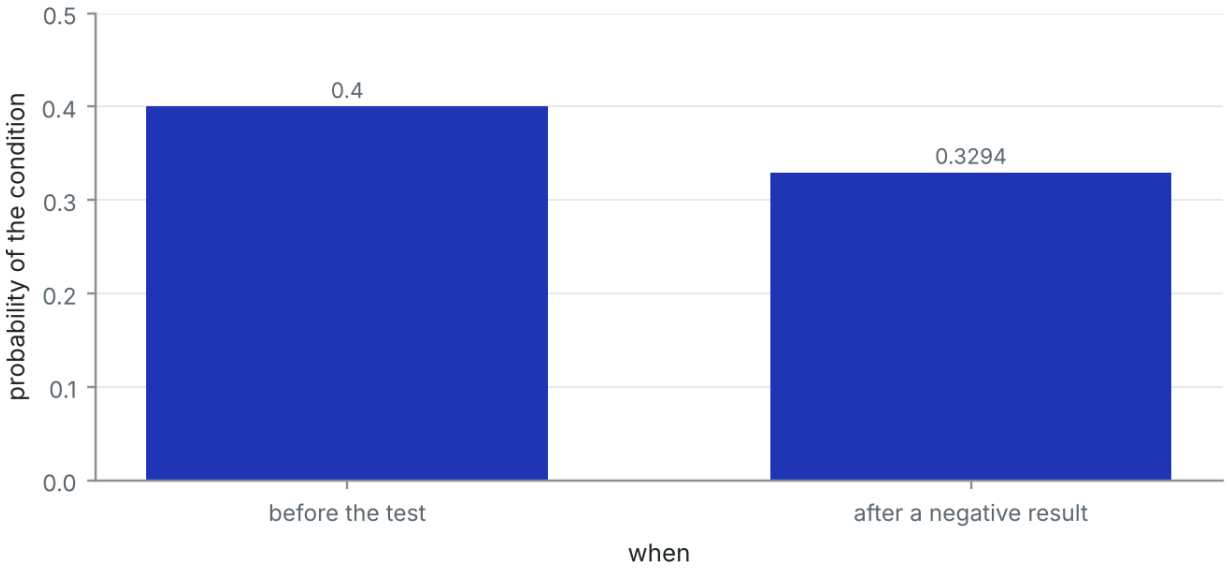
Bayes' theorem does the moving. It needs three numbers: how common the condition is, how often the test is positive in people who have it, and how often in people who do not.



The arm tape's positive and negative predictive values against the share of men in the 25-to-29.9 band, with sensitivity 0.86 and specificity 0.74 held fixed. The PPV climbs from 0.148 at a prevalence of 0.05 to 0.908 at 0.75; the NPV falls from 0.990 to 0.638. At 0.2 the PPV is 0.453, Book 0's 45 per cent; at 0.32 it is 0.609, Book 0's 61 per cent.



At a prevalence of 0.2, the share of all men flagged, 0.38, is made of two parts: 0.172 from men in the band (0.86 times 0.2) and 0.208 from men outside it (0.26 times 0.8). The PPV is the first part's share of the whole: 0.172 divided by 0.38, about 0.4526.



A made-up test with sensitivity 0.3 and specificity 0.95, used where 0.4 of people have the condition. A negative result moves the probability only from 0.4 to about 0.3294, because the test misses seven in ten of the people who have it. One person in three who tests negative still has the condition.

Illustration 1. Book 0 found by counting that the arm-circumference test of Kiran and colleagues (Heliyon 2022, https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/) has a positive predictive value of about 45 per cent in a group of 1,000 young men where one in five has a body-mass index from 25 to 29.9. A man is flagged when his arm measures 31.3 cm or more. The sensitivity is 86 per cent and the specificity 74 per cent. Now get the same answer with no table, from the formula.

Write down the three inputs first, and say each one aloud as a conditional.

```
prevalence = 0.2
sensitivity = 0.86
1 minus 0.74 = 0.26
1 minus 0.2 = 0.8
```

The 0.26 is the false-positive rate, the share of men outside the band whom the tape wrongly flags. Work out the top line, then the bottom line, then divide.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
0.172 divided by 0.38 = 0.4526
```

The PPV is about 0.45, Book 0's 45 per cent. The bottom line, 0.38, is the share of all men flagged, which the last section worked out. The top line is the part of it made of men in the band.

Do the NPV the same way. The top line is the men outside the band whom the tape clears. The bottom line is every man it clears.

```
0.74 times 0.8 = 0.592
1 minus 0.86 = 0.14
0.14 times 0.2 = 0.028
0.592 plus 0.028 = 0.62
0.592 divided by 0.62 = 0.9548
```

The NPV is about 0.95. Of the men the tape clears, 95 in 100 are outside the band.

Now change only the prevalence. Among Kiran's 131 medical students, 42 were in the band, about 0.32. The same formulas, worked the same way at each prevalence, give this.

prevalence	PPV	NPV
0.05	0.148	0.990
0.1	0.269	0.979
0.2	0.453	0.955
0.32	0.609	0.918
0.5	0.768	0.841
0.75	0.908	0.638

At 0.32 the PPV is about 0.61. That is Book 0's 61 per cent for the medical students, recovered by the formula. Book 0's table gave 36 divided by 59, which is 0.610; the small gap is the table's rounding to whole men. The tape is the same tape in every row. Only the prevalence moved, and the PPV went from about one in seven to about nine in ten.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. Every row uses the sensitivity and specificity measured on 131 medical students in one college, and they fit the 42 men in the 25-to-29.9 band, not every man at 25 or more. The table assumes the tape flags men in and out of the band at those same rates in every group. In a group with heavier men, or a different measurer, it may not, and then the row for that prevalence is wrong even though the arithmetic is right.

Illustration 2. A negative result moves the probability too, and how far depends on the sensitivity. Take a made-up test with sensitivity 0.3 and specificity 0.95, used on people of whom 0.4 have the condition. All three numbers are made up. A person tests negative. What is the probability they have the condition now?

That is 1 minus the NPV: the condition and a negative result, divided by all negative results.

```
1 minus 0.3 = 0.7
0.7 times 0.4 = 0.28
1 minus 0.4 = 0.6
0.95 times 0.6 = 0.57
0.28 plus 0.57 = 0.85
0.28 divided by 0.85 = 0.3294
```

Before the test the probability was 0.40. After a negative result it is about 0.33. The test missed seven in ten of the people with the condition, so a negative result is common among them, and it

says little. Saying "the test was negative, so it is ruled out" would discharge a group of whom about one in three has the condition, and would send home seven in ten of everyone who has it.

WHERE THIS PICTURE BREAKS. The test is made up, to show the shape of the result. A real test's sensitivity is an estimate from one study, and it can be lower in early or mild disease, which is where a negative result is most often relied on.

MUST KNOW POINTS FOR YOU.

- Sensitivity is P(positive | condition). The patient's question is P(condition | positive).
- Never quote a PPV or NPV without the prevalence it was worked out at. Recompute it from the sensitivity, the specificity and your own prevalence before you tell a patient what a positive result means.
- When the condition is rare, a positive result from a good test can still be wrong more often than right. Plan a confirming measurement for every positive before anyone is labelled.
- A negative result from a test with low sensitivity barely moves the probability. Do not write "ruled out" after a negative result unless the sensitivity is high.
- Bayes' theorem is exact. Its answer is only as good as its three inputs, and each is an estimate from some group.
- Before you substitute, say each number as a conditional aloud: "positive given the condition", "positive given no condition".

COMMON MISREADING. "The test is 86 per cent sensitive and 74 per cent specific, so its results are right about eight times in ten." Neither number is the chance that a result is right. That chance is the PPV for a positive result and the NPV for a negative one, and both move with the prevalence. The same tape's positive results are right about 15 times in 100 at one man in twenty, and about 91 times in 100 at three in four.

REPORTING IT. Kiran and colleagues report sensitivity 86 per cent and specificity 74 per cent for the 31.3 cm arm cut-off in medical students. Where 20 per cent of men have a body-mass index of 25 to 29.9, about 45 in every 100 men it flags would be in that range. The rest would be outside that range, some of them above it.

EXERCISE 1 (retrieval). From memory: derive Bayes' theorem from the two ways of writing P(A and B). Then write the PPV in terms of the sensitivity, the specificity and the prevalence, and say which of the four quantities, sensitivity, specificity, PPV and NPV, change when the prevalence changes.

EXERCISE 2 (teaching). A first-year resident, ten minutes, a whiteboard. She says: "The tape has 86 per cent sensitivity, so a man it flags is 86 per cent likely to be overweight." Plan what you

will draw and say, so that she can work out the right number herself at a camp where one man in five is in the band.

PRACTICE. Twelve problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. P(A) is 0.3, P(B | A) is 0.6 and P(B) is 0.25.

Work out P(A | B).

2. P(A) is 0.4, P(B | A) is 0.5 and P(B | not A) is 0.2.

Work out P(A | B).

3. A test has sensitivity 0.9 and specificity 0.8. The prevalence is 0.1.

Work out the PPV and the NPV.

4. Among Kiran and colleagues' 131 men, 42 had a body-mass index from 25 to 29.9. The 31.3 cm arm cut-off has sensitivity 86 per cent and specificity 74 per cent.

Work out the PPV and the NPV among these men from the formulas. Compare them with Book 0's rebuilt table, which has 36 of 59 flagged men in the band and 66 of 72 cleared men outside it.

5. A test has sensitivity 0.8 and specificity 0.9. In some group its PPV is 0.5.

Work out the prevalence in that group.

6. Altman and Bland's note on predictive values (BMJ 1994) uses a liver scan with sensitivity 0.895 and specificity 0.628, in patients of whom 0.75 had abnormal pathology.

Work out the PPV at that prevalence. Then work out the PPV and NPV at a prevalence of 0.25, and compare them with the note's printed 0.45 and 0.95.

7. Suppose the arm tape, with sensitivity 86 per cent and specificity 74 per cent, is used where one man in twenty is in the 25-to-29.9 band. The one in twenty is made up.

Work out the PPV and the NPV, and say in one sentence what a positive result means there.

8. Mohan and colleagues' Indian Diabetes Risk Score (J Assoc Physicians India 2005), at a score of 80 or more, has sensitivity 15.1 per cent and specificity 95.0 per cent for undiagnosed diabetes (Table 2). The paper's discussion takes the prevalence of undiagnosed diabetes as 10.1 per cent.

A man scores below 80. Work out the probability that he has undiagnosed diabetes, and compare it with the prior.

9. Here is a worked answer. Find the step that broke.

```
arm tape: sensitivity 0.86, specificity 0.74, prevalence 0.2
PPV = 0.74 times 0.2 divided by (0.74 times 0.2 plus 0.14 times 0.8)
= 0.148 divided by (0.148 plus 0.112)
= 0.148 divided by 0.26 = 0.569
```

10. Here is a worked answer. Find the step that broke.

```
among Kiran's medical students the arm tape's PPV was about 0.61
our camp uses the same tape on young men, of whom about one in five is in the band
the tape is the same, so its PPV is the same
so about 61 in every 100 men we flag will be in the band
```

11. A patient leaflet for a screening camp says this.

> *The arm tape finds 86 per cent of young men who are overweight. If your result is positive, you are overweight.*

The camp expects about one man in five to be in the 25-to-29.9 band; that figure is made up. The tape's specificity is 74 per cent. Decide what to compute, compute it, and say what your answer does not establish.

12. A colleague planning a rural camp says this.

> *In the sixth National Family Health Survey, NFHS-6 (2023-24), 23.0 per cent of rural men aged 15 to 49 were overweight or obese. The arm tape has a PPV of 61 per cent in Kiran's study, so about 61 per cent of the men it flags at our camp will be overweight.*

Decide what to compute, compute it, and say what your answer does not establish.

6 · The base-rate case that defeats most clinicians

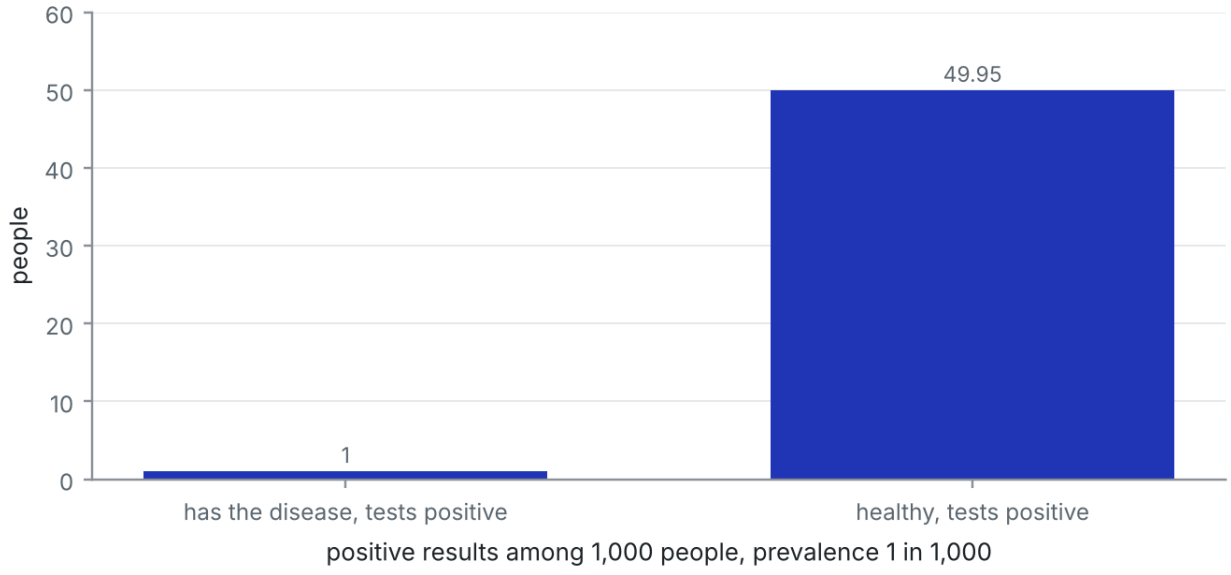
DEFINITION. The base-rate case is a measured finding about how clinicians read a positive test. The base rate is the prevalence of the condition in the group being tested, before any test is done: the prior probability of section 5.

In 1978 Casscells, Schoenberger and Graboys put one question to 60 students and doctors met in hallway encounters at four Harvard teaching hospitals. A test for a disease whose prevalence is 1/1000 has a false-positive rate of 5 per cent. What is the chance that a person with a positive result has the disease, knowing nothing else about them?

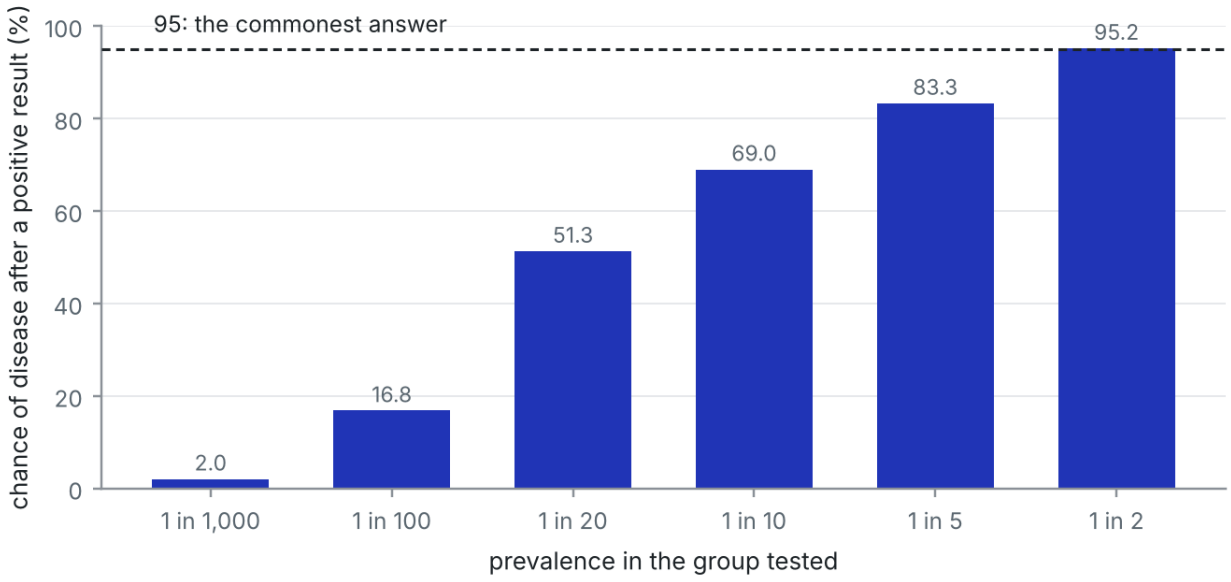
If the test catches everyone who has the disease, Bayes' theorem gives about 2 per cent: one person with the disease among about 51 positives. 11 of the 60 gave that answer. The commonest answer, given by 27, was 95 per cent.

The answer 95 per cent is 1 minus the false-positive rate. That is P(test negative | no disease), the specificity. Given as the answer, it is read as P(disease | test positive). The prevalence has dropped out of the reasoning altogether. [39, 40, 41, 42, 42, 43, 43, 44, 44]

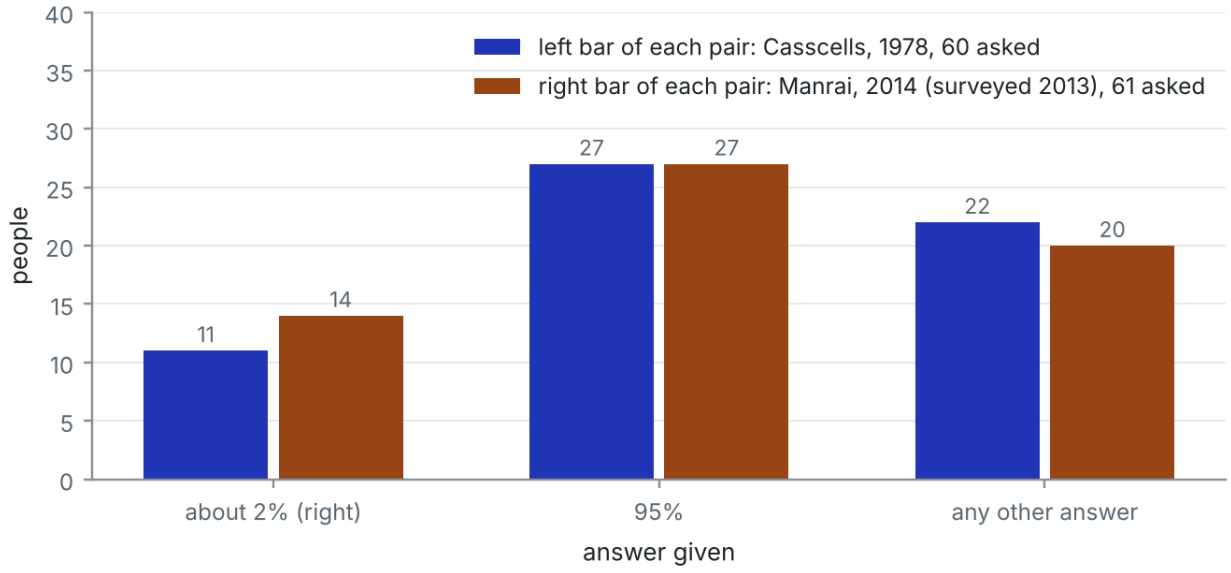
IN PLAIN TERMS. Most doctors asked this question give an answer near 95 per cent. The right answer is about 2 per cent. The disease is so rare that the few ill people who test positive are swamped by the many healthy people who test positive too.



The Casscells question counted in 1,000 people. The 1 person with the disease tests positive, and so do 5 per cent of the 999 without it, 49.95 people. One person with the disease among about 51 positives: a chance of 0.0196, about 2 per cent, not 95.



The same test at each prevalence of the table. It catches everyone with the disease and wrongly flags 5 per cent of the rest. The chance of disease after a positive result is 2.0 per cent at 1 in 1,000, the Casscells question, 16.8 per cent at 1 in 100, and reaches 95.2 per cent only at 1 in 2. The dashed line marks 95, the commonest answer.



Answers to the same question in two studies. In 1978 Casscells and colleagues found 11 of 60 answering about 2 per cent and 27 answering 95 per cent. Manrai and colleagues, 2014 (surveyed 2013), found 14 of 61 and 27. The rest, 22 and 20, gave other answers.

Illustration 1. Answer the question yourself before you read on. Write your answer down, and how sure you are of it, as a per cent. This is the question exactly as Casscells and colleagues put it in the *New England Journal of Medicine* in 1978 (https://doi.org/10.1056/NEJM197811022991808):

> *If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?*

If your answer is near 95 per cent, you are in good company. 27 of their 60 students and doctors said exactly that. Now count it. Take 1,000 people like the person in the question.

```
1000 divided by 1000 = 1
1000 minus 1 = 999
0.05 times 999 = 49.95
```

One person has the disease. The question does not say how often the test catches it, so assume, as the authors did, that it always does. That one person tests positive. Of the 999 without the disease, 5 per cent test positive anyway: about 50 people.

```
1 plus 49.95 = 50.95
1 divided by 50.95 = 0.0196
```

About 51 people test positive, and one of them has the disease. The chance is about 2 per cent, the authors' "one in 51". Only 11 of the 60 gave it.

The same answer comes from Bayes' theorem in section 5, with sensitivity 1, prevalence 0.001 and specificity 0.95.

```
1 times 0.001 = 0.001
0.05 times 0.999 = 0.04995
0.001 plus 0.04995 = 0.05095
0.001 divided by 0.05095 = 0.0196
```

The denominator, 0.05095, is the share of all people tested who come back positive. That is the law of total probability of section 4.

Figures quoted in this illustration are from [45, 46].

WHERE THIS PICTURE BREAKS. The 2 per cent holds only for a person about whom nothing else is known, drawn from a group where 1 in 1,000 has the disease. Casscells and colleagues say so themselves: if the person were known to have symptoms, the chance "would be higher". A patient with symptoms belongs to a group with a higher prevalence.

It also assumes the test catches everyone with the disease. The question does not say so. A test that missed half the cases would give about 1 per cent, not 2. That is still nowhere near 95.

Illustration 2. Where does 95 per cent come from? Take 1 from the false-positive rate.

```
1 minus 0.05 = 0.95
```

That is the specificity: P(test negative | no disease). It is a share of the 999 healthy people. The question asked for a share of the 51 people who tested positive. The two share no denominator at all. This is the trap of Book 0's two-way table, Book 0, D3, in its most costly form.

Here is a second way to see it. Ask what prevalence would make 95 per cent the right answer. Keep the test the same, and work out the chance of disease after a positive result at several prevalences.

prevalence	prevalence (%)	chance of disease after a positive result (%)
1 in 1,000	0.1	2.0
1 in 100	1	16.8
1 in 20	5	51.3
1 in 10	10	69.0
1 in 5	20	83.3
1 in 2	50	95.2

Work one row to see the pattern. At 1 in 100, count 1,000 people again.

```
0.01 times 1000 = 10
1000 minus 10 = 990
0.05 times 990 = 49.5
10 plus 49.5 = 59.5
10 divided by 59.5 = 0.168
```

About 16.8 per cent. The test has not changed. Only the group has.

The answer of 95 per cent is right only when about half the people tested have the disease. Read that way, the commonest answer quietly assumed a prevalence near 50 per cent. The question said 1 in 1,000.

WHERE THIS PICTURE BREAKS. Every row assumes the test catches everyone with the disease and wrongly flags 5 per cent of the rest. Those two numbers are carried unchanged from one group to the next. That is itself an assumption: a test can behave differently in a group with more advanced disease.

The table describes this made-up test, as the question did. It says nothing about any real test.

Illustration 3. Is this a quirk of a few hallways in 1978? Manrai and colleagues asked the same question in 2013. Their report is “Medicine’s uncomfortable relationship with math”, in *JAMA Internal Medicine*, 2014 (https://pmc.ncbi.nlm.nih.gov/articles/PMC4955674/).

They asked a convenience sample of 61 people at one Boston-area hospital: 24 attending physicians, 26 house officers, 10 medical students and one retired physician. A convenience sample is whoever was easy to reach, not a random draw.

study	answered about 2%	answered 95%	any other answer	asked
Casscells, 1978	11	27	22	60
Manrai, 2014 (surveyed 2013)	14	27	20	61

```
60 minus 11 minus 27 = 22
61 minus 14 minus 27 = 20
14 divided by 61 is about 0.230
11 divided by 60 is about 0.183
```

About 23 per cent right in 2013, against about 18 per cent in 1978. In both studies the commonest answer was 95 per cent, given by 27 people each time. Thirty-five years on, the error had barely moved.

Read what some of the 2013 respondents wrote beside their answers. Search the report for "prevalence".

> *one attending cardiologist wrote that "PPV does not depend on prevalence," and a resident wrote "better PPV when prevalence is low."*

PPV is the positive predictive value of Book 0, the share of positive results that are right. Both remarks are false, and section 5 shows why. These respondents knew the name of the quantity. They did not know what it depends on.

What would overturn the finding? A fresh sample of clinicians, given the same question, in which most give an answer near 2 per cent. Neither study found that.

Figures quoted in this illustration are from [46, 45].

WHERE THIS PICTURE BREAKS. Both studies are small, and neither drew its people at random: hallway encounters at four Harvard teaching hospitals in 1978, a convenience sample at one Boston-area hospital in 2013. Neither was done in India. The 18 and 23 per cent describe those 60 and 61 people. They are not the share of all doctors who can do this.

Both used one abstract question with no patient, no time pressure and no consequence. They show that the mistake is common. They do not show how often it changes what a doctor does with a real patient. Casscells and colleagues say that "is less clear".

MUST KNOW POINTS FOR YOU.

- The 95 per cent is a share of the healthy people. The question is about a share of the positives.
- You cannot say what a positive result means without a prevalence.
- Expect this mistake from most of the people around you, including senior ones. Do not take a confident number from a ward round as a check on your own working.
- Knowing the name of a quantity is not knowing what it depends on. When you teach this, make the trainee count a group of people, not recite a definition.
- The 2 per cent is for a person about whom nothing else is known. A patient with symptoms comes from a group where the disease is commoner, and the same positive result means more.
- These two studies were small, in two American teaching settings, on one abstract question. They do not measure how often it harms a patient, and neither was done in India.

COMMON MISREADING. "The test has a false-positive rate of only 5 per cent, so a positive result is 95 per cent likely to be right." The 5 per cent is a share of the people without the disease. The question is about the people with a positive result. At a prevalence of 1 in 1,000, about 50 well

people test positive for every one ill person, so a positive result means about a 2 per cent chance of disease. A rate that ignores the prevalence cannot answer a question that depends on it.

REPORTING IT. In a group where 1 in 1,000 people has the disease, a positive result on a test with a 5 per cent false-positive rate means about a 2 per cent chance of disease, not 95 per cent: about 50 healthy people test positive for every one person who has it.

EXERCISE 1 (retrieval). From memory: state the Casscells question in your own words. Give the right answer and the commonest wrong one. Say what the wrong answer actually is, written as a conditional probability.

EXERCISE 2 (teaching). A first-year resident, ten minutes, a whiteboard. She has just said in a ward round that a positive result on a test with a 5 per cent false-positive rate means "95 per cent he has it". Plan what you draw and say. Use the Casscells numbers.

EXERCISE 3 (design). You want to know whether the residents in your own department do better than the clinicians in these two studies. Specify a small survey: who you ask, the exact question, what counts as correct, and what your result could and could not show.

PRACTICE. Eight problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A condition has a prevalence of 1 in 500. A test for it catches everyone who has it and has a false-positive rate of 2 per cent. Count 10,000 people. How many test positive, and what is the chance that a person who tests positive has the condition?

2. A test catches everyone with a condition and has a false-positive rate of 5 per cent. At what prevalence is the chance of the condition after a positive result exactly 50 per cent?

3. Take the Casscells test: it catches everyone with the disease and has a false-positive rate of 5 per cent. It is now used in a clinic where 1 in 20 patients tested has the disease. Count 1,000 patients and give the chance of disease after a positive result.

4. Kiran and colleagues report that an arm circumference of 31.3 cm or more had sensitivity 86 per cent and specificity 74 per cent for a body-mass index from 25 to 29.9 in young men. Use it in a stated group where one man in five is in that band. A colleague asked "what does a positive tape mean?" answers "74 per cent". Count 1,000 men in that group, work out the right answer, and name the quantity the colleague gave.

5. Here is a worked answer to the Casscells question. Find the step that broke.

```
false-positive rate = 5 per cent
so the test is right 95 per cent of the time
the person tested positive, and the test is right 95 per cent of the time
so the chance the person has the disease = 95 per cent
```

6. Here is a second worked answer to the Casscells question. It does count people. Find the step that broke.

```
take 1,000 people: 1 has the disease, 999 do not
the false-positive rate is 5 per cent
so 5 per cent of the positive results are false
so 95 per cent of the positive results are true: the chance is 95 per cent
```

7. At a screening camp a colleague says: “Our test has a false-positive rate of only 5 per cent, so 95 per cent of the people we flag really have it. Refer them all straight to the specialist clinic.”

Decide what to compute, compute it, and say what your answer does not establish.

8. A newspaper story says: “Doctors can’t do maths. When asked what a positive test meant, only 18 per cent of doctors got it right.”

Decide what to compute or check, do it, and say what the 18 per cent does not establish.

7 · Natural frequencies: Bayes by counting people

DEFINITION. Natural frequencies are counts of people, got by following one group through each split in turn. Start with a group of a stated size. Split it by the condition, using the prevalence. Split each part by the test result, using the sensitivity among those with the condition and the false-positive rate among those without. The four counts at the end are the true positives, false negatives, false positives and true negatives.

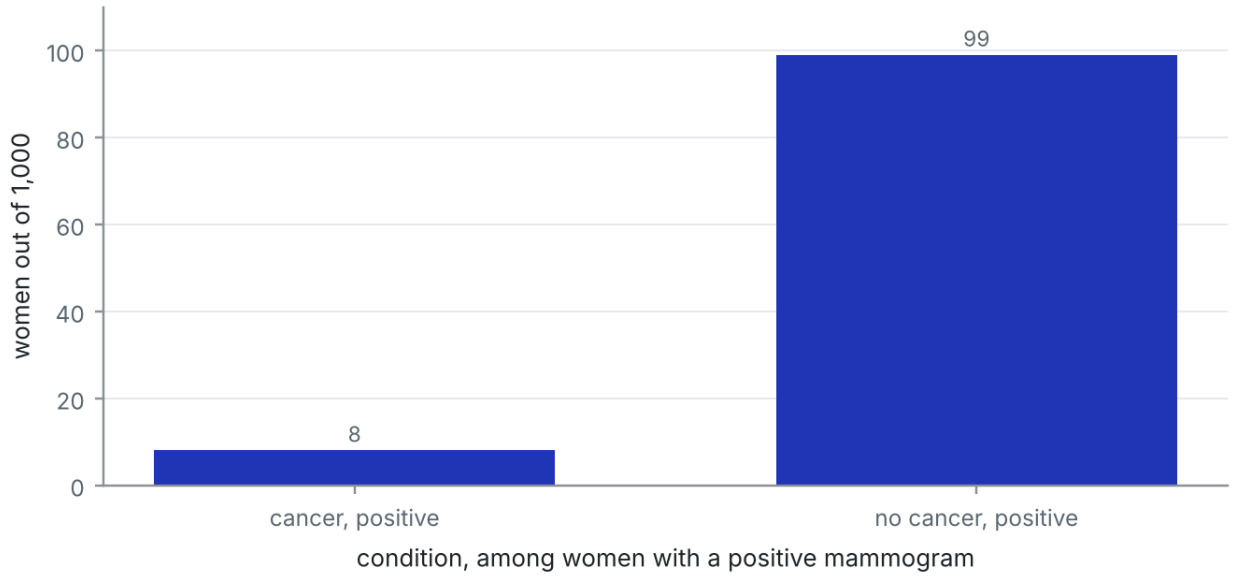
Laid out as a diagram that branches at each split, the counts form a frequency tree.

Write TP for the count of true positives, FP for false positives, TN for true negatives and FN for false negatives.

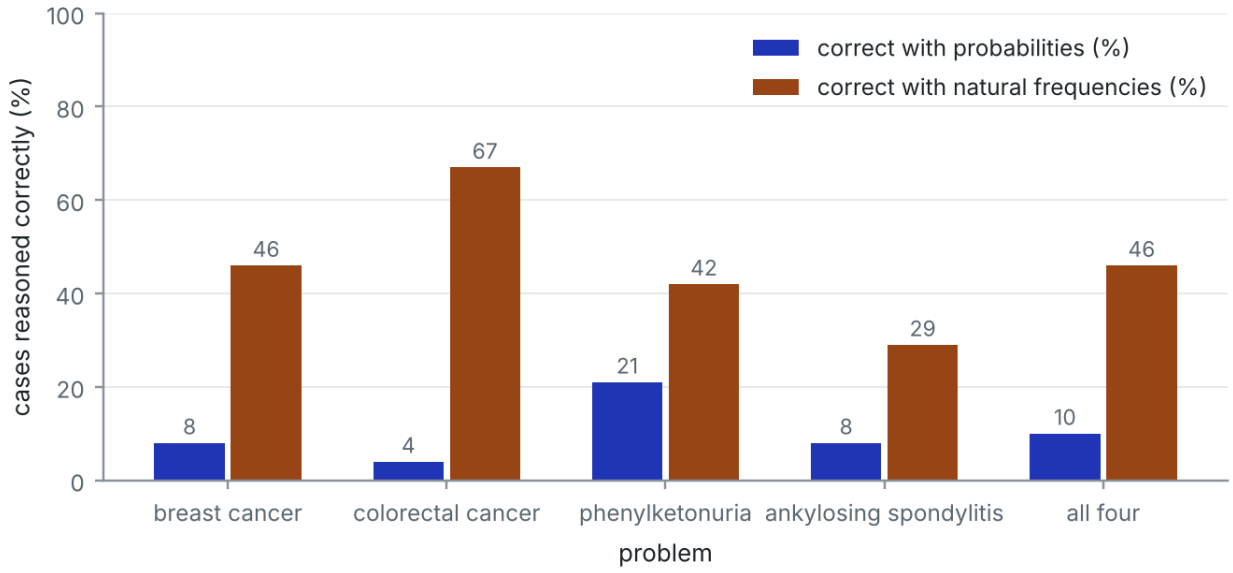
```
PPV = TP ÷ (TP + FP)
NPV = TN ÷ (TN + FN)
```

This is the same arithmetic as Bayes' theorem, with the multiplying already done. [47, 48, 48, 49, 50, 51]

IN PLAIN TERMS. Instead of three percentages, count people. Take 1,000 people. Then look at everyone who tested positive and ask how many of them have it.



Eddy's mammography problem as natural frequencies: the 107 of 1,000 women of 40 who have a positive mammogram. 8 have breast cancer and 99 do not. Of the 107 positives, 8 have cancer: about 7.5 per cent.



Hoffrage and Gigerenzer, 1998: per cent of cases in which 48 physicians reasoned correctly, with the same information given as probabilities (left bar of each pair) or as natural frequencies (right bar). Natural frequencies won in every problem. Breast cancer 8 against 46, colorectal cancer 4 against 67, phenylketonuria 21 against 42, ankylosing spondylitis 8 against 29; all four, 10 against 46.

Illustration 1. Try this first, the way Hoffrage and Gigerenzer gave it to physicians in "Using natural frequencies to improve diagnostic inferences", *Academic Medicine*, 1998 (https://doi.org/10.1097/00001888-199805000-00024). The numbers are from David Eddy's 1982 study, rounded, as Hoffrage and Gigerenzer report them.

> *For a woman at age 40 who participates in routine screening, the probability of breast cancer is 1%. If a woman has breast cancer, the probability is 80% that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 10% that she will still have a positive mammogram.*

A woman of 40 has a positive mammogram. What is the chance she has breast cancer? Write your answer down before you read on.

In Eddy's study, 95 of 100 physicians said between 70 and 80 per cent. Eddy concluded they had confused the sensitivity, 80 per cent, with the PPV. The correct answer is 7.5 per cent.

Now read the same information the way the 1998 paper rewrote it.

> *Ten out of every 1,000 women at age 40 who participate in routine screening have breast cancer. Of these ten women with breast cancer, eight will have a positive mammogram. Of the remaining 990 women without breast cancer, 99 will still have a positive mammogram.*

Every number is a count of women. Get the four counts at the ends of the frequency tree.

```
0.01 times 1000 = 10
1000 minus 10 = 990
0.80 times 10 = 8
10 minus 8 = 2
0.10 times 990 = 99
990 minus 99 = 891
```

	positive mammogram	negative mammogram	total
breast cancer	8	2	10
no breast cancer	99	891	990
total	107	893	1,000

Now ask the question of the positives only.

```
8 plus 99 = 107
8 divided by 107 is about 0.0748
```

107 women test positive, and 8 of them have cancer: about 7.5 per cent. The negative results are as quick.

```
2 plus 891 = 893
891 divided by 893 is about 0.998
```

The NPV is about 99.8 per cent. Compare the work with the formula of section 5. There you multiply the prevalence into each term to put the base rate back. Here the counts 8 and 99 already carry it. Whether 1,000 women or one probability, the answer is the same 7.5 per cent.

Does the rewrite help real doctors? Hoffrage and Gigerenzer gave 48 physicians in Munich and Düsseldorf four such problems. Each physician had two in probabilities and two in natural frequencies, with the formats and order varied. An answer counted as correct only if it was within five percentage points and the physician's notes and interview showed it was not a guess.

problem	correct with probabilities (%)	correct with natural frequencies (%)
breast cancer	8	46
colorectal cancer	4	67
phenylketonuria	21	42
ankylosing spondylitis	8	29
all four	10	46

Hoffrage and colleagues restated the colorectal row in *Science* in 2000: "Only 1 out of 24 physicians" was correct with probabilities, and 16 out of 24 others with natural frequencies.

With probabilities, the commonest wrong answers were the sensitivity itself, or the sensitivity minus the false-positive rate. Both ignore the prevalence. With natural frequencies, the commonest wrong answers ignored the test and gave a base rate instead.

Figures quoted in this illustration are from [52, 53, 54].

WHERE THIS PICTURE BREAKS. The 46 per cent is a share of cases, not of physicians, among 48 German physicians given written problems. More than half the answers were still wrong. Nothing here measures what those physicians then did with a patient.

The physicians in the study did not turn probabilities into counts on their own. The paper says so: they "did not spontaneously translate". The benefit came from being handed counts, or from being taught to make them.

The counts describe what would happen on average among 1,000 women. A real group of 1,000 would not give exactly 8 and 99: draw group after group, and the counts would scatter around them.

Illustration 2. Counting people has one trap: people come whole, and the arithmetic does not. Here it is in its plainest form. Take the Casscells question of section 6, at a prevalence of 1 in 1,000, and count only 100 people.

```
100 divided by 1000 = 0.1
```

One tenth of a person has the disease. Round it to zero and nobody has it, so a positive result means a 0 per cent chance. That is wrong. The fix is a bigger group: with 1,000 people the smallest count is 1.

The papers that made these trees met the same problem and rounded. Gigerenzer and Hoffrage's version of the mammography problem, in "How to improve Bayesian reasoning without instruction: frequency formats", *Psychological Review*, 1995 (https://doi.org/10.1037/0033-295X.102.4.684), used a false-positive rate of 9.6 per cent. Search it for "95 out of every 990".

```
0.096 times 990 = 95.04
```

They printed 95. The 2000 *Science* paper's colorectal tree used a false-positive rate of 3 per cent among 9,970 people.

```
0.03 times 9970 = 299.1
```

They printed 300. Of its 10,000 people, 30 have colorectal cancer and 15 of them test positive. Neither rounding moves the answer much.

```
15 divided by 315 is about 0.0476
15 divided by 314.1 = 0.0478
```

About 4.8 per cent either way.

Rounding can do more harm in the other direction: building counts back from rounded percentages. The arm-circumference table of Book 0, D3 took Kiran's printed sensitivity of 86 per cent and specificity of 74 per cent and rebuilt 36, 23, 6 and 66 men from them. Those are whole numbers, and they look like data. They are a reconstruction. Of the two group sizes Book 0 tried, only one rounds back to both percentages.

Book 0 also built a 1,000-man version at one in five. Read it now as a frequency tree, from the top.

```
0.20 times 1000 = 200
1000 minus 200 = 800
0.86 times 200 = 172
200 minus 172 = 28
0.74 times 800 = 592
800 minus 592 = 208
```

	arm 31.3 cm or more	arm under 31.3 cm	total
in the band	172	28	200
not in the band	208	592	800
total	380	620	1,000

The PPV is 172 of 380, the 45 per cent Book 0 reached. Now read off the NPV, which section 5 reached by formula.

```
28 plus 592 = 620
592 divided by 620 is about 0.955
```

Of 620 men the tape clears, 592 are outside the band: about 95.5 per cent.

Figures quoted in this illustration are from [53, 54, 10].

WHERE THIS PICTURE BREAKS. Choose a group large enough that the smallest count is at least a few whole people, round only at the last split, and say that you rounded. A count made from a percentage is an expected number, not a count anyone made.

The 1,000-man table uses Kiran's sensitivity and specificity, which come from 131 medical students in one college, and a prevalence of one in five that was stated for the example. Its 45 and 95.5 per cent hold for that stated group only.

MUST KNOW POINTS FOR YOU.

- When you need a PPV or NPV in your head or on a whiteboard, count 1,000 people (or 10,000).
- Split by the condition first, never by the test. Sensitivity applies only to the people with the condition and the false-positive rate only to the people without.
- When you explain a positive result to a patient, or write a handout, give counts: "of 1,000 people like you who test positive, about this many have it". In Hoffrage and Gigerenzer's study the same physicians got the PPV right in 10 per cent of cases from probabilities and 46 per cent from natural frequencies.
- Do not expect colleagues to translate on their own. Hand them the counts, or teach the four steps.
- Pick a group size that makes the smallest count at least a few whole people, round only at the end, and say that the counts are expected numbers.

- Counts rebuilt from rounded percentages look like data and are not. Never cite a rebuilt count as if a study measured it.
- Natural frequencies make the arithmetic easy. They do not make the inputs right.

COMMON MISREADING. "Natural frequencies give a different, more realistic answer than Bayes' theorem." They give the same answer. Eight out of 107 is the 7.5 per cent the formula gives. What changes is how easy the answer is to reach and to explain. A wrong prevalence or a misapplied sensitivity gives a wrong count just as surely as it gives a wrong probability.

REPORTING IT. Of every 1,000 people tested in this setting, about (the true positives plus the false positives) will have a positive result, and about (the true positives) of those will have the condition. A positive result therefore means about a (the true positives divided by all the positives, as a per cent) per cent chance of the condition, and is a reason for a confirmatory test, not a diagnosis.

EXERCISE 1 (retrieval). From memory: list the four steps that turn a prevalence, a sensitivity and a false-positive rate into a PPV by counting. Then say what result Hoffrage and Gigerenzer found when 48 physicians were given the same problems as probabilities and as natural frequencies.

EXERCISE 2 (teaching). A journalist, three minutes, no jargon. She is writing about a screening camp and asks: "If the test is 80 per cent accurate and I test positive, I've probably got it, right?" Use Eddy's mammography numbers as your example. Give her one sentence she can quote.

EXERCISE 3 (build). Draft the frequency-tree panel of your clinic handout. Choose a test used in your clinic, with a published sensitivity and specificity, and a prevalence for your patients. Write the 1,000-person counts as a table, the two sentences a patient will read, and one line saying where each input came from.

PRACTICE. Nine problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A condition has a prevalence of 10 per cent. A test has a sensitivity of 90 per cent and a false-positive rate of 20 per cent. Count 1,000 people and give the four counts at the end of the frequency tree.

2. From the counts in problem 1 (true positives 90, false negatives 10, false positives 180, true negatives 720), work out the PPV and the NPV.

3. Of 2,000 people, 40 have a condition. 30 of those 40 test positive. 98 of the people without it also test positive. Work out the prevalence, sensitivity, specificity, PPV and NPV.

4. Gigerenzer and Hoffrage's 1995 mammography problem gives a prevalence of 1 per cent, a sensitivity of 80 per cent and a false-positive rate of 9.6 per cent. Turn it into natural frequencies on 10,000 women and give the chance that a woman who tests positive has breast cancer.

5. An arm circumference of 31.3 cm or more had sensitivity 86 per cent and specificity 74 per cent for the body-mass index band 25 to 29.9 in Kiran's young men. A camp screens men among whom you expect 1 in 10 to be in the band. Build the 1,000-man frequency tree and give the PPV and the NPV, each with what it is a share of.

6. Here is a worked answer. Find the step that broke.

A condition has a prevalence of 1 per cent. A test has a sensitivity of 90 per cent and a false-positive rate of 9 per cent.

```
take 1,000 people
true positives: 0.90 times 1000 = 900
false positives: 0.09 times 1000 = 90
PPV: 900 divided by 990 = 0.91
```

7. Here is a worked answer to the Casscells question: prevalence 1 in 1,000, a test that catches everyone with the disease, false-positive rate 5 per cent. Find the step that broke.

```
take 100 people
with the disease: 100 divided by 1000 = 0.1, rounded to 0
false positives: 0.05 times 100 = 5
true positives: 0
chance of disease after a positive result: 0 divided by 5 = 0
```

8. A screening leaflet, made up for this problem, says: “This test picks up 90 per cent of people with the condition and gives a false alarm in only 9 per cent of people without it. So if your result is positive, you very likely have the condition.” The condition affects 1 in 100 of the people screened.

Decide what to compute, compute it, rewrite the last sentence as counts, and say what your rewrite does not establish.

9. A conference slide says: “Natural frequencies made doctors almost five times better at Bayes’ theorem.”

Decide what to compute or check, do it, and say what the claim does not establish.

8 · The odds form of Bayes' theorem and the likelihood ratio

DEFINITION. The odds of an event A are its probability divided by the probability that it does not happen, as in Book 0, D1: odds = P(A) ÷ (1 − P(A)). Going back, P(A) = odds ÷ (1 + odds).

Let A be a hypothesis, such as "has the condition", and B the evidence, such as a test result. Write Bayes' theorem of section 5 twice: once for A, once for "not A".

```
P(A | B) = P(B | A) × P(A) ÷ P(B)
P(not A | B) = P(B | not A) × P(not A) ÷ P(B)
```

Divide the first line by the second. P(B) is on the bottom of both, so it cancels.

```
P(A | B) ÷ P(not A | B) = [P(B | A) ÷ P(B | not A)] × [P(A) ÷ P(not A)]
```

The left side is the posterior odds of A, also called the post-test odds: its odds once you know B. The last bracket is the prior odds, also called the pre-test odds: the odds before you knew B. The middle bracket is the likelihood ratio (LR) of the evidence B.

```
LR = P(B | A) ÷ P(B | not A)
posterior odds = prior odds × LR
```

The positive likelihood ratio, LR+, uses "tests positive" as B. The negative likelihood ratio, LR−, uses "tests negative".

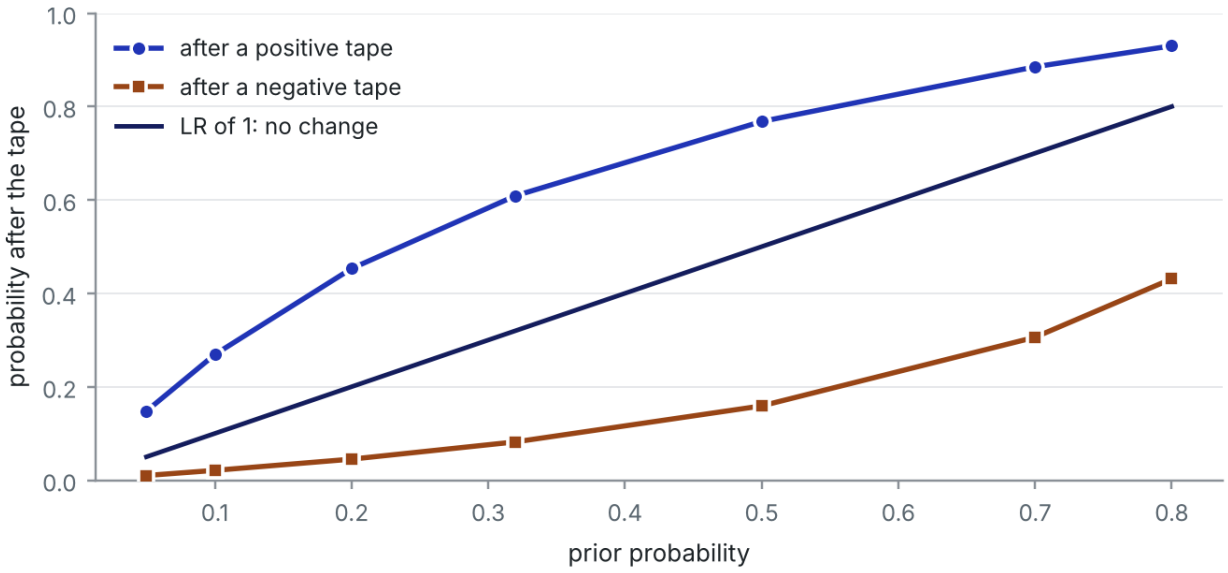
```
LR+ = sensitivity ÷ (1 - specificity)
LR- = (1 - sensitivity) ÷ specificity
```

An LR above 1 raises the odds, an LR below 1 lowers them, and an LR of 1 leaves them where they were.

Two results B and C can be applied one after the other, each multiplying the odds by its own LR. This is allowed only if B and C are conditionally independent given A and given not A (section 3). [55, 56, 57, 58, 59, 60, 61]

IN PLAIN TERMS. Odds say how many times more likely something is than not. A test result multiplies those odds by a fixed number, its likelihood ratio.

Turn the probability you start with into odds. Multiply by the likelihood ratio. Then turn the odds back into a probability.



What the arm tape does to the probability that a man is in the 25 to 29.9 band, from priors between 0.05 and 0.8. A positive tape multiplies the odds by about 3.31, a negative one by about 0.19. From 0.2, a positive tape gives 0.453 and a negative one 0.045. The solid diagonal line is a result with an LR of 1, which changes nothing.

Illustration 1. Take the arm tape of Book 0, Book 0, D3, again. In Kiran and colleagues' study of medical students in Mangaluru, a mid-upper arm circumference of 31.3 cm or more had a sensitivity of 86 per cent and a specificity of 74 per cent for men in the body-mass index band 25 to 29.9. Work out the two likelihood ratios.

```
1 minus 0.74 = 0.26
0.86 divided by 0.26 ≈ 3.3077
1 minus 0.86 = 0.14
0.14 divided by 0.74 ≈ 0.1892
```

So LR+ is about 3.31: a positive tape is about 3.3 times as common among men in the band as among men outside it. LR− is about 0.19.

Now use them. Start with the 131 men of the study, 42 of them in the band. The prior probability is 42 divided by 131. The prior odds are the men in the band against the men outside it, 42 against 89.

```
131 minus 42 = 89
42 divided by 89 ≈ 0.4719
0.4719 times 3.3077 = 1.5609
1.5609 divided by 2.5609 = 0.6095
```

The last line turns odds back into a probability: odds divided by (1 plus the odds). After a positive tape the probability is about 0.61. That is Book 0's 61 per cent, reached without the table. Book 0's 36 of 59 gives 0.6102, a shade higher, because its counts were rounded to whole men.

Move the tape to a group of men in which one in five is in the band, as Book 0 did. The prior odds are 1 against 4.

```
0.2 divided by 0.8 = 0.25
0.25 times 3.3077 = 0.8269
0.8269 divided by 1.8269 = 0.4526
0.25 times 0.1892 = 0.0473
0.0473 divided by 1.0473 = 0.0452
```

A positive result raises the probability from 0.2 to about 0.45, Book 0's 45 per cent. A negative result lowers it from 0.2 to about 0.045. The likelihood ratios did not change between the two groups. Only the prior did.

The table below runs the same two multiplications from a range of priors. The figure draws it.

prior probability	after a positive tape	after a negative tape
0.05	0.148	0.010
0.1	0.269	0.021
0.2	0.453	0.045
0.32	0.609	0.082
0.5	0.768	0.159
0.7	0.885	0.306
0.8	0.930	0.431

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The likelihood ratios stay fixed only while the sensitivity and specificity do. They were measured on 131 young men in one medical college. On older men, or men measured by someone else, the tape may behave differently, and then so do its likelihood ratios.

The 86 and 74 are rounded and, as Book 0 found, fit the men in the 25 to 29.9 band, not everyone at 25 or more. The likelihood ratios inherit both limits.

Illustration 2. A likelihood ratio does not need a test with only two results. Deeks and Altman show this in their note "Diagnostic tests 4: likelihood ratios" in the *BMJ* (https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/). They took a study of smoking history in people tested for obstructive airway disease, and gave each band of pack-years its own likelihood ratio. A pack-year is a pack of cigarettes a day for a year.

Start with the band of 40 or more pack-years. Its LR is 20.4 in their table. Their worked box carries more digits, 20.43, and takes a prior probability of 0.1. Do it yourself.

```
0.1 divided by 0.9 = 0.1111
0.1111 times 20.43 = 2.27
2.27 divided by 3.27 = 0.6942
```

The probability rises from 0.1 to about 0.69, as the note says. Now look at their box's last line as printed: "2.27/3.37 = 0.69". That division gives 0.674, not 0.69. The 3.37 is a slip for 3.27, the 1 plus

the odds. Their answer is right and one figure on the way is not. Check the arithmetic in a published worked example, as you would your own.

Now the band of 20 to 40 pack-years. Its LR is 1.01.

```
0.1111 times 1.01 = 0.1122
0.1122 divided by 1.1122 = 0.1009
```

The probability goes from 0.1 to 0.1009. Asking this question moved nothing. A likelihood ratio near 1 is a result that turns up about as often with the disease as without it, and it carries almost no information.

Figures quoted in this illustration are from [62].

WHERE THIS PICTURE BREAKS. Deeks and Altman count an LR above 10 or below 0.1 as strong evidence "in most circumstances". That is a convention, not a law. An LR of 20 still leaves a probability of only about 0.17 if the prior was 0.01: the prior odds are about 0.0101, times 20 is 0.202, and 0.202 divided by 1.202 is about 0.168. Always finish the multiplication before you call a result decisive.

Illustration 3. Here is the trap with two tests in a row. A man's arm tape is positive in a camp where one man in five is in the band. A colleague, unsure, measures the same arm again with the same tape, gets a positive again, and multiplies by LR+ a second time.

```
0.25 times 3.3077 times 3.3077 = 2.7352
2.7352 divided by 3.7352 = 0.7323
```

About 0.73. It is wrong. The second reading is the same arm, the same tape and the same hands. Among men in the band, a man whose first reading was positive is very likely to read positive again. The same holds among men outside it. The two readings are not conditionally independent (section 3). The second one adds almost nothing, and the probability stays near 0.45.

The rule needs a second test whose errors have nothing to do with the first one's, inside each group. Suppose, as a made-up case, a second test exists whose results are independent of the tape's among men in the band and among men outside it, with an LR+ of 4. Then the multiplication is allowed.

```
0.25 times 3.3077 times 4 = 3.3077
3.3077 divided by 4.3077 = 0.7679
```

Two positives now take the probability from 0.2 to about 0.77. Two measures of the same thing, such as an arm tape and a neck tape, both read body size. Do not assume they are independent inside each group. Ask for a study that reports both results on the same people.

WHERE THIS PICTURE BREAKS. The second test with an LR+ of 4 is made up, and its independence is assumed, not measured. Real paired tests are rarely fully independent inside each group.

Then the true combined likelihood ratio lies below the product, and only a study that applied both tests to the same people can give it.

MUST KNOW POINTS FOR YOU.

- "An LR of 3 triples the chance" is the error. A likelihood ratio multiplies the odds, not the probability. Turn the probability into odds before you multiply, and back after.
- When you read a test's accuracy in a paper, write down LR+ and LR− from its sensitivity and specificity.
- A result whose LR is near 1 changes almost nothing, whatever its sensitivity sounds like.
- Multiply the likelihood ratios of two tests only when their results are independent among people with the condition and among people without it. Two measures of the same thing, or the same measure repeated, fail this.
- An LR of 10 on a prior of 0.01 still leaves the probability under 0.1. Say the posterior probability, not just the size of the LR.
- A likelihood ratio is built from a sensitivity and a specificity measured on one group of people. Say which study the LR came from.

COMMON MISREADING. "The tape's LR+ is 3.3, so a positive tape makes a man 3.3 times as likely to be in the band." The LR multiplies the odds. From a probability of 0.2 the positive tape gives 0.45, which is about 2.3 times the prior, not 3.3 times. From 0.5 it gives 0.77, about 1.5 times. The higher the prior, the less a fixed LR can add to the probability, because the probability cannot pass 1.

REPORTING IT. At a cut-off of 31.3 cm, the arm tape's positive likelihood ratio is about 3.3. Its negative likelihood ratio is about 0.19. Both come from the sensitivity of 86% and specificity of 74% found in 131 male medical students. Where one man in five is in the band, a positive result raises the probability to about 0.45, and a negative result lowers it to about 0.05.

EXERCISE 1 (retrieval). From memory: write the odds form of Bayes' theorem, and the formulas for LR+ and LR− in terms of sensitivity and specificity. Then say when you may multiply two likelihood ratios together.

EXERCISE 2 (calculation). In a clinic you estimate that one adult man in ten has a body-mass index in the band 25 to 29.9. The arm tape's sensitivity is 86 per cent and its specificity 74 per cent. Find the probability that a man is in the band after a positive tape, and after a negative tape, by the odds form. Say before you start how sure you are that the positive result lifts the probability above one half.

Record your confidence as a percentage before turning to the answer.

EXERCISE 3 (teaching). A journalist has read that a new screening question "has a likelihood ratio of 2, so a positive answer doubles your risk". In three minutes and without jargon, explain what is right and wrong in that sentence, and give one sentence they can quote.

PRACTICE. Eleven problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Turn a probability of 0.2 into odds. Turn odds of 1.5 into a probability.

2. A test has a sensitivity of 0.9 and a specificity of 0.8. Find LR+ and LR−.

3. The prior probability is 0.1 and the result has an LR of 6. Find the posterior probability.

4. The prior probability is 0.05. What likelihood ratio would a result need to bring the posterior probability to 0.5?

5. Mohan and colleagues built the Indian Diabetes Risk Score in Chennai. At a score of 60 or more, it had a sensitivity of 72.5 per cent and a specificity of 60.1 per cent for undiagnosed diabetes. The paper's discussion assumes a prevalence of undiagnosed diabetes of 10.1 per cent. Find LR+ and LR−, and the probability of undiagnosed diabetes after a score of 60 or more and after a lower score, at that prevalence.

6. In Deeks and Altman's smoking example, people with fewer than 20 pack-years had an LR of 0.55 for obstructive airway disease, and people who had never smoked had an LR of 0.76. Starting from a prior probability of 0.1, find the posterior probability for each. Do either rule the disease out?

7. Altman and Bland's liver scan had a sensitivity of 0.895 and a specificity of 0.628, in a sample where the prevalence of abnormal pathology was 0.75. Find LR+, the pre-test odds, the post-test odds after a positive scan and the post-test probability. Then compare the probability with what the scan's LR+ suggests about how useful it is.

8. Here is a worked answer. Find the step that broke.

```
arm tape: sensitivity 0.86, specificity 0.74, so LR+ = 0.86 divided by 0.26 = 3.31
one man in five is in the band, so the prior is 0.2
after a positive tape: 0.2 times 3.31 = 0.66
```

9. Here is a worked answer. Find the step that broke.

```
arm tape: sensitivity 0.86, specificity 0.74
LR- = specificity divided by (1 - sensitivity) = 0.74 divided by 0.14 = 5.29
a negative tape multiplies the odds by 5.29
so a negative tape makes being in the band more likely
```

10. A health app says of a made-up screening question: "Our question has a likelihood ratio of 2. Answer yes, and your chance of having the condition doubles." Decide what to compute, compute it for a person whose chance before the question was 0.01, for one at 0.5 and for one at 0.6, and say what the claim does not establish.

11. A colleague in a screening camp says: “His arm tape was positive, and the neck tape was positive too. The arm tape’s LR+ is 3.3 and the neck tape’s is about 4.5. Multiply: nearly 15. With one man in five in the band, he’s almost certainly in it.”

Decide what to compute, compute it, and say what the answer does not establish. Treat the 4.5 as the colleague’s figure, not one you have checked.

9 · The prosecutor's fallacy: P(evidence | innocent) is not P(innocent | evidence)

DEFINITION. In this section only, let H be a hypothesis and E the evidence. P(E | H) is the probability of the evidence if the hypothesis is true. P(H | E) is the probability of the hypothesis once the evidence is known. The transposed conditional is the error of reading one as the other: the two conditions are swapped, so the bottom of the fraction changes (section 2).

In a trial, let H be "the accused is innocent" and E the evidence against them. Reading a small P(E | innocent) as a small P(innocent | E) is called the prosecutor's fallacy. The Royal Statistical Society used that name for it in 2001, of the Sally Clark case, a case told below.

The same error takes other forms. In a clinic it reads the false-positive rate, P(positive | no disease), as P(no disease | positive), the case of section 6. In research it reads a p-value, which has the shape P(data like these | no effect), as P(no effect | data).

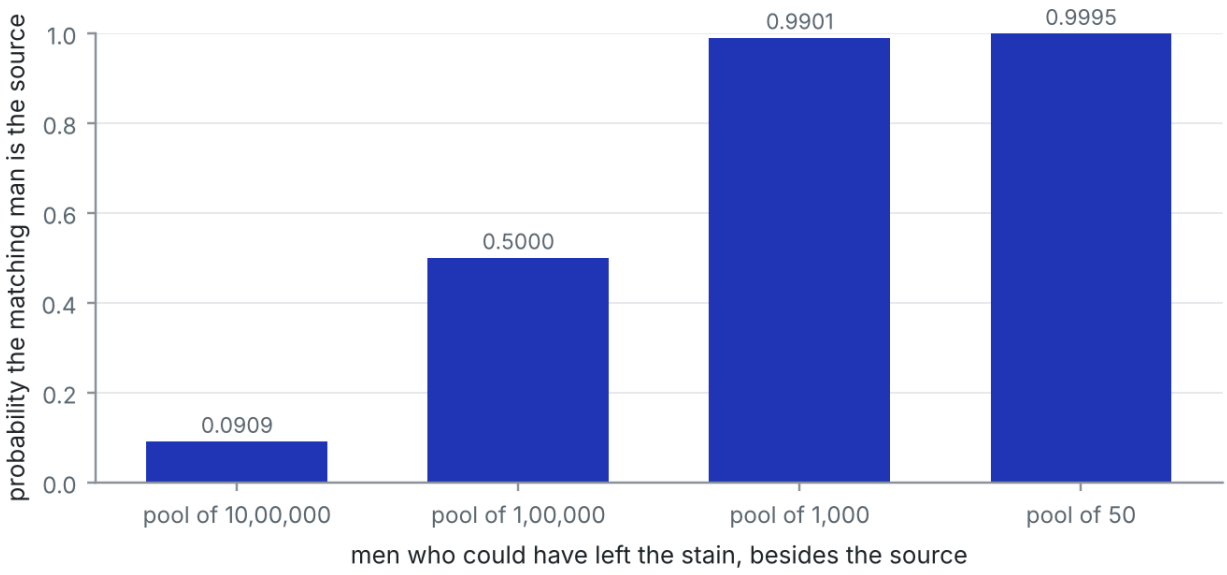
The mirror error is made by the defence. It counts how many other people the evidence would also fit, concludes the evidence is nearly worthless, and forgets that the other evidence in the case may already have shut most of those people out. It ignores the prior in the other direction.

Four steps correct the transposition on the spot.

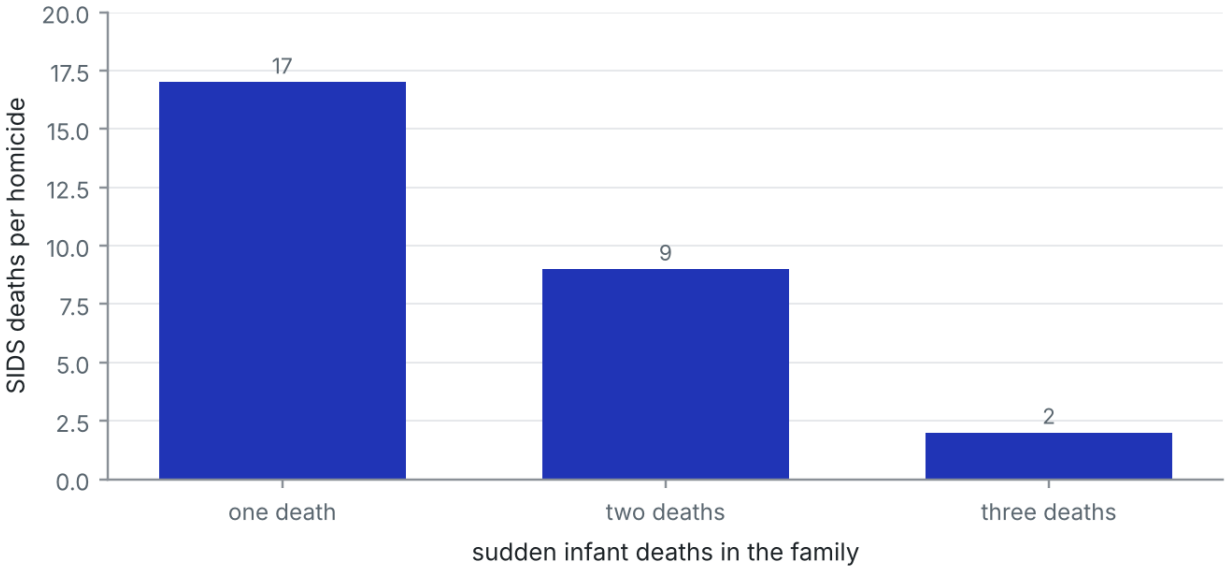
1. Say both conditionals in words: "the chance of this evidence if innocent" and "the chance of innocence given this evidence".
2. Ask of each: "out of whom?"
3. Count a population, 1,000 or 10 lakh people, through both splits (section 7).
4. Say which number answers the question being asked. [63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86]

IN PLAIN TERMS. "If he were innocent, this evidence would be very unlikely" sounds the same as "given this evidence, he is very unlikely to be innocent". It is not the same. The first counts among innocent people. The second counts among everyone the evidence fits.

The fix is to count. Take a large group of people, follow them through the evidence, and look at who is left.



The same random match probability, 1 in 1,00,000, gives very different answers depending on how many men besides the source could have left the stain. With 10,00,000 the matching man is the source with probability 0.0909; with 1,00,000, 0.5000; with 1,000, 0.9901; with 50, 0.9995. The pools are made up.



Hill's estimates of how much more common sudden infant death syndrome (SIDS) was than homicide among families with one, two and three sudden infant deaths: about 17 to 1, 9 to 1 and 2 to 1. Hill put the true two-death ratio between 4.5 to 1 and 9 to 1. Suspicion rises with each death, but SIDS stays the likelier explanation. The figures are rough estimates from small counts, for England in the 1990s.

Illustration 1. Start with a DNA match, in the words an Indian court has used. In *Manoj v State of Madhya Pradesh* (2022) the Supreme Court quoted the Law Commission of India's 185th Report. A match "does not mean the identity is conclusively proved". The expert, the Report says, gives a number for how often a profile like this turns up, and gives an example: "the relevant profile is found in 1 person in every 100,000" (1,00,000 in Indian grouping). The Supreme Court's quoted laboratory article calls this kind of figure the Random Match Probability (RMP): "the frequency of the particular DNA profile in a population".

An RMP of 1 in 1,00,000 is P(match | not the source). Here is the transposition. “The chance of a match if he is not the source is 1 in 1,00,000” becomes “the chance he is not the source is 1 in 1,00,000”. Count to see what the second sentence needs.

Suppose, as a made-up case, that 10 lakh men in a district besides the source could have left the stain, and nothing else in the case points to any of them. How many would match by chance?

```
10,00,000 divided by 1,00,000 = 10
```

About 10 innocent men match, and one man who is the source. The man in the dock is one of about 11 who match. If nothing else tells them apart, the chance that he is the source is about 1 in 11, not 99,999 in 1,00,000.

```
1 divided by 11 ≈ 0.0909
```

Stephen Watkins made the same count for the United Kingdom in the *BMJ* in 2000 (https://pmc.ncbi.nlm.nih.gov/articles/PMC1117305/). A chance match was “quoted as 20 million to one”. The accurate statement, he wrote, was “that the defendant or two other unknown people in the United Kingdom could have committed the offence”.

Now the mirror error. The defence says: “10 men in the district match, so the match is worth no more than a 1 in 11 chance.” That is right only if nothing else narrows the field. Suppose the other evidence, such as who was in the village that night, leaves 50 men who could have left the stain.

```
50 divided by 1,00,000 = 0.0005
1 divided by 1.0005 ≈ 0.9995
```

Now an innocent chance match among the 50 is very unlikely, and the probability that the matching man is the source is about 0.9995. The table runs the same count over four sizes of pool. The figure draws it.

men who could have left the stain, besides the source	innocent men expected to match	probability the matching man is the source
10,00,000	10	0.0909
1,00,000	1	0.5000
1,000	0.01	0.9901
50	0.0005	0.9995

This is the odds form of section 8. If the source always matches, the likelihood ratio of a match is 1 divided by 1 in 1,00,000, which is 1,00,000. The prior odds come from the rest of the case: 1 to 10 lakh in a whole district, 1 to 50 in the village. The RMP fixes the likelihood ratio. It never fixes the answer on its own.

Figures quoted in this illustration are from [87, 88].

WHERE THIS PICTURE BREAKS. The district of 10 lakh and the village of 50 are made up, and so is the assumption that every man in the pool was equally likely to be the source before the DNA.

Real pools are not even: relatives share more of their DNA, and a stain can come from a source unconnected with the crime.

The count also takes the laboratory's report as correct. A sample mixed up or contaminated gives a match that no RMP describes. *Manoj* gave the report little weight partly for that reason.

Illustration 2. The same error is common in a clinic. In the Casscells question of section 6, a test has a false-positive rate of 5 per cent, for a disease with a prevalence of 1/1000. A doctor says: "The chance that this positive is a false positive is 5 per cent." That sentence has transposed the conditional. The 5 per cent is P(positive | no disease). The sentence states it as P(no disease | positive).

Count 1,00,000 people. Assume, as the usual answer to the question does, that the test catches everyone who has the disease.

```
1,00,000 divided by 1,000 = 100
1,00,000 minus 100 = 99,900
99,900 times 0.05 = 4,995
100 plus 4,995 = 5,095
4,995 divided by 5,095 ≈ 0.98
```

Of 5,095 positive results, 4,995 are false. The chance that a positive is a false positive is about 98 per cent, not 5 per cent.

Research has its own version. A study reports p = 0.03, and someone writes "there is only a 3 per cent chance that the drug does not work". A p-value is computed by assuming that the drug does not work, so it has the shape P(data like these | no effect). The sentence turns it into P(no effect | data). The next book of this subject defines the p-value and works this through. For now, notice that the sentence has the same shape as the two above.

Figures quoted in this illustration are from [45].

WHERE THIS PICTURE BREAKS. The 98 per cent assumes a test that misses nobody. With a test that misses some cases, the true positives are fewer than 100, and the share of false positives is higher still.

The p-value paragraph only names the shape of the error. How far apart the two quantities are in research depends on things this book does not cover.

Illustration 3. Here is the case where the error went to a jury. In 1999 Sally Clark, a solicitor in Cheshire, England, was convicted of murdering her two baby sons. Both had died suddenly. The defence said both deaths were sudden infant death syndrome (SIDS): "the sudden death of a baby that is unexpected by history and in whom a thorough post-mortem examination fails to demonstrate an adequate cause of death", in the definition Ray Hill quotes in "Multiple sudden infant deaths – coincidence or beyond coincidence?" (*Paediatric and Perinatal Epidemiology*, 2004, https://doi.org/10.1111/j.1365-3016.2004.00560.x).

A paediatrician, Professor Meadow, gave the jury a figure of 1 in 8,543 for a SIDS death in a family like the Clarks'. Then, in the words the Court of Appeal later set out (*R v Clark* [2003] EWCA Crim 1020, para 96):

> *Yes, you have to multiply 1 in 8,543 times 1 in 8,543 and I think it gives that in the penultimate paragraph. It points out that it's approximately a chance of 1 in 73 million.*

Check the arithmetic first.

```
8,543 times 8,543 = 7,29,82,849
```

About 73 million. The arithmetic is right. Three steps were wrong.

First, squaring assumed independence (section 3). Multiplying two probabilities is only as good as the claim that knowing one tells you nothing about the other. The Royal Statistical Society said the squaring "would only be valid if SIDS cases arose independently within families". Families share genes and homes. From the published studies, Hill estimated that the risk of SIDS "is between 5 and 10 times greater" for a baby whose brother or sister died of SIDS. Undo the squaring with a factor of 5 and with 10.

```
7,29,82,849 divided by 5 ≈ 1,45,96,570
7,29,82,849 divided by 10 ≈ 72,98,285
```

About 1 in 15 million, or 1 in 7 million: five to ten times less rare, from this step alone. Hill also objected that 1 in 8,543 picked out low-risk features of the family, and that such a family is "atypical for murder" too.

Second, the transposition. Some press reports took 1 in 73 million as the chance that the two deaths were accidental. The Royal Statistical Society called this the Prosecutor's Fallacy. 1 in 73 million, even if it were right, is P(two deaths | SIDS). The jury's question was P(SIDS | two deaths).

Third, the comparison that was missing. Two deaths by murder are also very rare. What matters, the Society wrote, is "the relative likelihood of the deaths under each explanation". Hill made that comparison from the Care of Next Infant study. Among families with two sudden infant deaths, he found 18 with both deaths SIDS against 2 with both deaths homicide.

```
18 divided by 2 = 9
18 divided by 20 = 0.9
```

Among families with two such deaths, double SIDS outnumbered double homicide by about 9 to 1, in Hill's estimate from small counts. Hill thought this might be too high, and put the true ratio between 4.5 to 1 and 9 to 1. For one death his estimate was about 17 to 1, and for three about 2 to 1. The figure draws the three. More deaths do raise suspicion, but nowhere near what 1 in 73 million suggested.

The convictions were quashed in 2003. The main ground was medical evidence that had never been disclosed to the defence. But the Court of Appeal wrote of the 1 in 73 million that "it grossly overstates the chance of two sudden deaths within the same family from unexplained but natural causes". Had there been a challenge, "the wisest course would have been to exclude it altogether".

Figures quoted in this illustration are from [89, 88, 90].

WHERE THIS PICTURE BREAKS. Hill's 9 to 1 rests on 18 and 2 families, and his 17 to 1 and 2 to 1 on his own estimates; he calls such figures rough. They compare double SIDS with double homicide only, leaving out accidents, explained natural deaths and mixtures. They describe families in England in the 1990s, not any family today or in India.

None of these numbers says whether any one mother is guilty. The point is narrower: a probability of the evidence under one explanation, standing alone, cannot say that.

Illustration 4. Now the Indian setting. Since 1 July 2024 the law of evidence in Indian courts is the Bharatiya Sakshya Adhiniyam, 2023 (BSA), the Act that repealed the Indian Evidence Act, 1872 (s.170). Read three of its sections in the Gazette copy (https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf).

Section 39(1) makes the opinions of persons "specially skilled" in a science relevant facts, and calls such persons experts. A DNA analyst giving a match and its RMP is one. Section 45 makes the grounds of an opinion relevant too. Section 2(1)(j) says a fact is proved when, "after considering the matters before it, the Court either believes it to exist, or considers its existence so probable that a prudent man ought, under the circumstances of the particular case, to act upon the supposition that it exists".

Read that definition as this book would. It asks how probable the fact is after "the matters before it", the whole of the evidence. That is a P(fact | evidence). An RMP is a P(evidence | the fact is false). The Act asks the court the first question. The expert's number answers the second. The transposition is the step that turns one into the other without the prior.

The Supreme Court's own words point the same way. In *Pattu Rajan v State of Tamil Nadu* (2019), it held that "opinion evidence is advisory in nature, and the Court is not bound by the evidence of the experts". The expert must give reasons "so that the Court may form its independent judgment". In *Manoj* (2022), it faulted the expert for not choosing "to elaborate on the random occurrence ratio". It held that the DNA report there "cannot have a clinching, or high degree of probative value", for reasons that included doubts about how the samples were recovered and kept. Both judgments were decided under the 1872 Act.

No Indian judgment or official report held for this book names the prosecutor's fallacy, and this book does not say that any Indian court has made it. What you can do is spot its shape. When a report, a chargesheet or a news story says "there is a 1 in 1,00,000 chance that the accused is not the source", run the four steps. Out of whom is the 1 in 1,00,000? Out of people who are not the source. That sentence needs a count of who else could have left the stain before it can be said at all.

Figures quoted in this illustration are from [91, 92, 93].

WHERE THIS PICTURE BREAKS. This book reads "so probable that a prudent man ought … to act" as a probability given the evidence. That is a reading of the words for teaching, not a holding of any court, and the Act sets no number. How Indian courts weigh a DNA match in a given case is for a lawyer.

The BSA sections quoted are as enacted in 2023. Check the current text before you rely on one in a report or in court.

MUST KNOW POINTS FOR YOU.

- "The chance of this evidence if he were innocent is 1 in a million, so the chance he is innocent is 1 in a million" is the error. Before you repeat any small probability, say aloud whose probability it is and "out of whom".
- In clinic, never say "the chance this is a false positive is 5 per cent" from a false-positive rate of 5 per cent. That rate is out of people without the disease. Out of people who test positive, the share of false positives can be 98 per cent when the disease is rare. Give the patient the second number.
- To correct a colleague on the spot, use four steps. Keep it to three sentences when you say it aloud.
- A very small probability built by multiplying should make you ask what ties the events together, as in section 3. Refuse to repeat a product until someone has shown the parts are independent.
- Weigh the evidence under both explanations, never under one. The question is which is less rare, and by how much: the likelihood ratio, with a prior.
- The defence's mirror error is as wrong as the prosecutor's. Count the pool the rest of the evidence leaves, not the whole country.
- If you give expert evidence, give your number with what it is conditional on, and do not state the probability of guilt or innocence: that question is the court's, and your number does not answer it.
- Counting a population exposes the transposition. It does not supply the prior. Say the pool you assumed.

COMMON MISREADING. "A random match probability of 1 in 1,00,000 means a 99,999 in 1,00,000 chance that the accused is the source." The RMP is out of people who are not the source. How likely the accused is to be the source depends on how many such people could have left the stain, which the RMP does not tell you. In a district of 10 lakh with nothing else to go on, about 10 innocent men would match, and the chance is about 1 in 11.

REPORTING IT. The DNA profile found at the scene matched the accused. The laboratory reports that such a profile is found in about 1 in 1,00,000 unrelated persons. This is the probability of a match if the accused is not the source. It is not the probability that the accused is not the source, which depends on how many other persons could have left the sample.

EXERCISE 1 (retrieval). From memory: what is the prosecutor's fallacy, written with the bar of conditional probability? Name the four steps that correct it, and say what the defence's mirror error leaves out.

EXERCISE 2 (critique). Here is a made-up sentence from a case summary. "The DNA expert deposed that the profile occurs in 1 in 10 lakh persons. Therefore there is only a one in ten lakh chance that the accused is not the person who left the stain."

Find the specific flaw, and rewrite the sentence so that it says only what the expert's number supports.

EXERCISE 3 (teaching). A journalist asks you, three minutes before a deadline, why the "1 in 73 million" in the Sally Clark case was wrong if the multiplication was right. Answer without jargon, and give one sentence they can quote.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. In a group of 1,000 people, 900 have H. Among the 900, the event E happens to 9. Among the 100 without H, it happens to 90. Find P(E | H) and P(H | E).

2. In 10,000 people, P(B) is 0.02, P(A | B) is 0.95 and P(A | not B) is 0.10. A colleague says that P(B | A) must therefore be 0.95. Count the people and find P(B | A).

3. One event has a probability of 1 in 1,000. A second event of the same kind is 8 times as likely after the first has happened. Find the probability of both, and compare it with the probability you get by squaring 1 in 1,000.

4. Watkins quotes 1 in 2.75 million, the population risk of a cot death squared: the chance that two babies in one family both die of it, if the deaths were independent. He also gives 3,78,000 second or later births a year in England. Find how often such a pair of deaths would turn up somewhere in England. Then do it again if a second death is 5 times as likely after a first.

5. A DNA profile is found, in the Law Commission's example quoted in *Manoj*, in 1 person in every 1,00,000. In a made-up case, 5 lakh men besides the source could have left the stain, and nothing else distinguishes them. Find the expected number of innocent matches, and the probability that a matching man is the source. Then do the same if other evidence leaves 50 men besides the source.

6. Hill counts, in the Care of Next Infant study, 18 families in which both of two sudden infant deaths were SIDS and 2 in which both were homicide. He also puts the chance of one SIDS death at about 1 in 1300, and the risk of a second at between 5 and 10 times that. Find the odds of double SIDS against double homicide in those families. Find the probability that a double death was SIDS, among the two explanations. Then find the probability of two SIDS deaths in one family, taking the second death as 5 times as likely.

7. Here is a worked answer. Find the step that broke.

```
test for a disease with prevalence 1 in 1,000
false-positive rate 5 per cent, so specificity 95 per cent
a person tests positive
the chance this positive is a false positive is 5 per cent
so the chance the person has the disease is 95 per cent
```

8. Here is a worked answer, from the expert evidence at Sally Clark's trial as the Court of Appeal set it out. Find the step that broke.

```
chance of two SIDS deaths in a family like the Clarks' = 1 in 73 million
about 7,00,000 live births a year in England, Wales and Scotland
7,30,00,000 divided by 7,00,000 = about 104
so by chance this happens about once every hundred years
```

9. A news story on a made-up case says: “The forensic report found a random match probability of one in ten lakh. In other words, there is a 99.9999 per cent chance the accused left the stain.” Decide what to compute, compute it for a pool of 20 lakh men and for a pool of 20 men besides the source, and say what the story does not establish.

10. In a ward round a colleague says: “Her Indian Diabetes Risk Score is 60 or more. Only about 40 per cent of people without diabetes score that high, so there’s about a 60 per cent chance she has undiagnosed diabetes.”

At 60 or more, Mohan and colleagues report a sensitivity of 72.5 per cent and a specificity of 60.1 per cent. Their discussion takes a prevalence of undiagnosed diabetes of 10.1 per cent. Decide what to compute, compute it, correct the colleague in three sentences, and say what your answer does not establish.

10 · Where the prior comes from: pre-test probability in your own clinic

DEFINITION. The prior probability of a condition for a patient, also called the pre-test probability, is the probability that a person like this patient, in this setting, has the condition before the test result is known.

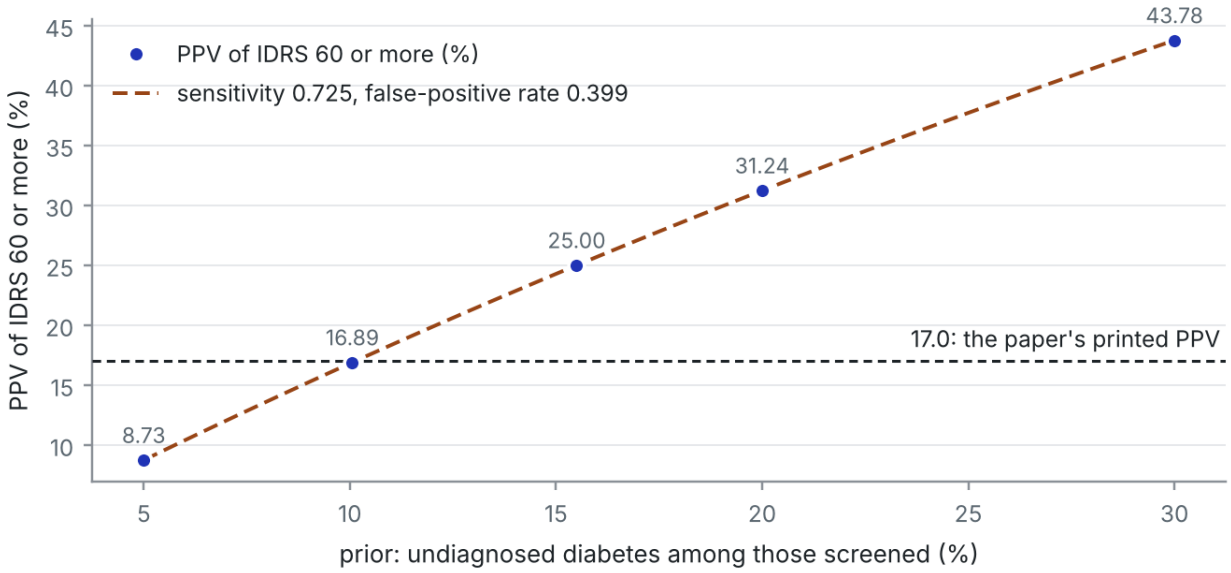
You estimate it as a proportion, in the sense of Book 0, A5. Divide the number of people who had the condition by the number examined, in a stated group over a stated period. Four things fix which proportion is the right one.

1. The condition must be defined exactly as it was in the study that measured the sensitivity and specificity: the same reference test and the same cut-off.
2. The denominator must be the people to whom the test will be applied.
3. The people must be like this patient in what changes the prevalence: age, sex, setting, and why they came.
4. The count must come from a stated period, with every person tested by the reference test, not only those who were suspected. [94, 95, 96, 97]

IN PLAIN TERMS. Bayes' theorem needs a starting chance. That chance is how common the condition is among people like your patient, where you see them.

Nobody hands you that number. You count it. Take the people your clinic would test. Count how many turned out to have the condition. Divide by how many there were.

A national figure, a figure from a study and a figure from your own register are three different groups. Use the group your patient belongs to.



The positive predictive value of an Indian Diabetes Risk Score of 60 or more, at five priors, with sensitivity 72.5 per cent and specificity 60.1 per cent held fixed. At the prior worked out from the paper's counts, about 10.06 per cent, it is 16.89, close to the printed 17.0. The overall prevalence of 15.5 per cent, which includes people already known to have diabetes, would give 25.00.

Illustration 1. Start with a printed predictive value and try to rebuild it. Open "A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects" by Mohan and colleagues (2005), at https://repository.ias.ac.in/80170/. The Indian Diabetes Risk Score (IDRS) adds points for age, waist, physical activity and family history. Search for "optimum sensitivity".

> *An IDRS value ≥ 60 had the optimum sensitivity (72.5%) and specificity (60.1%) for determining diabetes. The positive predictive value was 17.0%*

The sign ≥ is read "greater than or equal to". So a score of 60 or more has sensitivity 72.5 per cent and specificity 60.1 per cent, and a PPV of 17.0 per cent. Here PPV is the positive predictive value of Book 0, D3: the share of positive results that are right. A PPV belongs to a prior. Which prior gives 17.0?

The first number you meet is in the Results. Search for "overall prevalence".

> *Overall, 365 of the 2350 study subjects had diabetes (overall prevalence: 15.5%*

Try 15.5 per cent as the prior, in the PPV formula of section 5. The false-positive rate is 1 minus the specificity.

```
1 minus 0.601 = 0.399
1 minus 0.155 = 0.845
0.725 times 0.155 = 0.112375
0.399 times 0.845 = 0.337155
0.112375 plus 0.337155 = 0.44953
0.112375 divided by 0.44953 = 0.24998…
```

About 25.0 per cent. The paper says 17.0. Something about the prior is wrong.

Read on one sentence.

> *For developing the diabetes risk score, 143 subjects with known diabetes in Phase 3 were excluded.*

The 365 includes people who already knew they had diabetes. Nobody screens them, and the score was never applied to them. Take them out of both the top and the bottom.

```
365 minus 143 = 222
2350 minus 143 = 2207
222 divided by 2207 = 0.10058…
```

About 10.06 per cent. The legend of the paper's Fig. 1 gives the same number analysed, n = 2,207. Now the formula again.

```
1 minus 0.1006 = 0.8994
0.725 times 0.1006 = 0.072935
0.399 times 0.8994 = 0.3588606
0.072935 plus 0.3588606 = 0.4317956
0.072935 divided by 0.4317956 = 0.16891…
```

About 16.9 per cent. The printed sensitivity and specificity are rounded, so 16.9 against 17.0 is as close as they allow. The paper's own Discussion uses the same prior, from the Chennai Urban Rural Epidemiology Study (CURES), a population survey in Chennai: "the prevalence of undiagnosed diabetes … observed in CURES i.e. 10.1%".

So the 17.0 per cent is the PPV among people in Chennai not yet known to have diabetes. It is tied to a prior of about 10 per cent. The 15.5 per cent was a real prevalence too. It was the prevalence of the wrong group.

The table puts the same score at five priors. Each row is the PPV formula with 0.725 and 0.399, and the prior in per cent.

prior (%)	PPV of IDRS 60 or more (%)
5	8.73
10.06	16.89
15.5	25.00
20	31.24
30	43.78

The score's sensitivity and specificity do not change down the table. The PPV changes more than fivefold.

Figures quoted in this illustration are from [98].

WHERE THIS PICTURE BREAKS. The formula reproduces 17.0 only on the assumption that the sensitivity and specificity hold at every prior. They were measured on one Chennai sample,

CURES Phase 3, at one time. In a clinic where the people tested differ in how the score behaves, the table is a guide, not a measurement.

The 10.06 per cent is worked out from the paper's counts. The paper never prints it as the prior of its PPV; the match with 17.0 is what shows it was.

Illustration 2. Now the arm tape of Book 0, D3, and three places you might take its prior from. In Kiran and colleagues' study the tape's cut-off for men was 31.3 cm, with sensitivity 86 per cent and specificity 74 per cent. Those figures fit the 42 men with a body-mass index from 25 to 29.9, as Book 0 found. So the condition the tape was tested against is "in the 25 to 29.9 band".

First, the study's own men. 42 of 131 were in the band, about 32 per cent, and Book 0's rebuilt table gave a PPV of about 61 per cent. That prior describes medical students in one Mangaluru college. Your clinic is not that college.

Second, the national survey. The India fact sheet of the sixth National Family Health Survey (NFHS-6, 2023-24) gives 27.3 per cent of men aged 15 to 49 as having overweight or obesity (a body-mass index of 25 or more). Put it in the formula and you get a number.

```
1 minus 0.273 = 0.727
0.86 times 0.273 = 0.23478
0.26 times 0.727 = 0.18902
0.23478 plus 0.18902 = 0.4238
0.23478 divided by 0.4238 = 0.55398…
```

About 55 per cent. Before you use it, run the four checks of the definition. The survey counts a body-mass index of 25 or more. The tape's figures fit the 25 to 29.9 band only. So the first check fails: the prior and the test are about different conditions. The survey's men are all of India's men aged 15 to 49, so the third check fails too for most clinics. The 55 per cent is arithmetic on a prior that answers a different question.

Third, your own register. Suppose, as a made-up example, that your urban health centre measured every man aged 18 to 29 who came to its outpatient clinic from April to September. The register shows 480 men, all weighed and measured, and 96 of them in the 25 to 29.9 band.

```
96 divided by 480 = 0.2
```

The prior is 0.2, one in five. That is the stated group of Book 0, so its PPV is already known: about 45 per cent.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
0.172 divided by 0.38 = 0.45263…
```

Three sources gave about 61, 55 and 45 per cent for the same tape. Only the last answers the question a man at your clinic asks. It has the right condition, the right people and a stated period, and every man in it was measured.

Figures quoted in this illustration are from [10, 32].

WHERE THIS PICTURE BREAKS. The register is made up, and so is its count of 96 of 480. A real register gives the right prior only if everyone in it was measured. A register that records body-mass index only for men who looked heavy has counted the suspected, and its prevalence is too high.

Every PPV here also assumes the tape's sensitivity and specificity carry from the Mangaluru students to your clinic. That is the subject of the next illustration.

Illustration 3. The prior is not the only estimate in the formula. The tape's sensitivity of 86 per cent is 36 of the 42 men in the band, from Book 0's rebuilt table. Its specificity of 74 per cent is 66 of the 89 men outside it. Each is a count from one small sample, so each would have come out a little differently in another sample of students.

How much does one man matter? Suppose the tape had caught 35 of the 42, or 37. Keep the specificity at 66 of 89, and your clinic's prior at 0.2.

```
35 divided by 42 = 0.8333…
37 divided by 42 = 0.88095…
66 divided by 89 = 0.74157…
1 minus 0.7416 = 0.2584
0.8333 times 0.2 = 0.16666
0.2584 times 0.8 = 0.20672
0.16666 plus 0.20672 = 0.37338
0.16666 divided by 0.37338 = 0.44635…
0.8810 times 0.2 = 0.1762
0.1762 plus 0.20672 = 0.38292
0.1762 divided by 0.38292 = 0.46015…
```

One man either way moves the PPV from about 44.6 to about 46.0 per cent. Let one man move in the specificity too, 65 or 67 of 89, and the nine combinations run from about 43.6 to about 47.1 per cent. Another sample of the same size could easily differ by more than one man.

So write the PPV as "about 45 per cent", not "45.26 per cent". The digits after the point are arithmetic, not knowledge. How wide the honest range is comes from the sampling distribution of the sensitivity and specificity, how they would vary from one sample of the same size to the next, which the next book of this subject teaches.

WHERE THIS PICTURE BREAKS. Moving one man is an illustration of size, not an interval. It does not say how far another sample would actually land; that needs the sampling distribution, taught in the next book of this subject.

The 36 and 66 are Book 0's rebuilt counts, not counts the paper prints. And a sample of medical students is not your clinic, which no count from it can repair.

MUST KNOW POINTS FOR YOU.

- "Use the national prevalence as the prior for my patient" is the error.
- Before you put any prevalence into Bayes' theorem, run four checks aloud. Same condition and cut-off as the test's study? Denominator only the people you would test? People

like your patient? A stated period, with everyone tested? If one fails, say so beside the answer.

- Count your own prior from your register: the number with the condition, the number examined, and the months covered.
- A prevalence that includes people already known to have the condition is the wrong prior for a screening test.
- A register that records the condition only for people who were suspected gives a prior that is too high. Then every PPV you work from it is too high too.
- The PPV you work out carries the error of three estimates: the prior, the sensitivity and the specificity.
- When a paper prints a PPV, find the prior it belongs to before you quote it.

COMMON MISREADING. "The test's PPV is 17 per cent, so 17 per cent of my positives will have diabetes." The 17.0 per cent belongs to the prior it was worked out at, about 10 per cent in the IDRS paper. At a prior of 5 per cent the same score gives about 8.7 per cent; at 20 per cent, about 31 per cent. Find your own prior, then work out your own PPV.

REPORTING IT. Pre-test probability: (the count) of (the number examined) (the people, defined as the test's study defined them), examined at (the clinic) from (month) to (month), had (the condition), or (the per cent) per cent. With the sensitivity and specificity reported by (the study), a positive result raises this to about (the per cent) per cent. Both figures are estimates; treat them as approximate.

EXERCISE 1 (retrieval). From memory: what is the prior probability for a patient, how do you estimate it, and what four checks must the number pass before you put it into Bayes' theorem?

EXERCISE 2 (design). You are preparing the Bayes handout for one test used in your own clinic. Plan how you will get its prior.

Write down six things.

1. The condition, defined as the test's study defined it.
2. The register you will use.
3. Who goes in the denominator, and who you will leave out.
4. The period.
5. How you will check that everyone in the denominator was examined.
6. The one sentence the handout will print about where the prior came from.

EXERCISE 3 (interpretation). Three numbers are offered as "the prevalence of overweight in young men" for the arm-tape handout. NFHS-6 gives 27.3 per cent. Kiran and colleagues' men give 32 per cent. A made-up clinic register gives 20 per cent, 96 of 480 men.

For each, say what kind of number it is, a survey estimate of a national parameter or a proportion in one sample, and which group it describes. Then say which one belongs in the handout, and why.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A register lists 600 people examined in a year, and 45 of them had the condition.

Work out the prevalence as a proportion and as a per cent.

2. A test has sensitivity 0.8 and specificity 0.9.

Work out its PPV at a prior of 0.05 and at a prior of 0.2.

3. A test has sensitivity 0.9 and specificity 0.8. A paper reports its PPV as 0.5.

Work out the prior the paper's PPV belongs to.

4. A made-up register from a primary health centre covers April to June. It lists 300 adults screened with a glucose test. Of these, 40 already knew they had diabetes, and 27 others were newly found to have it.

You want the prior for the Indian Diabetes Risk Score in this centre. Work it out as a per cent, and say who is in the denominator.

5. The Indian Diabetes Risk Score at 60 or more has sensitivity 72.5 per cent and specificity 60.1 per cent (Mohan and colleagues, 2005). In a rural camp your prior for undiagnosed diabetes is 5 per cent.

Work out the PPV of a score of 60 or more in the camp, as a per cent.

6. In Book 0's rebuilt table, Kiran and colleagues' arm tape caught 36 of 42 men in the 25 to 29.9 band and cleared 66 of 89 men outside it. Your clinic's prior is 0.2.

Keep the sensitivity at 36 of 42. Work out the PPV with the specificity at 65 of 89 and at 67 of 89. Then say how many digits of the PPV you would print.

7. Here is a worked answer from a draft handout. Find the step that broke.

```
IDRS 60 or more: sensitivity 0.725, specificity 0.601
prior: diabetes in CURES Phase 3, 365 of 2350 = 0.155
0.725 times 0.155 = 0.112375
0.399 times 0.845 = 0.337155
PPV = 0.112375 divided by 0.449530 = 0.25
so a quarter of people with a score of 60 or more have undiagnosed diabetes
```

8. Here is a worked answer. Find the step that broke.

```
arm tape for men, cut-off 31.3 cm: sensitivity 0.86, specificity 0.74
prior: NFHS-6, men aged 15-49 with overweight or obesity, 0.273
0.86 times 0.273 = 0.23478
0.26 times 0.727 = 0.18902
PPV = 0.23478 divided by 0.4238 = 0.55
so more than half the men the tape flags in my clinic have a body-mass index of 25
or more
```

9. A camp coordinator writes this in a note to the medical officer.

> *The IDRS paper gives a positive predictive value of 17 per cent. So in our camp at the diabetes awareness rally, about one in six people with a score of 60 or more will have undiagnosed diabetes.*

Suppose, for this problem only, that undiagnosed diabetes among people who come to such a rally is 20 per cent. Decide what to compute, compute it, and say what your answer does not establish.

10. A colleague says this at a review meeting.

> *Our handout should use the national figure. NFHS-6 says 30.7 per cent of women have overweight or obesity. That's the most reliable number we have, so it's the right prior for every woman we measure with the arm tape.*

For women, Kiran and colleagues' cut-off is 28.5 cm, with sensitivity 88 per cent and specificity 83 per cent. Decide what to compute, compute it, and say what the answer does not establish.

11 · Counting arrangements: factorials and “n choose k”

DEFINITION. The counting rule. Suppose a task is done in stages, and the number of ways to do each stage does not depend on how the earlier stages were done. Then the number of ways to do the whole task is the product of the numbers of ways for each stage.

For a whole number n, n factorial, written n! and read “n factorial”, is the product of the whole numbers from n down to 1.

```
n! = n × (n - 1) × (n - 2) × … × 2 × 1
```

By the counting rule, n different things can be put in a row in n! orders. The choices go down by 1 to a single one for the last. 0! is defined to be 1, which makes the formulas below hold when k is 0 or n.

An ordered selection of k things from n different things, with k from 0 to n, can be made in n × (n − 1) × ... × (n − k + 1) ways: k factors, starting at n and going down by 1. This equals n! ÷ (n − k)!, because the factors from (n − k) down to 1 cancel top and bottom.

A choice of k things from n when order does not matter, a set rather than a row, is counted by “n choose k”. Each set of k can be put in a row in k! ways. So every set is counted k! times among the ordered selections. Divide by k!.

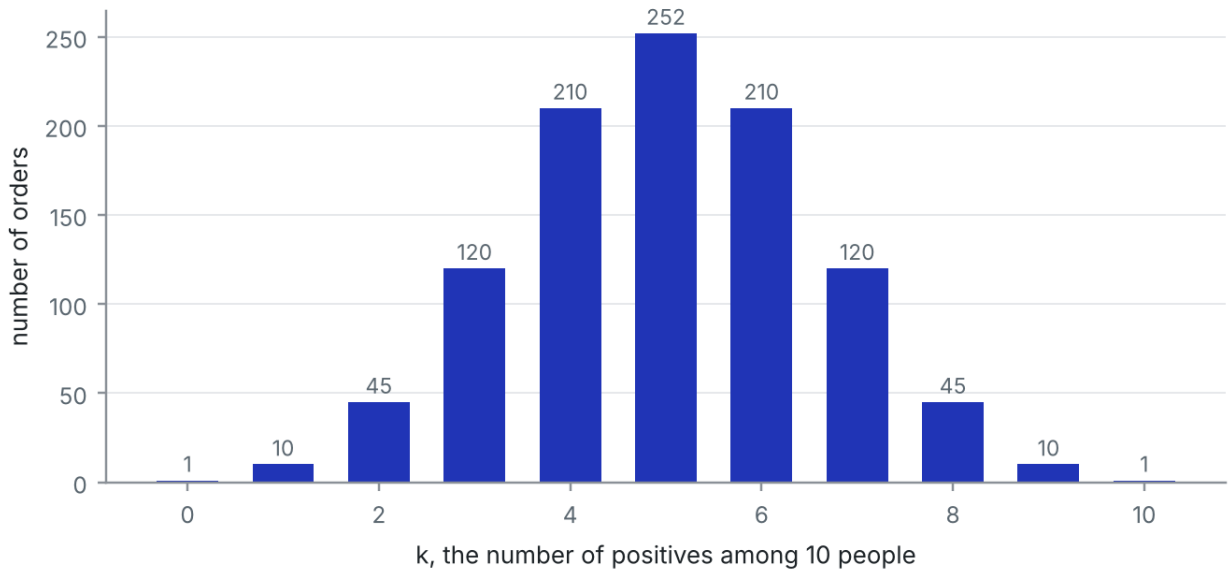
```
n choose k = [n! ÷ (n - k)!] ÷ k! = n! ÷ [k! (n - k)!]
```

Choosing which k to take is the same as choosing which n − k to leave, so n choose k equals n choose (n − k).

Its main use in this book: take n trials, n runs of a chance process each sorted into one of two kinds, a success (the kind being counted) or a failure. The number of different orders in which exactly k successes and n − k failures can fall is n choose k, because an order is fixed by choosing which k of the n trials are the successes. [99, 100, 101, 102, 103, 104, 105, 106]

IN PLAIN TERMS. Count the ways to do something in steps by multiplying the number of choices at each step.

If the order matters, that product is your answer. If it does not, you have counted every group several times, once for each order it can be put in. Divide by that number.



The number of orders in which k positives can turn up among 10 people screened, for k from 0 to 10: 10 choose k. The counts are 1, 10, 45, 120, 210, 252, 210, 120, 45, 10 and 1, and they add up to 1,024, every order of positive and negative among 10 people. The row reads the same from both ends, because choosing k positives is choosing 10 − k negatives.

ILLUSTRATION. You must choose 3 of the 5 subcentres in your block for a supervision visit this month. Call them A, B, C, D and E. How many different choices are there?

Any of 5 can be first, any of the 4 left can be second, any of the 3 left can be third.

```
5 times 4 times 3 = 60
```

That gives 60, and it is the answer to a different question. It counts routes: A then B then C is one, and C then B then A is another. For the visit list, those are the same three subcentres. List the choices themselves, keeping each set in alphabetical order so none is written twice.

sets with A	ABC	ABD	ABE	ACD	ACE	ADE
sets without A	BCD	BCE	BDE	CDE		

Ten sets. Take one set, A, B and C. It can be visited in 3 × 2 × 1 orders: 3 choices for the first stop, 2 for the second, 1 for the last.

```
3 times 2 times 1 = 6
```

Every one of the 10 sets turns up 6 times among the 60 routes.

```
60 divided by 6 = 10
```

That is the whole method. Count in order, then divide by the number of orders each set has.

Four people are screened, and each is positive (+) or negative (−). In how many orders can exactly 2 positives turn up? Choose which 2 of the 4 places hold the +. That is 4 choose 2.

```
4 times 3 = 12
2 times 1 = 2
12 divided by 2 = 6
```

The six are ++−−, +−+−, +−−+, −++−, −+−+ and −−++. When the next section works out the chance of exactly 2 positives, it adds up the chance of each of these six orders.

For large numbers, cancel before you multiply. In Book 0's rebuilt table of Kiran and colleagues' study, the arm tape missed 6 of the 42 men in the band. In how many ways could 6 of 42 men be the ones missed? That is 42 choose 6. The factors of 36! cancel from top and bottom, leaving 6 factors on top.

```
42 times 41 times 40 times 39 times 38 times 37 = 3776965920
6 times 5 times 4 times 3 times 2 times 1 = 720
3776965920 divided by 720 = 5245786
```

5,245,786 different sets of six men.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The counts hold only when the things are all different and each is used at most once.

The 6 missed men are Book 0's rebuilt count, not a number the paper prints.

MUST KNOW POINTS FOR YOU.

- "Multiply the choices at each step" counts ordered arrangements. Before you multiply, ask aloud: is A then B different from B then A here? If not, divide by k!.
- Check a counting formula on a case small enough to list. Write out every set for 3 from 5, or every order of 2 positives among 4 people.
- Cancel before you multiply. For 42 choose 6, keep six factors on top and 6! below.
- n choose k equals n choose (n − k).
- A count of ways is not a probability. It becomes one only if every way is equally likely, or after each way is weighted by its own chance.
- The formulas assume the n things are all different and none is used twice.

EXERCISE 1 (retrieval). From memory: what does n! mean, and what is 0!? Write the formula for n choose k, and explain in two sentences why it divides by k!.

EXERCISE 2 (teaching). A first-year resident asks: "Isn't the chance of exactly 2 positives among 4 people just p times p times (1 − p) times (1 − p)?"

Answer in about ten minutes at a whiteboard. Use the four-person example and list what you need to list.

PRACTICE. Twelve problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Work out 5!, 0! and 6!.

2. Four things are labelled A, B, C and D. List every way to choose 2 of them when order does not matter. Then check your count with the formula for 4 choose 2.

3. Work out 8! ÷ 5! without working out 8!.

4. Work out 9 choose 2 and 9 choose 7.

5. The counts of ways to choose k from 5, for k from 0 to 5, are 1, 5, 10, 10, 5, 1.

Write the counts for 6, for k from 0 to 6. Use only those six numbers and addition, then check 6 choose 3 with the formula. Hint: a set of k from 6 either leaves out the sixth thing or includes it.

6. Your block has 12 subcentres. This week you can visit 3.

How many different sets of 3 subcentres could you choose? How many different routes, if the order of visits matters? Give each with its unit.

7. Ten people are screened at a camp. Each result is positive or negative.

In how many different orders can exactly 2 positives turn up? What share of all possible orders of results is that?

8. In Book 0's rebuilt table of Kiran and colleagues' study, the arm tape caught 36 of the 42 men in the 25 to 29.9 band.

In how many different ways could 36 of 42 men be the ones caught? Give the count with its unit, and say which smaller choose count gives the same number.

9. Here is a worked answer. Find the step that broke.

```
a review committee of 3 is to be chosen from 8 medical officers
first member: 8 choices
second member: 7 choices
third member: 6 choices
number of committees = 8 times 7 times 6 = 336
```

10. Here is a worked answer. Find the step that broke.

```
number of ways to choose 3 of 10 = 10! divided by 3!
10! = 3628800
3! = 6
3628800 divided by 6 = 604800
```

11. A district officer says this about a pilot.

> *We picked 3 of our 6 primary health centres (PHCs) for the pilot. There are only 20 possible sets of three, so the chance that we happened to pick the three best-run PHCs was 1 in 20. The pilot's good results are not about which PHCs we chose.*

Decide what to compute, compute it, and say what the answer does not establish.

12. A report on a screening camp says this.

> *Each person screened has a 0.2 chance of a positive result. So the chance that exactly 2 of 10 people are positive is 0.2 times 0.2 times 0.8 to the power 8, about 0.007. Two positives among ten would be very unusual.*

Decide what to compute, compute it, and say what your answer does not establish.

12 · The Bernoulli and binomial distributions: counting successes in independent trials

DEFINITION. A trial is one run of a chance process whose result is sorted into two kinds. The kind being counted is called a success and the other a failure.

A Bernoulli random variable takes the value 1 for a success, with probability p, and 0 for a failure, with probability 1 − p. For trial j, call it Xj.

```
E[Xj] = 0 × (1 - p) + 1 × p = p
Var(Xj) = (0 - p)^2 × (1 - p) + (1 - p)^2 × p = p(1 - p)
```

The second line simplifies because $p^2 (1 - p) + (1 - p)^2 p$ takes out the common factor $p(1 - p)$, leaving $p + (1 - p)$, which is 1.

Now run n trials. The count of successes, X, has a binomial distribution with parameters n and p when four conditions hold:

1. the number of trials, n, is fixed before you start;
2. each trial is a success or a failure;
3. the trials are independent, in the sense of section 3;
4. the probability of a success, p, is the same on every trial.

A model's parameters are the fixed numbers that pick out one distribution: here n and p.

Any one order of k successes and n − k failures has probability $p^k (1 - p)^{(n - k)}$, by multiplying across independent trials. There are n choose k such orders (section 11), and no two orders can both happen, so their probabilities add (rule 3, for any number of pieces, as in section 1).

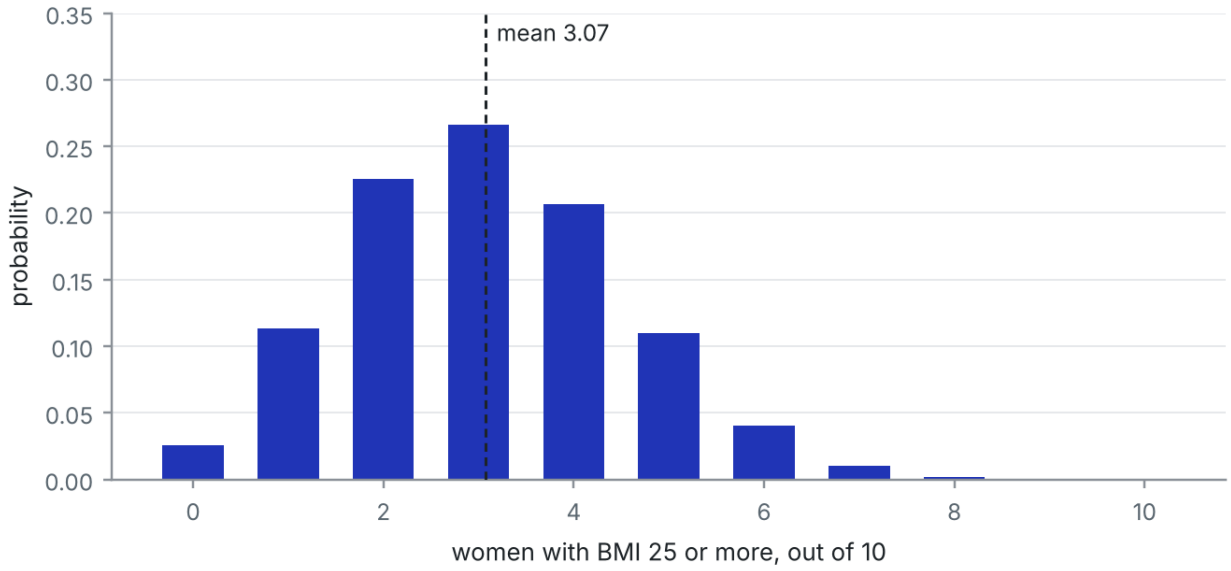
```
P(X = k) = (n choose k) × p^k × (1 - p)^(n - k)
```

The count X is the sum of n Bernoulli variables, one per trial: X = X1 + X2 + … + Xn. The expected value of a sum is the sum of the expected values, with no condition. The variance of a sum of independent variables is the sum of their variances (Book 2, section 15). SD(X) is the standard deviation of X, called σ in Book 2, section 15: the square root, written √, of Var(X). So:

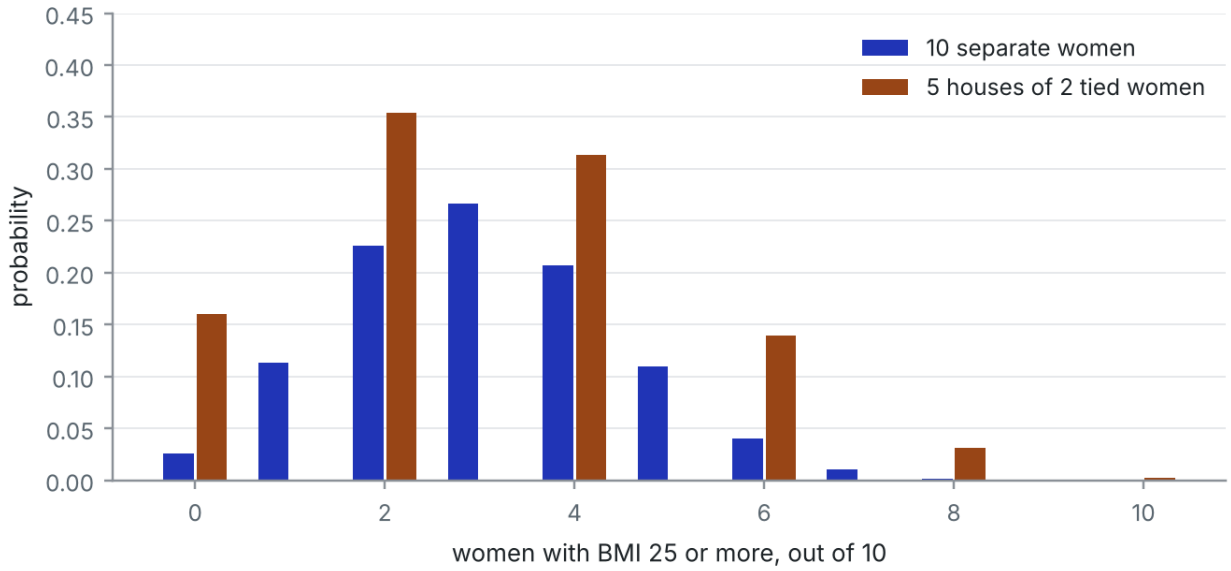
```
E[X] = p + p + … + p = np
Var(X) = p(1 - p) + … + p(1 - p) = np(1 - p)
SD(X) = √(np(1 - p))
```

The mean np needs only conditions 1, 2 and 4. The variance np(1 − p) needs independence too. [107, 108, 109, 110, 111, 112, 113]

IN PLAIN TERMS. Ask ten women, picked at random, one question with a yes-or-no answer. Count the yeses. Suppose every woman has the same chance of a yes. Suppose too that no answer tells you anything about another. Then the count follows a fixed pattern: the binomial.



The binomial distribution of how many of 10 randomly picked women aged 15-49 have a body-mass index of 25 or more, when each has a chance of 0.307 (NFHS-6, 2023-24, India). Exactly 3, the count nearest the mean, has a chance of about 0.27; nearly three times in four the count is something else. The expected count is 3.07, a long-run average, not a count any group can have.



The same 10 women, two ways. Left bar of each pair: 10 separate women, each with a chance of 0.307. Right bar: 5 houses of 2 women whose statuses always match, a made-up extreme. The mean is 3.07 either way. Tied women spread the count wider, and make a count of 0 about six times as likely: 0.1598 against 0.0255.

Illustration 1. The national figure first. Open the India fact sheet of the National Family Health Survey's sixth round (NFHS-6, 2023-24), from https://www.nfhsiips.in, and find indicator 76. In it, 30.7% of women aged 15 to 49 had a body-mass index (BMI) of 25.0 kg/m^2 or more: overweight or obesity.

Take that as p for one woman picked at random from India's women aged 15 to 49. So p = 0.307 and 1 − p = 0.693. Now pick 10 women at random, and let X be how many of them have a body-mass index of 25 or more.

A colleague says: "Nearly one in three. So in any 10 women, 3 will be overweight." Test that before you agree.

Check the four conditions. The 10 is fixed. Each woman's index is 25 or more, or it is not. Women picked at random from a population of crores do not inform on each other, so treat them as independent. And each has the same chance, 0.307. The count is binomial, with n = 10 and p = 0.307.

Work out P(X = 3). The number of orders is 10 choose 3. One order has three successes and seven failures.

```
10 choose 3 = 120
0.307^3 = 0.028934
0.693^7 = 0.076760
120 times 0.028934 times 0.076760 = 0.2665
```

About 0.27. About three in four random groups of 10 have some other count. Here is the whole distribution, every value worked the same way.

```
women with BMI 25 or more, k    P(X = k)
0    0.0255
1    0.1132
2    0.2256
3    0.2665
4    0.2066
5    0.1098
6    0.0405
7    0.0103
8    0.0017
9    0.0002
10    0.0000
```

The last is not zero. It is below 0.00005, so it rounds to 0.0000.

Now the mean and spread, from the sum of 10 Bernoulli variables.

```
10 times 0.307 = 3.07
10 times 0.307 times 0.693 = 2.1275
√2.1275 = 1.459
```

The expected count is 3.07 women. No group of 10 can have 3.07 women; the expected value is a long-run average over many groups (Book 2, section 14). The standard deviation is about 1.46 women.

The chance of at least one is fastest through the complement (Book 0, D2): one minus the chance of none.

```
0.693^10 = 0.0255
1 minus 0.0255 = 0.9745
```

About one random group of 10 in forty has no woman with an index of 25 or more.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The 0.307 is itself a survey estimate for all of India's women aged 15 to 49, with pregnant women and those who gave birth in the last two months left out (footnote 15 of the fact sheet). Treating it as the fixed p of the model is a choice, and every probability above inherits it. For urban women alone NFHS-6 gives 42.8%, and for rural women 25.5%: a group drawn from one of those has a different p.

The binomial also needs the 10 to be picked at random from that whole population. Ten women who came to one clinic, or who live in one lane, are not.

Illustration 2. Now break condition 3 and watch what changes. A survey team, for convenience, picks 5 households at random and measures 2 women in each, 10 women in all. Two women who cook from one kitchen and live in one house are not independent.

How strongly tied they are is not known. So take the extreme, as a made-up model: in every house, the two women either both have an index of 25 or more, or neither does. Each house is then one trial with p = 0.307, and the count of women is twice the count of houses. Call the count of houses Y. It is binomial with n = 5 and p = 0.307.

```
5 choose 1 = 5
5 choose 2 = 10
0.693^5 = 0.1598
5 times 0.307 times 0.693^4 = 0.3540
10 times 0.307^2 times 0.693^3 = 0.3137
10 times 0.307^3 times 0.693^2 = 0.1390
5 times 0.307^4 times 0.693 = 0.0308
0.307^5 = 0.0027
```

Those are P(Y = 0) to P(Y = 5), so the chances of 0, 2, 4, 6, 8 and 10 women. Set them beside 10 separate women.

women with BMI 25 or more	10 separate women	5 houses of 2 tied women
0	0.0255	0.1598
1	0.1132	0
2	0.2256	0.3540
3	0.2665	0
4	0.2066	0.3137
5	0.1098	0
6	0.0405	0.1390
7	0.0103	0
8	0.0017	0.0308

9	0.0002	0
10	0.0000	0.0027

The mean has not moved. Each woman still has a 0.307 chance, and the sum rule for expected values needs no independence.

```
2 times 5 times 0.307 = 3.07
```

The spread has. The count is 2Y, and Var(2Y) is 4 times Var(Y) (Book 2, section 15).

```
5 times 0.307 times 0.693 = 1.0638
4 times 1.0638 = 4.2552
√4.2552 = 2.063
```

A standard deviation of 2.06 women, against 1.46 for separate women. A group with nobody at 25 or more is now far more likely.

```
0.1598 divided by 0.0255 = 6.27
```

About six times as likely: 0.1598 against 0.0255.

WHERE THIS PICTURE BREAKS. Real women in one house are tied, but not completely. The truth lies between the two columns, and the size of the tie is something data must show. At the fully tied extreme the spread grows. With weaker ties it grows less, and it still grows while the pair's statuses go together. The mean stays where it was.

MUST KNOW POINTS FOR YOU.

- "Nearly one in three, so 3 in any 10" is the error. Before you predict a count for one group, give the chance of a range.
- Check the four conditions before you use the formula: fixed n, two outcomes, independent trials, the same p.
- People drawn from one household, one clinic, one school or one measurer are not independent trials. The binomial then understates the spread, though it gets the mean right.
- The binomial gives the chance of each count when p is known. It cannot tell you whether the p you put in is right for your group.
- For "at least one", work out the chance of none, $(1 - p)^n$, and take it from 1.
- A survey figure used as p is an estimate of a population number, treated as if it were the model's fixed p. Say so when you quote a binomial chance built on it: "if the share is 30.7%, then ...". The "if" is part of the answer.
- When you teach this, say out loud that "success" means only "the outcome we count".

COMMON MISREADING. "The expected count is 3.07, so a group of 10 will have 3." The expected value is the average count over many random groups. Any one group can have 1, 2, 4 or 5, and to-

gether those are more than twice as likely as 3. When you predict for one group, give the chance of a range of counts, not the mean.

EXERCISE 1 (retrieval). From memory: state the four conditions for a count to be binomial. Write the probability of exactly k successes in n trials, and say what each of its three parts counts. Then give the mean and the variance, and say which condition the variance needs that the mean does not.

EXERCISE 2 (teaching). A first-year resident, ten minutes, a whiteboard. She knows E[X] and Var(X) from a single variable and has seen the binomial formula. She asks why the mean of a binomial count is np and its variance np(1 − p). Plan what you write on the board, line by line.

PRACTICE. Fourteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. X is a Bernoulli random variable with p = 0.3. Work out E[X], Var(X) and SD(X).

2. X is binomial with n = 5 and p = 0.4. Work out P(X = 2).

3. X is binomial with n = 6 and p = 0.1. Work out the probability that X is at least 1.

4. X is binomial with n = 50 and p = 0.2. Work out its mean, variance and standard deviation.

5. A binomial count has mean 6 and variance 4.2. Work out n and p.

6. In NFHS-6, 27.3% of men aged 15 to 49 in India had overweight or obesity (indicator 77). Eight men are picked at random from that population. Work out the chance that none of them has a body-mass index of 25 or more, and the chance that exactly 2 do.

7. A survey picks 50 women aged 15 to 49 at random from all of India. Take p as NFHS-6's 30.7% for a body-mass index of 25 or more. Work out the expected number with an index of 25 or more and the standard deviation. The survey finds 25. How many standard deviations above the mean is that?

8. A study picks 5 urban women and 5 rural women aged 15 to 49 at random. In NFHS-6, 42.8% of urban and 25.5% of rural women had a body-mass index of 25 or more. Is the count of all 10 women with an index of 25 or more binomial? Work out its expected value and its variance.

9. In the Chennai Urban Rural Epidemiology Study (CURES), an Indian Diabetes Risk Score of 60 or more had a sensitivity of 72.5% for undiagnosed diabetes. Eight people with undiagnosed diabetes, unknown to each other, are scored. Work out the expected number the score misses and the chance it flags all eight.

10. Here is a worked answer. Find the step that broke.

```
10 women picked at random, p = 0.307
P(exactly 2 with BMI 25 or more) = 0.307^2 times 0.693^8
= 0.094249 times 0.053194 = 0.0050
so a group of 10 with exactly 2 is rare, about 1 in 200
```

11. Here is a worked answer. Find the step that broke.

```
10 women picked at random, p = 0.307
the chance of at least one with BMI 25 or more = 10 times 0.307 = 3.07
3.07 is more than 1, so at least one is certain
```

12. Here is a worked answer. Find the step that broke.

```
a team visits 5 households at random and measures 2 women aged 15-49 in each
the count of the 10 with BMI 25 or more is binomial, n = 10, p = 0.307
SD = √(10 times 0.307 times 0.693) = 1.46
so a count of 0 is very unlikely: 0.693^10 = 0.026
```

13. A district training slide says: "Nationally nearly one woman in three is overweight or obese. So in any group of 10 women, 3 will be overweight." Decide what to compute, compute it, and say what your answer does not establish.

14. A medical officer measures 20 women at one anganwadi (a government child-care centre under the Integrated Child Development Services scheme) and finds 11 with a body-mass index of 25 or more. She writes: "Against the national 30.7%, this centre clearly has a problem." Decide what to compute, compute it, and say what the answer does not establish.

13 · The geometric and hypergeometric distributions: waiting for the first success, and drawing without replacement

DEFINITION. The binomial of section 12 fixes the number of trials and assumes independent trials with one p. Each distribution here relaxes one of those.

Geometric. Run independent trials, each a success with the same probability p, and stop at the first success. Let X be the number of trials up to and including the first success.

$$P(X = k) = (1 - p)^{(k - 1)} \times p$$

The chance that more than k trials are needed is the chance that the first k all fail.

$$P(X > k) = (1 - p)^{k}$$

The expected value is 1 ÷ p. One argument: the first trial is used either way. With probability 1 − p it fails, and the wait starts again from nothing, because the trials are independent and p does not change. So E[X] = 1 + (1 − p) × E[X]. Take (1 − p) × E[X] from both sides: p × E[X] = 1.

Some books count instead the failures before the first success, which is X − 1. Its probabilities are shifted down by one and its mean is (1 − p) ÷ p.

Hypergeometric. A group of N people contains K with some feature. Draw n of them at random without replacement: once drawn, a person cannot be drawn again. Let X be the number with the feature among the n. Every set of n people is equally likely, and there are N choose n sets.

$$P(X = k) = (K \text{ choose } k) \times (N - K \text{ choose } n - k) \div (N \text{ choose } n)$$

The draws are dependent, as section 3 showed for drawing without replacement. But each single draw, looked at alone, is a person with the feature with probability K ÷ N. This is because, before anything is seen, every person is equally likely to be the one drawn first, second, or at any place in the order. Expected values add with no condition (Book 2, section 14), so the mean is n × K ÷ N. That is the binomial's mean, with p = K ÷ N. The spread is smaller than the binomial's, because each draw removes someone from the group.

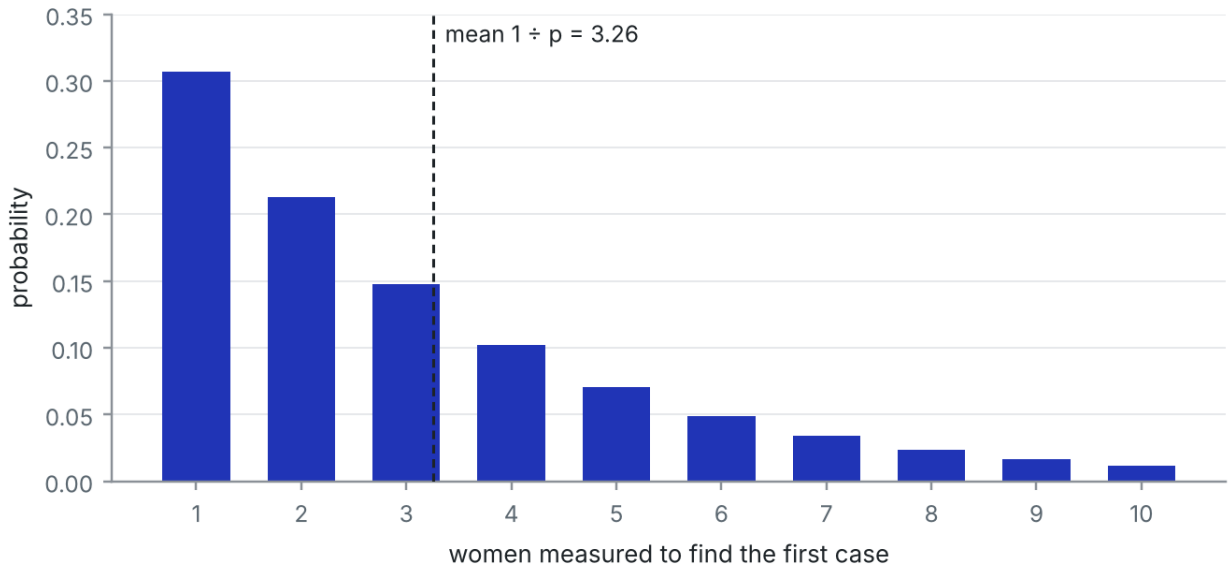
$$\mathrm{Var}(X) = n \times (K \div N) \times (1 - K \div N) \times (N - n) \div (N - 1)$$

The last factor, (N − n) ÷ (N − 1), is below 1 and is what makes the spread smaller.

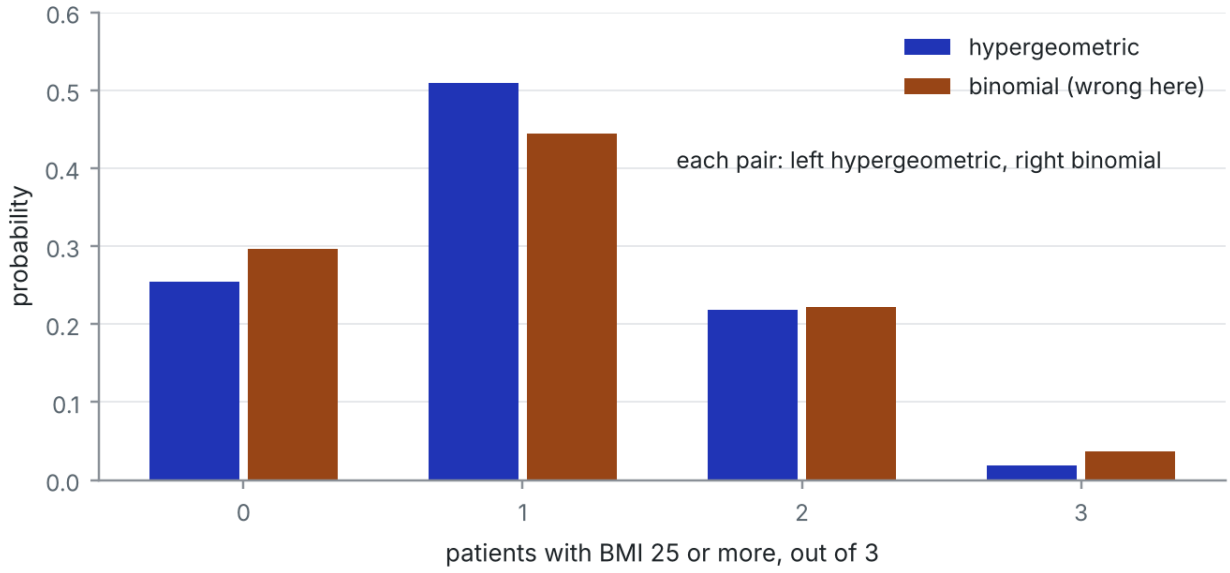
When N is large against n, removing a few people barely changes the share of the rest. The hypergeometric then comes close to the binomial with n and p = K ÷ N. A common rule of thumb: draws of under a tenth of the group behave nearly as independent. [114, 115, 116, 117, 118, 119, 120, 121, 122]

IN PLAIN TERMS. You keep going until you find the first one you are looking for, and count how many it took. That count is geometric.

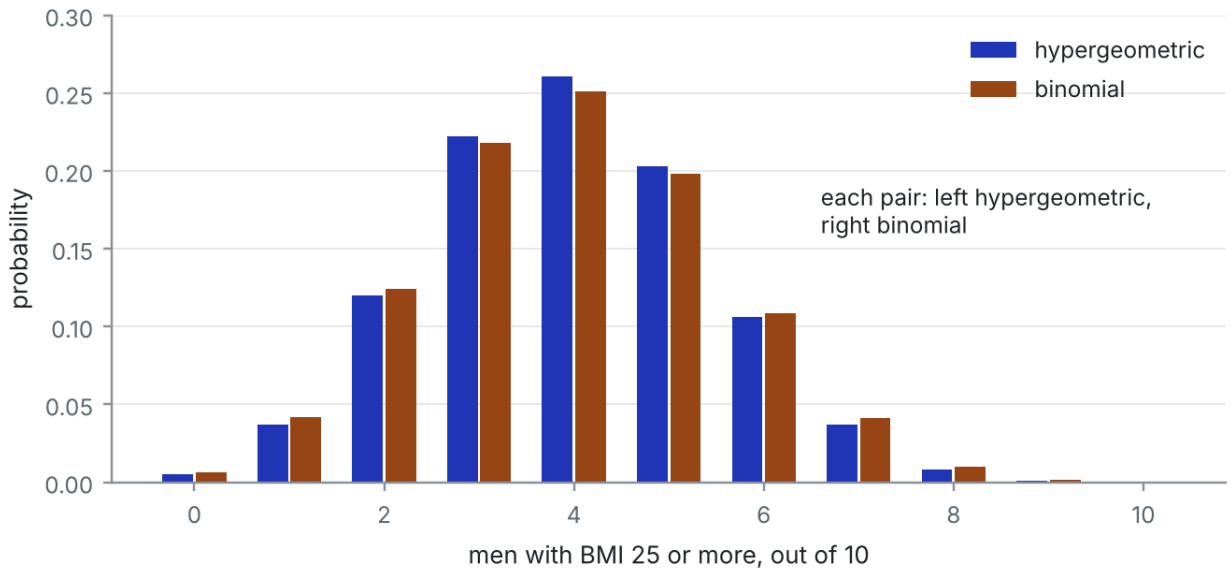
In the second you draw from a small group, and each person drawn leaves the group. The count is hypergeometric.



How many randomly picked women aged 15-49 must be measured to find the first with a body-mass index of 25 or more, when each has a chance of 0.307 (NFHS-6, 2023-24, India). The first woman is the most likely place, at 0.307; each later bar is 0.693 times the one before. The mean, 3.26, sits to the right of the most likely value.



A made-up ward of 12 patients, 4 with a body-mass index of 25 or more, and an audit of 3 picked at random. Drawing without replacement (left bar of each pair) against the binomial with p = 1/3 (right bar), which wrongly treats the picks as independent. The binomial is too high at the ends: 0.2963 against 0.2545 for none, 0.0370 against 0.0182 for all three. Both have mean 1.



Ten men drawn at random from Kiran and colleagues' 131, of whom 52 had a body-mass index of 25 or more. The draw is 7.6 per cent of the group, and the hypergeometric (left bar of each pair) and the binomial with p = 0.397 (right bar) differ by about 0.01 at most. Both have mean 3.97.

Illustration 1. A health worker goes through a list of women aged 15 to 49, picked at random from all of India, measuring each until she finds the first with a body-mass index (BMI) of 25 or more. Take p as the 30.7% of the National Family Health Survey (NFHS-6, 2023-24), so p = 0.307. Let X be the number of women she measures, counting the one she finds.

The tempting reading: "nearly one in three, so it will be the third woman". Work out the chances and see.

```
P(X = 1) = 0.307
0.693 times 0.307 = 0.2128
0.693^2 times 0.307 = 0.1474
```

Those are the chances that the first case is the first, second and third woman. The third is less likely than the first. The first woman is the single most likely place to find the first case, as it is for every geometric distribution. Here are the first ten.

```
women measured, k    P(X = k)
1    0.3070
2    0.2128
3    0.1474
4    0.1022
5    0.0708
6    0.0491
7    0.0340
8    0.0236
9    0.0163
10    0.0113
```

So where does "one in three" belong? In the mean.

```
1 divided by 0.307 = 3.26
```

On average she measures about 3.26 women to find the first case. The long tail of slow searches pulls the average past the most likely value.

Two more questions a supervisor asks. What is the chance she finds one within five women, and the chance she measures ten and finds none?

```
0.693^5 = 0.159833
1 minus 0.159833 = 0.840167
0.693^10 = 0.0255
```

About 0.84 within five. And about 0.026 for none in ten, which is the binomial chance of 0 out of 10 in section 12. It is the same event: ten failures in a row.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The geometric needs the same p for every woman, and independence. A health worker going door to door in one lane meets women who share a place; one who stops at the houses that look well-off first changes p as she goes. Then the shape and the mean 3.26 no longer hold.

The 0.307 is a national survey estimate for women aged 15 to 49, used here as if exact.

Illustration 2. Now a small group, and a mistake first. A made-up ward has 12 patients, and 4 of them have a body-mass index of 25 or more. An audit picks 3 patients at random. A trainee writes: “p is 4 out of 12, a third. The count is binomial with n = 3, so the chance of none is (2/3) cubed, 0.296.”

The trainee broke condition 3 of section 12. The audit cannot pick a patient twice, so each pick changes who is left. Count sets instead. There are 12 choose 3 possible sets of 3.

```
12 choose 3 = 220
8 choose 3 = 56
56 divided by 220 is about 0.2545
```

No patient at 25 or more means all 3 come from the other 8. The chance is about 0.2545, not 0.296. Now every count, the same way: choose k of the 4 and 3 − k of the 8.

```
4 choose 1 times 8 choose 2 = 4 times 28 = 112
4 choose 2 times 8 choose 1 = 6 times 8 = 48
4 choose 3 times 8 choose 0 = 4 times 1 = 4
56 plus 112 plus 48 plus 4 = 220
```

The four counts cover all 220 sets, as they must.

patients with BMI 25 or more, k	hypergeometric	binomial (wrong here)
0	0.2545	0.2963
1	0.5091	0.4444

2	0.2182	0.2222
3	0.0182	0.0370

The mean is the same for both: n × K ÷ N.

```
3 times 4 divided by 12 = 1
```

One patient on average either way. The spread is not the same. The binomial puts more weight on the ends, 0 and 3. Drawing without replacement pulls the count towards the middle: after one patient at 25 or more is drawn, fewer are left to draw. Work out the hypergeometric variance from its table, as in Book 2, section 15: each count's squared distance from the mean 1, which is $(k - 1)^2$, that is 1, 0, 1 and 4, times its chance.

```
1 times 0.2545 plus 0 times 0.5091 plus 1 times 0.2182 plus 4 times 0.0182 = 0.5455
```

The formula in the definition gives the same, with last factor 9 ÷ 11. The binomial's is np(1 − p), the formula of section 12.

```
3 times 1/3 times 2/3 times 9/11 is about 0.5455
3 times 4 divided by 12 times 8 divided by 12 is about 0.6667
```

A variance of about 0.545 without replacement, against 0.667 for the binomial.

WHERE THIS PICTURE BREAKS. The ward is made up. The hypergeometric is exact here because you know N, K and n and the 3 are picked at random from the 12. In real work K is usually the very thing you do not know. The formula then gives the chance of a count for each possible K, which is the start of inference, not the end of it.

Illustration 3. Last, a bigger group, where the two distributions meet. Kiran and colleagues measured 131 men, medical students in Mangaluru. In their Table 2, 42 men had a body-mass index of 25 to 29.9, and 10 had 30 or more. So 52 of the 131 had 25 or more.

Draw 10 of the 131 men at random, without replacement. The draw is about a thirteenth of the group.

```
10 divided by 131 is about 0.0763
52 divided by 131 is about 0.3969
```

About 7.6 per cent of the group, under the rule-of-thumb tenth. The hypergeometric has N = 131, K = 52 and n = 10. The binomial stand-in has n = 10 and p = 52 ÷ 131, about 0.397.

men with BMI 25 or more, k	hypergeometric	binomial
0	0.0050	0.0064
1	0.0371	0.0419
2	0.1199	0.1240
3	0.2221	0.2177
4	0.2609	0.2508

5	0.2031	0.1981
6	0.1061	0.1087
7	0.0367	0.0409
8	0.0080	0.0101
9	0.0010	0.0015
10	0.0001	0.0001

No row differs by more than about 0.01. Both means are the same.

```
10 times 52 divided by 131 is about 3.97
```

About 3.97 men either way. With a tenth of the group or less drawn, the binomial is close enough for most purposes. With a quarter drawn, as in the ward, it is not.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The 0.01 is for these numbers. The gap grows as the draw becomes a larger share of the group, and it is largest in the tails, where a rare count is exactly what you may want to judge.

Kiran's 52 is a count from one college. The paper's printed "overweight" count is the 25 to 29.9 band only; 52 adds the 10 men at 30 or more from the same table.

MUST KNOW POINTS FOR YOU.

- "One in three, so the third person" is the error. In a geometric wait the first person is always the single most likely place for the first success; 1 ÷ p is only the average.
- Is the number of people fixed, or do you stop at the first success? Are people drawn from a group they then leave, and is the draw more than about a tenth of it? Fixed and a small share: binomial. Stop at first: geometric. Fixed and a large share of a known group: hypergeometric.
- The binomial used on a draw from a small group gets the mean right and the spread too wide.
- The geometric holds only while p stays the same and the trials stay independent.
- The chance of no success in the first k trials is $(1 - p)^k$, whatever you call the distribution.
- The hypergeometric needs K, the number with the feature in the group. When K is the unknown, use the formula to ask how likely your count would be under each K you are weighing, not to state the count's chance.
- Counting trials including the success, the mean is 1 ÷ p. Counting failures before it, the mean is (1 − p) ÷ p.

COMMON MISREADING. "The mean wait is 1 ÷ p, so the first success comes at about the 1 ÷ p-th person." The mean is an average over many searches, pulled up by a long tail of slow ones. The

most likely place for the first success is always the first person. And with p = 0.307, about one search in forty still finds none in ten (0.026).

EXERCISE 1 (retrieval). From memory: write P(X = k) and the mean for the geometric distribution, counting trials up to and including the first success. Then write P(X = k) and the mean for the hypergeometric, and say when the binomial can stand in for it.

EXERCISE 2 (teaching). A first-year resident, ten minutes, a whiteboard. She is planning a record audit of 20 files from a ward register of 60. She wants to use the binomial for the number of files with an error. Explain whether she can, and what changes if she cannot.

PRACTICE. Eleven problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. X is geometric, counting trials up to and including the first success, with p = 0.2. Work out P(X = 3) and E[X].

2. X is geometric with p = 0.1. Work out P(X > 4) and P(X ≤ 4).

3. A group of 10 has 3 with a feature. Four are drawn at random without replacement. Work out the chance that exactly 1 of the 4 has the feature, and the expected number.

4. A geometric count of trials has mean 8. Work out p, and the chance that the first trial is a success.

5. In the Chennai Urban Rural Epidemiology Study (CURES), 17.0% of people with an Indian Diabetes Risk Score of 60 or more had undiagnosed diabetes (the positive predictive value). A clinic confirms such people one at a time with a blood test. Work out the expected number tested to find the first confirmed case, and the chance that the first 5 are all negative.

6. Kiran and colleagues measured 131 men. Of them, 52 had a body-mass index of 25 or more: 42 at 25 to 29.9 and 10 at 30 or more. Three men are picked at random. Work out the chance that all three have an index of 25 or more, without replacement, and compare it with the binomial answer.

7. Choose the distribution for each situation, and give the expected count. (a) A district survey picks 50 women aged 15 to 49 at random from lakhs and counts those with a body-mass index of 25 or more. (b) A health worker measures randomly picked women from the same population. She stops at the first with an index of 25 or more and counts the women she measured. (c) A teacher picks 10 of Kiran's 131 men at random and counts those with an index of 25 or more. Take p as NFHS-6's 30.7% for (a) and (b).

8. Here is a worked answer. Find the step that broke.

```
first woman with BMI 25 or more, p = 0.307
the chance it is the fourth woman = 0.693^4 times 0.307
= 0.230639 times 0.307 = 0.0708
```

9. Here is a worked answer. Find the step that broke.

```
a ward of 12 patients, 4 with BMI 25 or more; an audit picks 3 at random
p = 4 divided by 12 = 0.3333
chance all 3 have BMI 25 or more = 0.3333^3 = 0.0370
```

10. A programme note says: “Nearly one woman in three aged 15 to 49 is overweight or obese. So a health worker visiting homes will find one by the third woman she measures at the latest.” Decide what to compute, compute it, and say what your answer does not establish.

11. A colleague draws 10 of the 131 men’s records from Kiran and colleagues’ study at random. She finds none with a body-mass index of 25 or more, and says: “The paper’s 52 must be wrong.” Decide what to compute, compute it, and say what the answer does not establish.

14 · The Poisson distribution: counts of events at a steady rate

DEFINITION. A Poisson process produces events one at a time, in time or in space, under three conditions. Events happen at a constant average rate. The numbers of events in intervals that do not overlap are independent. Two events at exactly the same instant do not happen.

Let X be the number of events in a fixed interval of the process. The average number of events in that interval is written with the Greek letter lambda, λ. X then has a Poisson distribution with parameter λ. For k = 0, 1, 2 and so on, with no upper limit:

$$P(X = k) = e^{(-\lambda)} \times \lambda^{k} \div k!$$

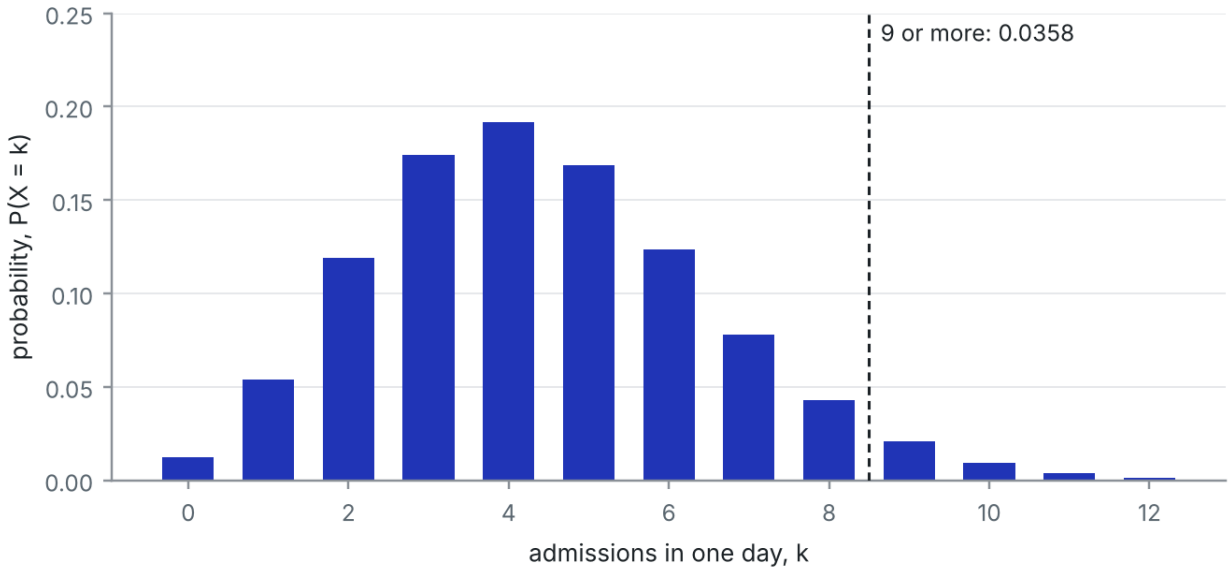
$$E[X] = \lambda$$
$$Var(X) = \lambda$$
$$SD(X) = \sqrt{\lambda}$$

The Poisson distribution is the limit of the binomial distribution of section 12 when the number of trials n grows and the probability p shrinks, with np held at λ. So a binomial count with large n and small p is close to a Poisson count with λ = np.

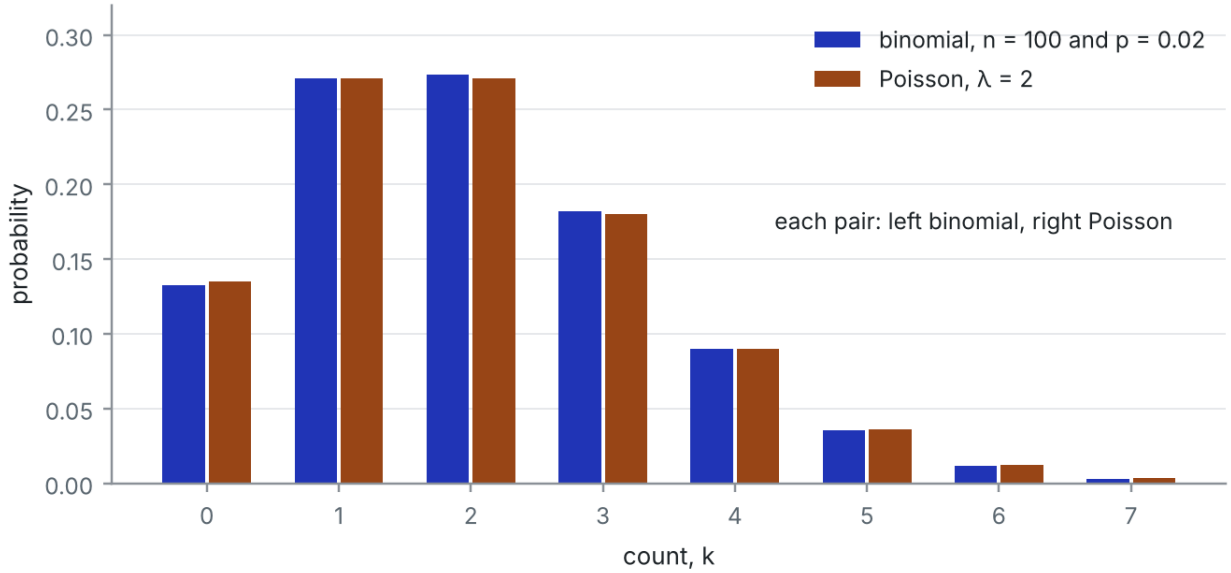
Counts whose variance is well above their mean are called overdispersed. [123, 124, 125, 126, 127, 128, 129, 130]

IN PLAIN TERMS. Some events turn up one at a time, at a steady average rate, each without regard to the others.

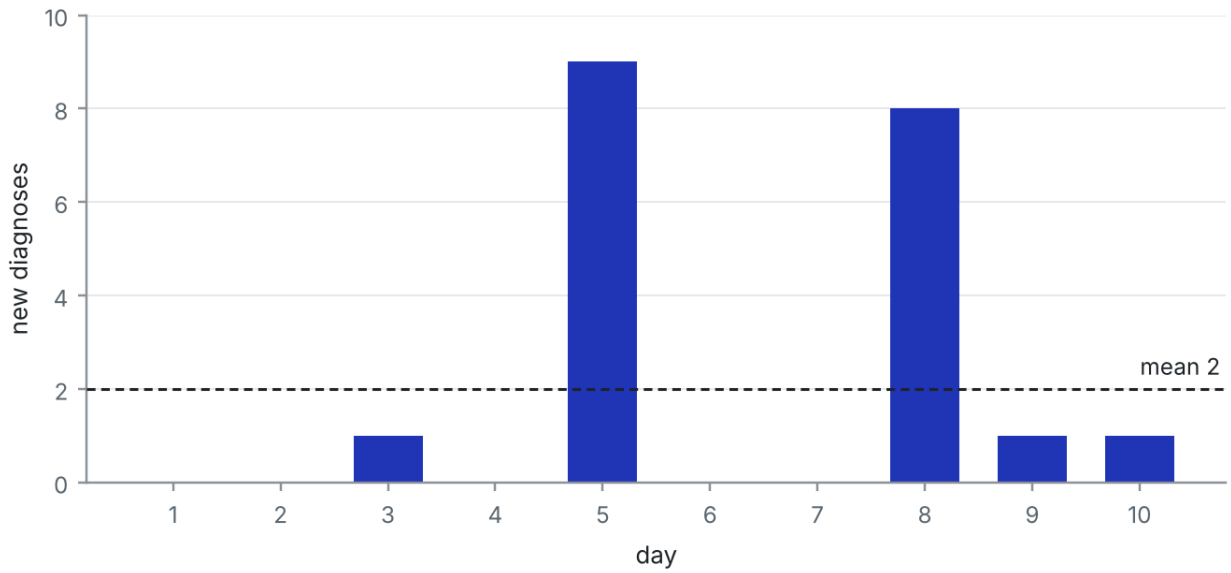
The Poisson distribution gives the chance of each count, 0, 1, 2 and so on. It needs one number, the average count, λ. Its spread is set by the same number: the variance is λ too.



The Poisson distribution with λ = 4.4, OpenIntro's average daily heart-attack admissions in New York City. The most likely counts are 3 to 5. The bars at 9 and above together hold 0.0358 of the probability: about 13 days in a year of 365.



A binomial count with n = 100 and p = 0.02 (left bars) beside a Poisson count with λ = 2 (right bars). The two never differ by more than 0.003: for rare events among many trials, only the product np matters.



Ten days of made-up counts of new diabetes diagnoses at a primary health centre. The mean is 2 a day, but the variance is 12, six times the Poisson's 2. If days 5 and 8, with 9 and 8, were camp days, the rate was not steady.

Illustration 1. Start with a remark you will hear. A cardiologist looks at one day's admissions and says: "Nine heart attacks today. We average about four. Something is going on."

OpenIntro *Statistics* (https://www.openintro.org/book/os/), §4.5, gives the numbers for a real setting. About 8 million people live in New York City. Admissions for acute myocardial infarction, a heart attack, average about 4.4 a day by historical records. Take that estimate as λ, the average count for one day. Work out the chance of each count, starting from 0.

```
P(X = 0) = e^(-4.4) = 0.0123
```

Each later chance is the one before times λ ÷ k. That is the formula: going from k − 1 to k multiplies in one more λ and divides by one more k.

```
P(X = 1) = 0.0123 times 4.4 = 0.054
P(X = 2) = 0.054 times 4.4 divided by 2 = 0.1188
```

Carry on the same way, keeping full precision in the calculator.

```
count k P(X = k)
0    0.0123
1    0.0540
2    0.1188
3    0.1743
4    0.1917
5    0.1687
6    0.1237
7    0.0778
8    0.0428
9    0.0209
10    0.0092
11    0.0037
12    0.0013
```

The counts 0 to 8 add up to 0.9642. So the chance of 9 or more on a given day is 1 − 0.9642, which is 0.0358. That is rare for one day. Now count the days in a year.

```
1 minus 0.9642 = 0.0358
365 times 0.0358 = 13.07
```

A city with a perfectly steady rate of 4.4 expects about 13 days a year with nine or more heart-attack admissions. Nine on one day is not a signal on its own. Two such days in a row are rarer. A year has 364 pairs of days next to each other.

```
0.0358 times 0.0358 = 0.00128
364 times 0.00128 = 0.47
```

So expect one such pair about every two years. A run longer or more frequent than that is the thing to look into.

Check the spread against the data. OpenIntro's year of 365 daily counts has a mean of 4.38 and a standard deviation of about 2. The Poisson standard deviation is √4.4, which is 2.10. The data spread about as much as the model says they should.

Figures quoted in this illustration are from [27].

WHERE THIS PICTURE BREAKS. OpenIntro says its 365 daily counts are simulated, and that in practice "we should check for an association between successive days". The 4.4 is a real historical average for one city. Heat waves, festivals and weekends can change the rate. Then a single λ for every day is wrong, and the 13 days a year is wrong with it.

The 0.0358 is the chance for one day picked before you look. It is not the chance that "something is going on" on a day you noticed because it was high.

Illustration 2. Where does the formula come from? Build it from the binomial of section 12. The numbers here are bare: an average of 2 events in an interval.

Cut the interval into n short slices. Make them short enough that two events in one slice never happen. Then each slice is a trial: one event, a success, or none. The average is 2 events across n

slices, so each slice has p = 2 ÷ n. Independence across slices is the second condition of the process. So the count is binomial, with n trials and p = 2 ÷ n.

Start with no events at all: every slice fails.

```
P(X = 0) = (1 - 2 ÷ n)^n
```

Watch it as the slices get thinner.

```
slices n P(X = 0)
10    0.1074
100    0.1326
1000    0.1351
10000    0.1353
```

The values settle at 0.1353, which is $e^{(-2)}$. For any average λ, $(1 - λ ÷ n)^n$ settles at $e^{(-λ)}$ as n grows. You can watch it happen on a calculator; the proof is beyond this book.

Now the step from one count to the next. For the binomial, divide the chance of k successes by the chance of k − 1, both from the formula of section 12. Take the two pieces in turn.

The counting pieces first. Write k! as k × (k − 1)!, and (n − k + 1)! as (n − k + 1) × (n − k)!. Then everything cancels except two factors.

```
(n choose k) ÷ (n choose (k - 1)) = [n! ÷ (k! (n - k)!)] × [(k - 1)! (n - k + 1)! ÷
n!] = (n - k + 1) ÷ k
```

The powers next. The top has one more p and one fewer (1 − p) than the bottom. So the ratio of the powers is p ÷ (1 − p). Together:

```
P(X = k) ÷ P(X = k - 1) = [(n - k + 1) ÷ k] × [p ÷ (1 - p)]
```

Put p = λ ÷ n. When n is huge, n − k + 1 is almost n, and 1 − p is almost 1. So the ratio is almost n ÷ k × λ ÷ n, which is λ ÷ k. Starting from $e^{(-λ)}$ and multiplying by λ ÷ 1, then λ ÷ 2, up to λ ÷ k gives the formula.

```
P(X = k) = e^(-λ) × (λ ÷ 1) × (λ ÷ 2) × … × (λ ÷ k) = e^(-λ) × λ^k ÷ k!
```

The mean and the variance come the same way. The binomial mean is np, which is λ for every n. The binomial variance is np(1 − p), which is λ × (1 − p). As p shrinks, 1 − p goes to 1, so the variance goes to λ.

How close is a real binomial? Take n = 100 and p = 0.02, so np = 2.

```
k binomial, n = 100 and p = 0.02 Poisson, λ = 2
0 0.1326    0.1353
1 0.2707    0.2707
2 0.2734    0.2707
3 0.1823    0.1804
4 0.0902    0.0902
5 0.0353    0.0361
6 0.0114    0.0120
7 0.0031    0.0034
```

No pair differs by more than 0.003. That is why a count of rare events among many people, such as new cases of a disease in a large district in a month, is usually treated as a Poisson count. Nobody needs to know n or p, only their product.

WHERE THIS PICTURE BREAKS. The match needs p small. With n = 10 and p = 0.2, np is still 2, but the binomial chance of no events is 0.8^{10}, about 0.107, against the Poisson's 0.135. And a binomial count can never exceed n; a Poisson count has no ceiling. The approximation is for rare events among many trials.

Illustration 3. Now a count that is not Poisson. A primary health centre logs new diagnoses of diabetes each day for ten days. The counts are made up for this example.

day	new diagnoses
1	0
2	0
3	1
4	0
5	9
6	0
7	0
8	8
9	1
10	1

The mean is 20 ÷ 10, which is 2 a day. Work out the sample variance as in Book 0, D5, dividing by n − 1. The squared deviations from 2 are 4, 4, 1, 4, 49, 4, 4, 36, 1 and 1.

```
4 plus 4 plus 1 plus 4 plus 49 plus 4 plus 4 plus 36 plus 1 plus 1 = 108
108 divided by 9 = 12
```

A Poisson count with mean 2 has variance 2. These counts have variance 12, six times as much. They are overdispersed.

Ask what happened on days 5 and 8. Suppose they were screening-camp days, when the centre tested many more people. Then the rate was not constant. It was low on ordinary days and high

on camp days. Two Poisson counts, one for each kind of day, may each fit. One Poisson count for every day does not.

Never fit a single Poisson to these data and call the camp days "unusual". With $\lambda = 2$, the chance of 8 or more on one day is about 0.0011. That number answers a question about a steady rate that the centre never had.

WHERE THIS PICTURE BREAKS. Ten days are few. A variance of 12 against a mean of 2 is a strong hint here only because the two big days stand far apart from the rest, and a reason for them is known. With ten counts that differ less, the sample variance itself scatters widely, and a gap between it and the mean proves little. Methods that model the extra spread are taught with regression models in a later book of the course.

MUST KNOW POINTS FOR YOU.

- A steady Poisson process throws up high days by chance. Before you call a day unusual, work out how often such a day comes by chance across all the days you looked at.
- Before you use a Poisson model, set the variance of the counts beside their mean. If the variance is well above the mean, the events cluster or the rate changes.
- The Poisson model holds only while the rate is steady and the events are independent. When the counts spread well beyond their mean, split the intervals by what changes the rate, as with the camp days, or leave it to a model for extra spread.
- λ belongs to an interval. An average of 12 an hour is $\lambda = 2$ for ten minutes and $\lambda = 288$ for a day. Rescale λ to the interval you are asking about before you use the formula.
- Look at many places, or many months, and one of them will show a "cluster" by chance. If one place has chance q, the chance that at least one of k independent places shows it is 1 minus $(1 - q)^k$, the no-success chance of section 13.
- The standard deviation of a Poisson count is $\sqrt{\lambda}$. So a count of 4 can easily move by 2. Do not read a jump from 2 to 4 cases as "doubling" without that spread in mind.

COMMON MISREADING. "λ is 4.4, the city's historical average." The 4.4 is an average of past days. It estimates λ, as OpenIntro's simulated year gives 4.38. λ itself is a parameter of a model, the steady rate the model assumes. Treat any λ you read as an estimate, and ask over how many days it was counted and whether the rate was steady across them.

EXERCISE 1 (retrieval). From memory: what three conditions make a count of events Poisson? Write the formula for $P(X = k)$. What are its mean and its variance? What does it mean when a set of counts has a variance well above its mean?

EXERCISE 2 (teaching). For a journalist, three minutes. A district reports four new childhood cancers in one block in a year. A block of that size averages about one a year. The numbers are

made up. The reporter asks you whether this is a cluster. Say what you would want to know before answering, and give one sentence they could quote.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. X has a Poisson distribution with $\lambda = 3$.

Work out $P(X = 0)$ and $P(X = 2)$.

2. X has a Poisson distribution with $\lambda = 3$.

Work out $P(X \leq 2)$ and $P(X \geq 3)$.

3. Events happen at an average of 12 an hour, at a steady rate. Work out λ for a 10-minute interval, and the chance of no events in 10 minutes.

4. For a Poisson count with $\lambda = 2.5$, work out $P(X = 4)$ divided by $P(X = 3)$, without working out either chance.

Then, for a different Poisson count, $P(X = 0)$ is 0.05. Work out λ.

5. OpenIntro gives about 4.4 heart-attack admissions a day in New York City. Treat the daily count as Poisson with $\lambda = 4.4$. Work out the chance of a day with no admissions, and how many such days a year to expect.

6. OpenStax *Introductory Statistics* 2e, §4.6, uses a reported 1.02 per cent chance of low seismic activity in a 48-hour period, over the next 200 days. Treat the count of days with low activity as binomial with $n = 200$ and $p = 0.0102$.

Work out the Poisson λ that approximates it. Then work out the chance of no such day both ways, binomial and Poisson. OpenStax uses the 48-hour chance as each day's chance; do the same.

7. Here is a worked answer. Find the step that broke.

```
Leah receives about 6 calls between 8 a.m. and 10 a.m.
question: the chance of more than one call in the next 15 minutes
λ = 6
P(X > 1) = 1 - P(X = 0) - P(X = 1)
= 1 - e^(-6) - 6 × e^(-6) = 1 - 0.0025 - 0.0149 = 0.9826
```

8. Here is a worked answer. Find the step that broke. The counts are made up.

```
new diagnoses a day for ten days: 0, 0, 1, 0, 9, 0, 0, 8, 1, 1
mean = 2 a day, so model the daily count as Poisson with λ = 2
P(X ≥ 8) at λ = 2 is about 0.0011
so days 5 and 8 are extraordinary events that need explaining
```

9. A district officer writes: "Three children were newly diagnosed with type 1 diabetes in one block last month. That block averages 0.5 a month. This is a cluster and needs a field investigation."

The district has 30 blocks of about the same size, each averaging 0.5 a month. All the numbers are made up. Decide what to compute, compute it, and say what your answer does not establish.

10. The clinic has averaged 4 new hypertension diagnoses a month over the past three years. This month there were 8. The clinic newsletter says: "New hypertension diagnoses doubled. Hypertension is rising fast in our area." The numbers are made up.

Decide what to compute, compute it, and say what your answer does not establish.

15 · The normal distribution: what generates it, and areas under it

DEFINITION. The normal distribution is a family of continuous distributions (Book 2, section 13) with two parameters: the mean μ, the Greek letter mu, and the standard deviation σ, the Greek letter sigma. It is written N(μ, σ). Its probability density is a symmetric bell centred on μ:

```
height at x = [1 ÷ (σ × √(2π))] × e^(-(x - μ)² ÷ (2σ²))
```

Here π, the Greek letter pi, is 3.14159, and e is the number of Book 2, section 4. You will not work out areas from this formula.

The z-score of a value x is its distance from the mean in standard deviations:

```
z = (x - μ) ÷ σ
```

If X is N(μ, σ), then Z = (X − μ) ÷ σ is N(0, 1), the standard normal distribution. Subtracting μ slides the bell to centre on 0, and dividing by σ rescales its width to 1; neither changes the shape. So one table of areas, for N(0, 1), serves every normal distribution. In a spreadsheet, NORM.S.DIST(z, TRUE) gives the area to the left of z, and NORM.S.INV(share) gives the z-score with that share to its left. This book states every area it uses.

The p-th percentile of a distribution is the value with p per cent of the distribution below it; the median is the 50th.

About 68 per cent of the area lies within one standard deviation of the mean, about 95 per cent within two, and about 99.7 per cent within three. 95 per cent lies within 1.96 standard deviations (to two decimals).

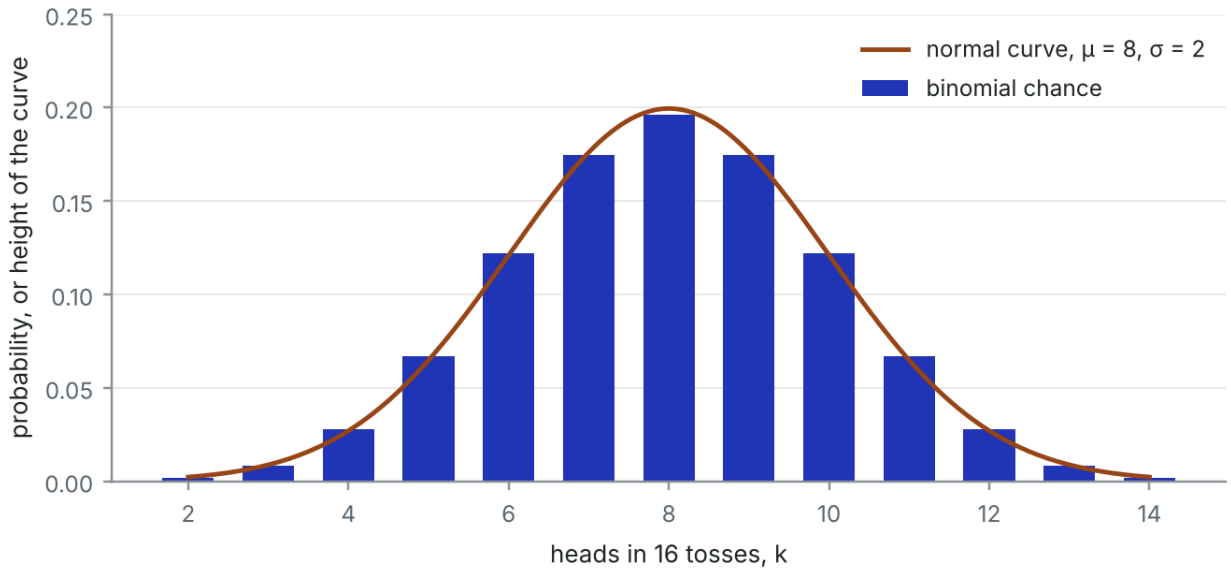
A binomial count with n trials and probability p is close to N(np, √[np(1 − p)]) when np and n(1 − p) are both at least 10. For a range of whole counts, widen the range by 0.5 at each end before taking the area, because the bar for the count k stands over the interval from k − 0.5 to k + 0.5.

No measured quantity is exactly normal. A normal probability plot shows how far a data set departs from normal. Sort the n values from smallest to largest. Give the i-th sorted value the share (i − 0.5) ÷ n, and plot it against the z-score that has that share of a normal distribution's area to its left. Near-normal data fall close to a straight line.

On a histogram, near-normal data make one hump, roughly symmetric, with tails thinning at the same rate on both sides; a long right tail, or a mean well above the median, says right skew. With few values a histogram is bumpy, and its shape depends on the bar width, so read it with the normal probability plot, not alone. [131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143]

IN PLAIN TERMS. Add up many small pieces that each push a little, up or down, independently. The total lands near the middle most of the time and far out only rarely. Its distribution is a bell, the same on both sides. That bell is the normal distribution.

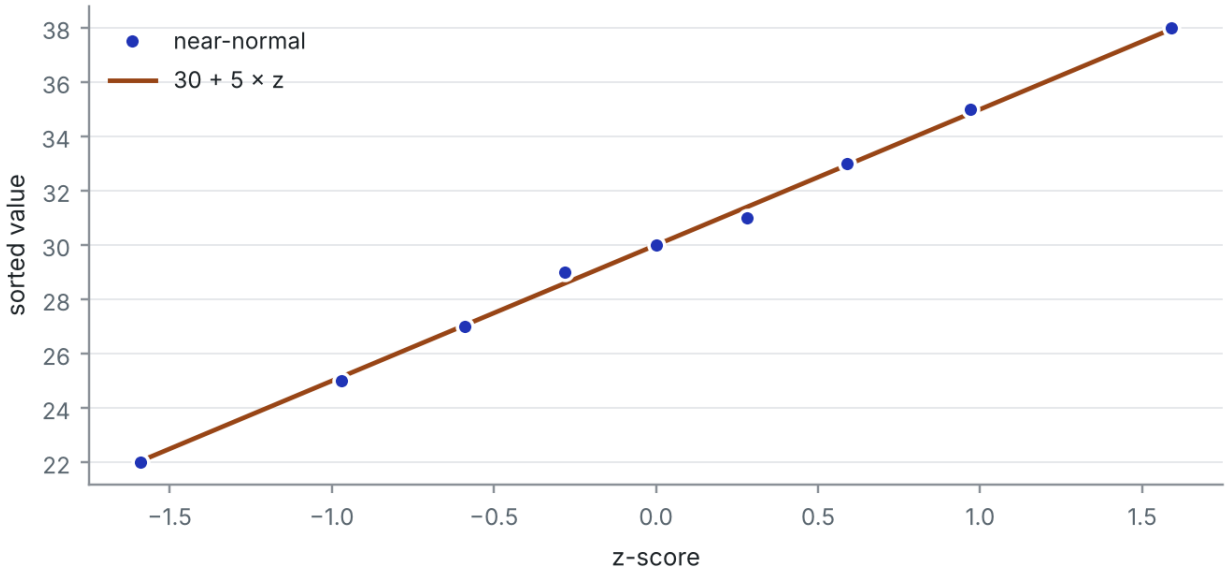
Two numbers fix it: where the middle is, the mean, and how wide the bell is, the standard deviation.



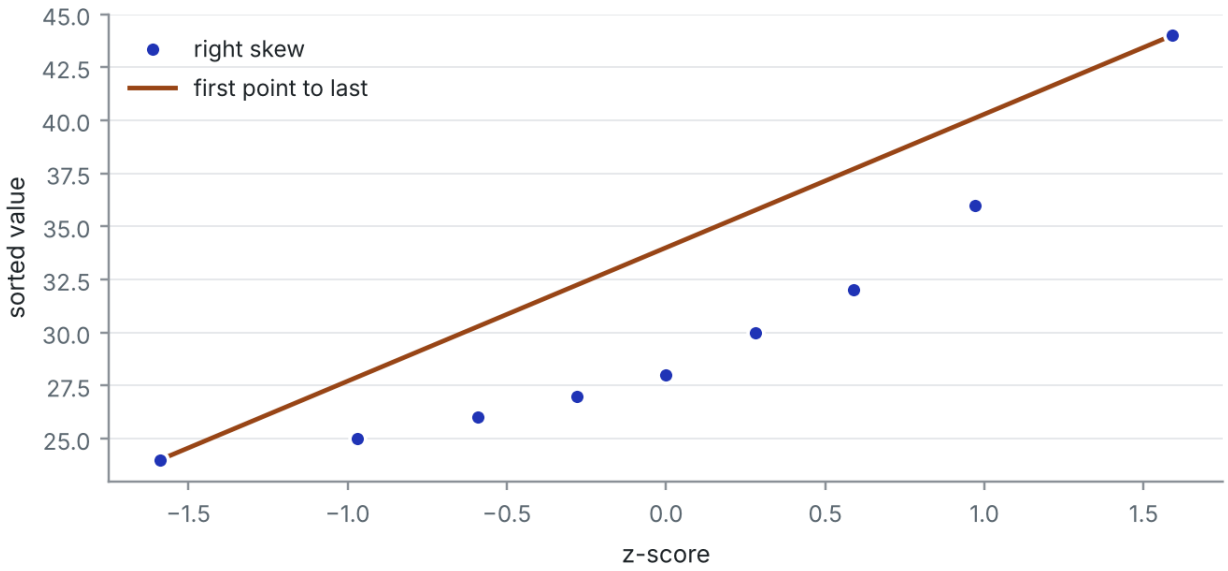
Heads in 16 tosses of a fair coin, a sum of 16 small independent pieces (bars), beside the normal curve with mean 8 and standard deviation 2 (line). They agree to about 0.003 at every count: the height at 8 is 0.1964 for the bars and 0.1995 for the curve.



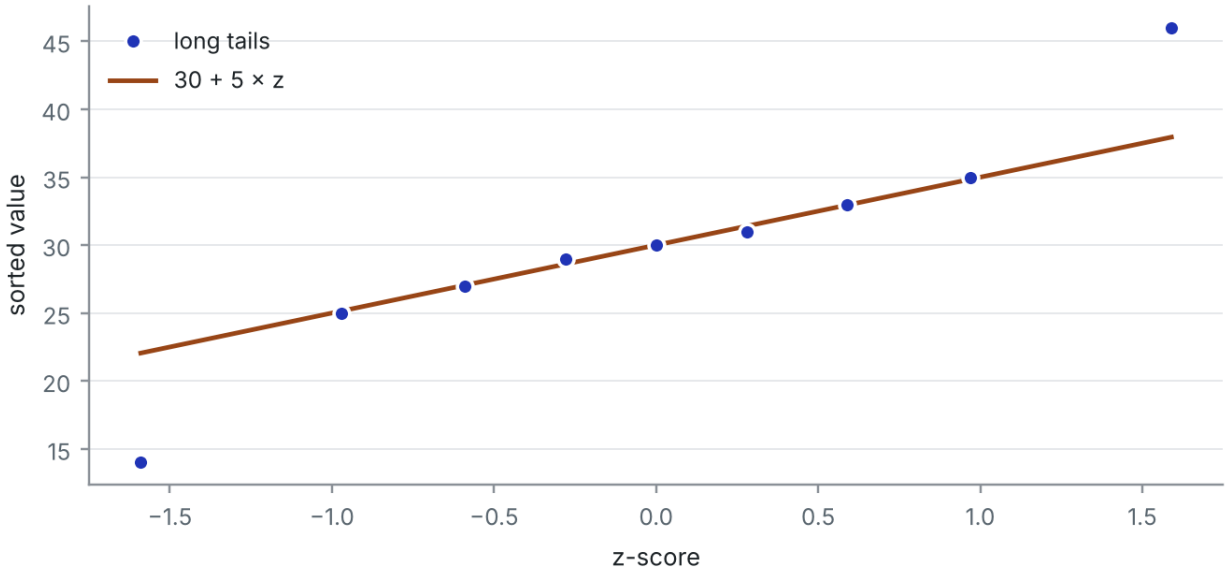
Body fat in Kiran 2022's 131 men. Left bars: the lower quartile, median and upper quartile a normal distribution with mean 22.1 and standard deviation 7.3 would give (17.18, 22.1, 27.02). Right bars: what the paper reports (16.7, 20.3, 28.8). The reported upper quartile sits 8.5 above the median, the lower only 3.6 below: a long right tail.



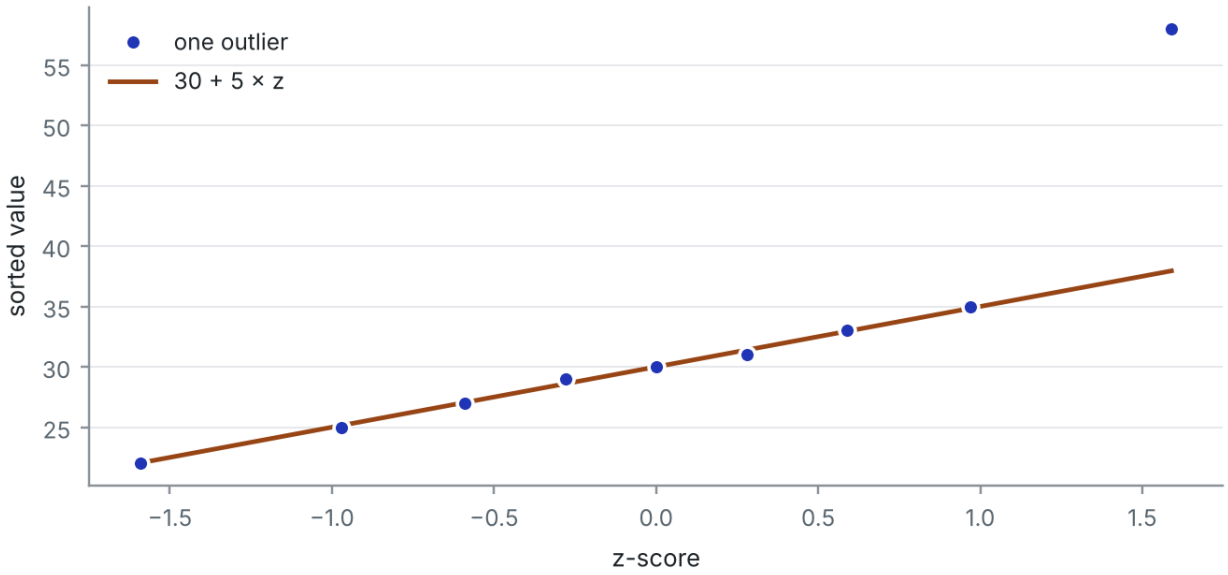
A normal probability plot of the made-up near-normal set: nine sorted values against their z-scores. The points sit close to the line 30 + 5 × z.



Right skew on a normal probability plot. The line runs from the first point (24 at z = −1.59) to the last (44 at z = 1.59). Every point in between lies below it, the signature NIST names for a right-skewed data set.



Long tails on a normal probability plot. The middle seven points sit on the line 30 + 5 × z; the lowest, 14, falls far below it and the highest, 46, rises far above it.



One outlier on a normal probability plot. Eight points sit on the line 30 + 5 × z; one value, 58, stands far above the end of it.

Illustration 1. Watch a bell build up from small pieces. Toss a fair coin 16 times and count the heads. Each toss adds 1 or 0, with equal chance, independently of the rest. The count is binomial (section 12), with n = 16 and p = 0.5.

```
16 times 0.5 = 8
16 times 0.5 times 0.5 = 4
```

The mean is 8 and the variance is 4, so the standard deviation is √4, which is 2. Now set the binomial chances beside the height of the normal curve with μ = 8 and σ = 2.

```
heads k binomial chance normal height, μ = 8 and σ = 2
2    0.0018    0.0022
3    0.0085    0.0088
4    0.0278    0.0270
5    0.0667    0.0648
6    0.1222    0.1210
7    0.1746    0.1760
8    0.1964    0.1995
9    0.1746    0.1760
10    0.1222    0.1210
11    0.0667    0.0648
12    0.0278    0.0270
13    0.0085    0.0088
14    0.0018    0.0022
```

Work out the height at the centre, k = 8, where x − μ is 0 and the power of e is 0.

```
2 times 3.14159 = 6.28318
√6.28318 = 2.5066
2 times 2.5066 = 5.0132
1 divided by 5.0132 = 0.1995
```

The bars and the curve agree to about 0.003 everywhere. Sixteen pieces is not many. The more pieces, and the smaller each one, the closer the match.

Here is the point for health data. OpenIntro notes that the heights of adult men in the United States closely follow the normal distribution. The usual explanation is that height sums many small contributions: many genes, many years of food and illness. So adult height within one sex is close to normal. A quantity driven by one big factor, or by pieces that multiply, is not.

WHERE THIS PICTURE BREAKS. The coin's pieces are identical and equal in size. Real contributions differ in size, and the total is near-normal only if none of them dominates. One big factor, such as sex for height, gives two overlapping bells, one for men and one for women, not one.

The table compares a chance with a height of a density. They can be set side by side here only because the counts are one unit apart, so each bar has width 1.

Illustration 2. Now use the bell to get an area. Take the National Family Health Survey's figure (NFHS-6, 2023-24): 30.7% of women aged 15 to 49 in India had overweight or obesity. Pick 100 women at random, and count those with overweight or obesity. How likely is a count of 21 or fewer?

The count is binomial with n = 100 and p = 0.307. Check the condition first.

```
100 times 0.307 = 30.7
100 times 0.693 = 69.3
```

Both are at least 10, so the normal approximation may be used. Its mean is 30.7. Its standard deviation is the square root of np(1 − p).

```
100 times 0.307 times 0.693 = 21.2751
√21.2751 = 4.612
```

"21 or fewer" counts the whole numbers up to 21. Widen by a half: take the area below 21.5. Turn 21.5 into a z-score.

```
21.5 minus 30.7 = -9.2
-9.2 divided by 4.612 = -1.99
```

The area to the left of z = −1.99 is 0.0233, from software or a table. So about 2 per cent of such samples would hold 21 or fewer women with overweight or obesity. The exact binomial answer, adding the 22 chances from 0 to 21, is 0.0204. The bell is close but a little high here: with np = 30.7, the binomial is still slightly lopsided.

Now use the 68-95-99.7 rule, with two standard deviations either side of 30.7. They run from 30.7 − 9.22 to 30.7 + 9.22.

```
2 times 4.612 = 9.22
30.7 minus 9.22 = 21.48
30.7 plus 9.22 = 39.92
```

About 95 per cent of random samples of 100 would hold between 22 and 39 such women.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The binomial needs independent picks. A survey that samples whole households or villages picks women who are alike, so its counts spread more than this (section 12), and the 95 per cent range is too narrow for it.

The 30.7 per cent is NFHS-6's estimate for all of India. A district's share can be quite different, and this calculation says nothing about any district.

Illustration 3. Now a measured quantity that is not normal. Open Kiran, Harshitha and Bhargava's study of medical students in Mangaluru (https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/) and find Table 1. For the 131 men, body fat had a mean of 22.1 per cent and a standard deviation of 7.3. The median was 20.3, and the quartiles were 16.7 and 28.8.

Suppose you take the mean and standard deviation and treat body fat as N(22.1, 7.3). That is the move to test. A normal distribution puts its median at its mean, and its quartiles at z = −0.674 and z = 0.674, the points with a quarter of the area beyond them.

```
0.674 times 7.3 = 4.92
22.1 minus 4.92 = 17.18
22.1 plus 4.92 = 27.02
```

The normal model says: median 22.1, quartiles about 17.2 and 27.0, each 4.92 from the middle. The paper says: median 20.3, quartiles 16.7 and 28.8.

```
20.3 minus 16.7 = 3.6
28.8 minus 20.3 = 8.5
```

The upper half of the middle spreads 8.5 points, the lower half only 3.6. The mean sits 1.8 above the median. Body fat in these men has a long right tail. It is not normal.

See what the normal model does to a cut-off. It puts 2.5 per cent of men below 22.1 − 1.96 × 7.3, which is 7.8 per cent body fat. With a long right tail and a short left one, that lower figure is misplaced, and so is the upper one, 22.1 plus 1.96 times 7.3 = 36.4. A "normal range" of mean plus or minus two standard deviations, worked out on a skewed quantity, is wrong at both ends.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The study is 131 medical students at one college, not men in general. The skew is a fact about these numbers. It shows that body fat need not be normal, not how body fat is distributed in Indian men.

The table gives a mean, a standard deviation and three quartiles. You cannot draw the shape from five numbers, only see that it is lopsided.

Illustration 4. A table of summaries hides the shape. With the raw values in hand, draw a normal probability plot. Here is how, by hand, for 9 values.

Sort the values from smallest to largest. Give the i-th value the share (i − 0.5) ÷ 9. That is the share of a normal distribution's area that should lie to its left. Turn each share into a z-score with NORM.S.INV. For the first value:

```
1 minus 0.5 = 0.5
0.5 divided by 9 = 0.0556
```

The z-score with 0.0556 of the area to its left is −1.59. The nine z-scores run −1.59, −0.97, −0.59, −0.28, 0.00, 0.28, 0.59, 0.97 and 1.59. The US National Institute of Standards and Technology's *Engineering Statistics Handbook* (NIST) uses a slightly different rule for the shares; the plot looks almost the same.

Plot each sorted value, up the side, against its z-score, along the bottom. If the data are normal, each value is about μ + z × σ. So the points fall near a straight line, with height μ where z is 0 and slope σ. Four made-up data sets show the four patterns you will meet.

z	near-normal	right skew	long tails	one outlier
−1.59	22	24	14	22
−0.97	25	25	25	25
−0.59	27	26	27	27
−0.28	29	27	29	29
0.00	30	28	30	30
0.28	31	30	31	31
0.59	33	32	33	33
0.97	35	36	35	35
1.59	38	44	46	58

Near-normal. The points sit on the line 30 + 5 × z, with height 30 at z = 0 and slope 5. The set's mean is 30.

Right skew. Draw the line from the first point to the last: from 24 at z = −1.59 to 44 at z = 1.59. Its slope is 20 divided by 3.18, which is 6.29, and it crosses z = 0 at 34. Every point in between lies below it, so the points bend upward. NIST calls this pattern "the signature of a significantly right-skewed data set". The values bunch low and stretch out high: 28 is the median, the mean is 30.2.

Left skew is the mirror: every point lies above the line from the first point to the last.

Long tails. The middle seven points sit on the line 30 + 5 × z. The lowest point, 14, lies far below it and the highest, 46, far above. Both ends peel away, in opposite directions. The set is symmetric, with mean 30 and median 30, but it has more extreme values than a normal set. Short tails peel the other way: the lowest point above the line, the highest below.

One outlier. Eight points sit on the line 30 + 5 × z, and one, 58, stands far above the end. The rest of the data are normal-looking. That one value needs checking: a mistyped 38, or a real person who is not like the others.

WHERE THIS PICTURE BREAKS. The four sets are made up, and nine points is few. Even data drawn from a true normal distribution wobble about the line, most at the ends, and with ten or twenty points the wobble can look like a pattern. Reading the plot is a judgement, not a test. OpenIntro puts it plainly: "There is no perfect way to check the normality condition".

The plot tests one shape, the normal. A straight line does not show the values are independent, or that they came from the population you meant.

MUST KNOW POINTS FOR YOU.

- "Two standard deviations above the mean means abnormal" is the error. Rare in a distribution is not the same as disease. A clinical cut-off is set from outcomes, not from a spread.
- Before you use a z-score, a percentile or a mean plus or minus two standard deviations, look at the data's shape. Compare the mean with the median, and the two quartiles with the median.
- Read a normal probability plot by its shape. Points bending below the line from the first point to the last mean right skew; above it, left skew. A single point off the end is an outlier. Say "no clear departure", never "the data are normal".
- Use the normal approximation to a binomial count only when np and n(1 − p) are both at least 10.
- For 95 per cent, the normal multiplier is 1.96, not 2. Two standard deviations hold 95.45 per cent. The difference is small.
- Before you assume a quantity is near-normal, ask what builds it. A sum of many small independent pieces gives a bell; one dominant factor (sex, for height) gives two bells, and pieces that multiply give a right tail (section 16).

COMMON MISREADING. "The data passed a normality check, so they are normal." No data set is exactly normal. A straight-enough plot, or a test that did not reject, says only that this sample showed no clear departure. With few points, large departures go unseen; with many, trivial ones

get flagged. So ask a different question. Is the departure big enough to change the answer you are about to work out, such as a share in a tail?

EXERCISE 1 (retrieval). From memory: what generates a near-normal distribution? Write the z-score. What share of a normal distribution lies within one, two and three standard deviations of the mean, and what multiplier gives 95 per cent? What does a normal probability plot put on each axis?

EXERCISE 2 (interpretation). A thesis draft has a normal probability plot of waist circumference for 160 women. Here is what the text says.

> *"Most points lie on the line, but the top ten points rise further and further above it, while the bottom points stay on the line. The data are therefore normally distributed, and we report mean ± SD."*

Say what the plot does show, what the authors should have done, and what sentence they should write instead.

PRACTICE. Fifteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. X is N(120, 8). Work out the z-score of x = 130.

2. The area to the left of z = 1.25 under the standard normal curve is 0.8944. Work out the area to the right of 1.25, the area to the left of −1.25, and the area between −1.25 and 1.25.

3. X is N(50, 6). Between what values do about 68, 95 and 99.7 per cent of the values lie?

4. X is N(100, 15). The z-score with 0.90 of the area to its left is 1.2816. Find the 90th percentile of X.

5. The area to the left of z = 2 is 0.9772 and to the left of z = −1 is 0.1587. Work out the chance that a standard normal value lies between −1 and 2.

6. Five values: 17, 12, 23, 15, 18. Set them up for a normal probability plot, using the share (i − 0.5) ÷ 5 for the i-th sorted value. The z-scores for the shares 0.1, 0.3, 0.5, 0.7 and 0.9 are −1.28, −0.52, 0.00, 0.52 and 1.28. List the five points.

7. OpenStax *Introductory Statistics* 2e, §6.1, gives the heights of 15-to-18-year-old males in Chile in 2009-10 as N(170, 6.28), in centimetres. Work out the z-score of a height of 168 cm, and the height with a z-score of 1.27.

8. OpenIntro *Statistics*, §4.1.4, models the heights of adult men in the United States as N(70.0, 3.3), in inches, from a sample of 100 men. The area to the left of z = 1.21 is 0.8869. Work out the chance that a man picked at random is at least 74 inches tall.

9. Pick 200 women aged 15 to 49 at random in India, with NFHS-6's 30.7% as the chance that each has overweight or obesity. Use the normal approximation to work out the chance that 75 or more of them do. The area to the left of z = 2.01 is 0.9778.

10. In Kiran 2022, Table 1, the 131 men's body fat had mean 22.1 per cent, standard deviation 7.3, and quartiles 16.7 and 28.8. If body fat were N(22.1, 7.3), what share of men would lie below 16.7,

and what share above 28.8? The area to the left of z = −0.74 is 0.2296 and to the left of z = 0.92 is 0.8212. Compare with what the quartiles say.

11. Here is a worked answer. Find the step that broke.

```
exam scores are N(63, 5); find the chance a student scored more than 65
z = (65 - 63) ÷ 5 = 0.4
the area to the left of z = 0.4 is 0.6554
so the chance of scoring more than 65 is 0.6554
```

12. Here is a worked answer. Find the step that broke.

```
a rare finding: 20 patients, each with chance 0.05, independently
count X is binomial; use the normal approximation
mean = 20 × 0.05 = 1; SD = √(20 × 0.05 × 0.95) = 0.975
P(X = 0): area below 0.5, z = (0.5 - 1) ÷ 0.975 = -0.51, area 0.305
so the chance of no patient with the finding is about 0.31
```

13. Here is a worked answer. Find the step that broke.

```
nine sorted values: 16, 24, 28, 30, 32, 33, 34, 35, 36
their z-scores: -1.59, -0.97, -0.59, -0.28, 0.00, 0.28, 0.59, 0.97, 1.59
the line from the first point to the last: 26 + 6.29 × z
the middle points all lie above that line, and the lowest point stands far below the
rest
a bend away from a straight line is right skew, so the values have a long right tail
```

14. A hostel medical officer writes: “In our students, BMI averages 24.2 with an SD of 3.9. Any student above two SDs, that is above 32.0, has abnormal BMI and should be referred. That will be about 2.5 per cent of students.”

The 24.2 and 3.9 are the men’s figures from Kiran 2022, Table 1, used here as if they were the hostel’s. Decide what to compute, compute it, and say what the answer does not establish.

15. A district officer says: “NFHS-6 found 30.7% of women aged 15 to 49 overweight or obese. Our district survey of 400 women found only 104. Fewer than 105 out of 400 would almost never happen by chance, so our district is genuinely better.”

Decide what to compute, compute it, and say what your answer does not establish. The area to the left of z = −1.98 is 0.0239.

16 · Skewed quantities and the log scale: when effects multiply

DEFINITION. A positive random variable X is log-normal when ln X, its natural logarithm (Book 2, section 4), has a normal distribution (section 15). A log-normal quantity can take only values above zero, and its distribution has a long right tail.

Where it comes from. Multiplicative effects give a log-normal shape: many small independent factors, multiplied.

Three numbers describe a log-normal quantity on its own scale.

- The median of X is e^{μ}, where μ, the Greek letter mu, is the mean of ln X.
- The multiplicative standard deviation is e^{σ}, where σ, the Greek letter sigma, is the standard deviation of ln X. It is a factor, with no unit, and it is always above 1.
- About 68.3 per cent of values lie between the median divided by the multiplicative standard deviation and the median multiplied by it. About 95.5 per cent lie between the median divided by the square of the multiplicative standard deviation and the median multiplied by that square.

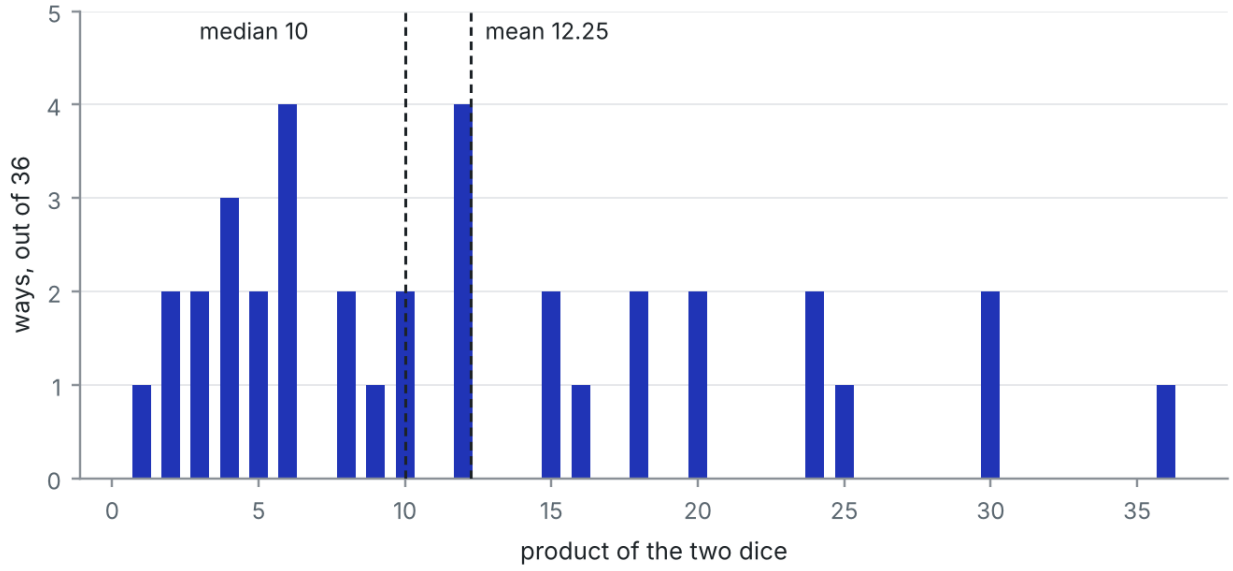
Here μ and σ are the mean and standard deviation of ln X, not of X as in section 15.

From a sample of n positive values x1 to xn, the geometric mean is the mean of the logarithms, back-transformed: take ln of each value, average them, and raise e to that average. It estimates the median of a log-normal quantity, not its mean. [144, 144, 145, 146, 147, 148, 149, 150, 151, 152]

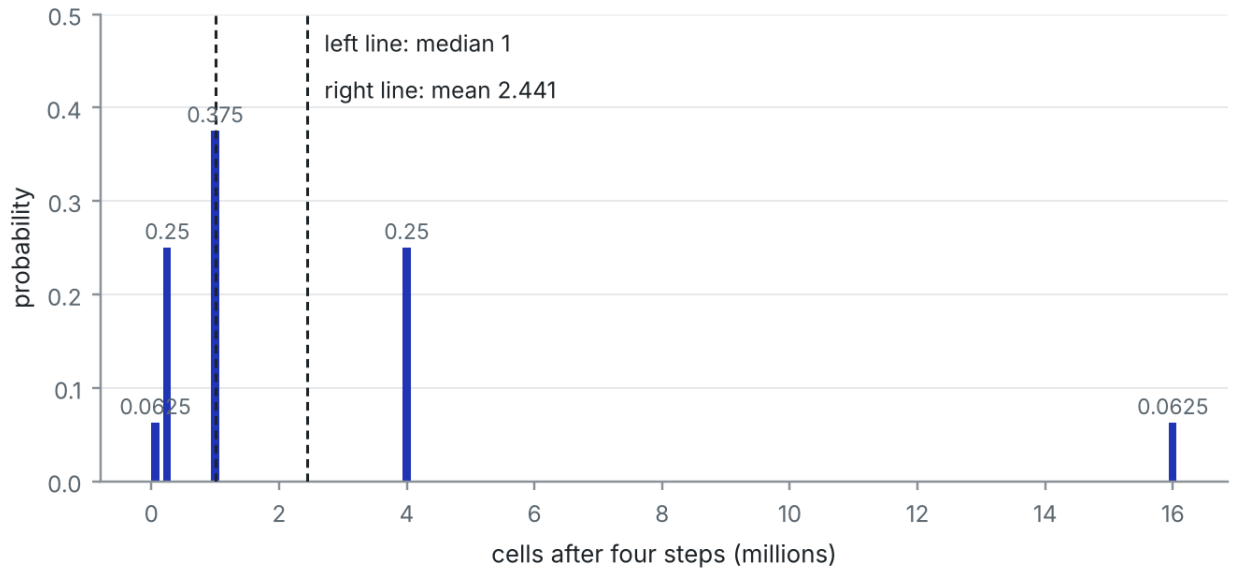
IN PLAIN TERMS. Some quantities grow by adding pieces.

Other quantities grow by multiplying. Each step multiplies what is already there. The result is lopsided: a crowd of modest values and a long tail of big ones.

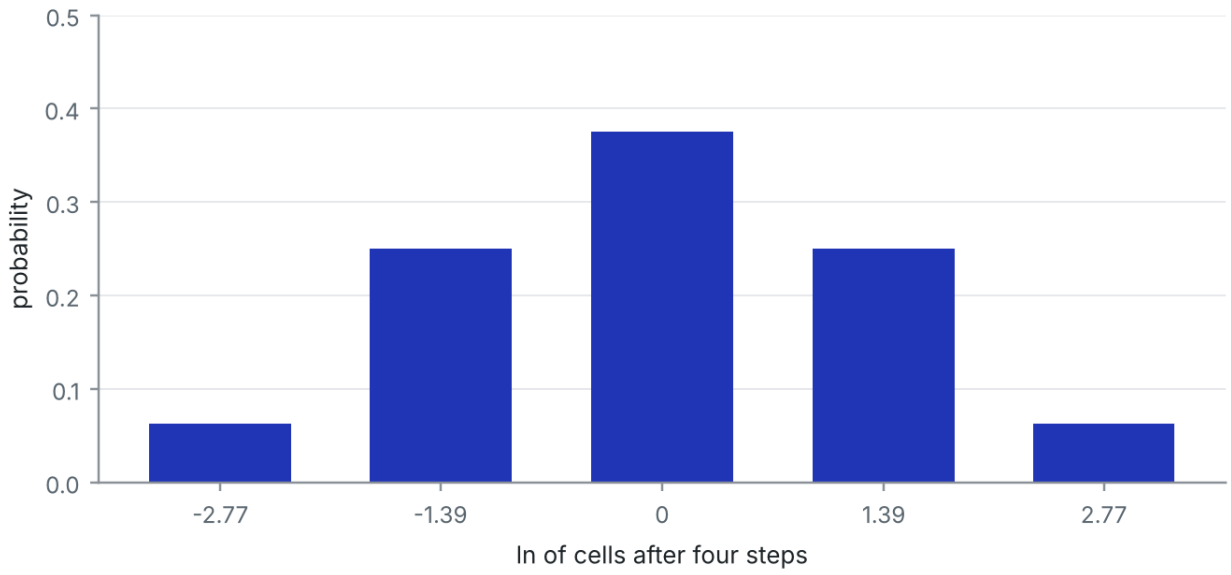
The fix is a logarithm. A logarithm turns multiplying into adding (Book 0, A7). So take the logarithm of every value. When you are done, turn the answer back with e to the power.



The 36 ways two dice can land, multiplied. The products bunch at the low end and thin out towards 36. The median product is 10 and the mean is 12.25: the few large products pull the mean up and leave the median where it is.



The made-up culture after four steps, each doubling or halving 1 million cells. On the original scale the five possible counts, 0.0625 to 16 million, bunch low with one far out at 16: a long right tail. The first three bars, at 0.0625, 0.25 and 1 million, stand at 0.0625, 0.25 and 0.375. The mean is about 2.44 million; the median is 1 million.



The same five counts on the log scale. Each doubling adds ln 2, 0.693, and each halving takes it away, so the logarithms sit at minus 2.77, minus 1.39, 0, 1.39 and 2.77, the same distance either side of 0. The skew has gone.

Illustration 1. Start with two ordinary dice, as Limpert, Stahel and Abbt do in "Log-normal distributions across the sciences: keys and clues" (*BioScience*, 2001), at https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf. List all 36 ways the two can land.

First add the two numbers. The totals run from 2 to 12. Count how many of the 36 ways give each total: one way gives 2, two give 3, and so on up to six ways for 7, then down again to one way for 12. The shape is the same on both sides of 7. The mean is 7 and the median is 7.

Now multiply the two numbers instead. The products run from 1 to 36, and they bunch at the low end.

product	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
ways	1	2	2	3	2	4	2	1	2	4	2	1	2	2	2	1	2	1

Find the median. Order the 36 products from smallest to largest. The 18th and 19th are both 10, so the median is 10. Now the mean. Multiply each product by its number of ways and add. The 36 products together come to 441. Each of the 36 ways is equally likely.

```
441 divided by 36 = 12.25
```

The mean, 12.25, sits above the median, 10. The few large products, 25, 30 and 36, pull the mean up. They leave the median where it is. The dice are the same dice. Only the way their effects combine changed.

Limpert and colleagues put it in one line.

> *Adding the two numbers, which is the principle of most games, leads to values from 2 to 12, with a mean of 7, and a symmetrical frequency distribution.*

> *Multiplying the two numbers, however, leads to values between 1 and 36 with a highly skewed distribution.*

Figures quoted in this illustration are from [153].

WHERE THIS PICTURE BREAKS. Two dice give neither a normal nor a log-normal distribution; the authors say so themselves. The dice show only the direction: adding gives a balanced shape, multiplying gives a long right tail. The near-normal shapes need many pieces, not two.

Illustration 2. Now watch the logarithm undo the skew. Limpert and colleagues describe a culture of about 10^6 bacteria, a million cells. One cell division more doubles it, to 2×10^6. One division fewer halves it, to 5×10^5. Search their paper for "one cell division more" to see the sentence.

Build a made-up model from that. Start with 1 million cells. In each of four time steps the count doubles or halves, each with probability one half, and each step is independent of the others. After four steps there are 16 equally likely paths, section 1's outcomes. Count them by the number of doublings: four doublings happen one way, three happen four ways, two happen six ways.

cells (millions)	ln of cells	paths	probability
0.0625	-2.77	1	0.0625
0.25	-1.39	4	0.25
1	0	6	0.375
4	1.39	4	0.25
16	2.77	1	0.0625

Look at the first column. Three of the five values sit below 2. The last one is 16. That is a long right tail.

Now take the natural logarithm of each count, the second column. Each doubling adds ln 2 to the logarithm, and each halving takes it away.

```
ln 2 = 0.693
4 times 0.693 = 2.77
```

So the logarithm of the final count is a sum of four pieces, each plus or minus 0.693. The second column is the same distance either side of 0, with the same probabilities. The skew has gone. With many more steps, that sum would be close to normal, by the reasoning of section 15, and the count would be close to log-normal.

The median count is 1 million: the middle path. Find the mean count: each value times its probability, added up, as in Book 2, section 14.

```
0.0625 times 0.0625 = 0.0039
0.25 times 0.25 = 0.0625
1 times 0.375 = 0.375
4 times 0.25 = 1
16 times 0.0625 = 1
0.0039 plus 0.0625 plus 0.375 plus 1 plus 1 = 2.441
```

The mean is about 2.44 million, more than twice the median. The single path to 16 million contributes as much to the mean as the four paths to 4 million. On the log scale the mean of the second column is 0, and e^0 is 1: the back-transformed mean of the logarithms is the median, 1 million, not the mean.

Figures quoted in this illustration are from [153].

WHERE THIS PICTURE BREAKS. The four-step model is made up, and its numbers hold only for it. Real cultures do not double or halve in lockstep. The point that survives is the arithmetic: repeated multiplying gives a right tail on the original scale and a balanced shape on the log scale.

Illustration 3. Now a real measured quantity. Kiran and colleagues measured percentage body fat by bioimpedance in 131 men and 151 women, medical students in Mangaluru. Their Table 1 is at https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/. Book 0, D5 used the men's figures to show a mean above a median. Read the median and the two quartiles for each sex.

group	first quartile (%)	median (%)	third quartile (%)	mean (%)
men	16.7	20.3	28.8	22.1
women	28.1	34.0	39.2	33.8

For the men, measure how far each quartile sits from the median, first in percentage points.

```
20.3 minus 16.7 = 3.6
28.8 minus 20.3 = 8.5
```

The upper quarter-band is more than twice as wide as the lower one. That is a right tail. Now ask the log-normal question: is the lopsidedness one of ratios? Divide instead of subtracting.

```
20.3 divided by 16.7 = 1.216
28.8 divided by 20.3 = 1.419
```

Compare like with like: take logs of the ratios, then divide the wider side by the narrower.

```
ln 1.216 = 0.196
ln 1.419 = 0.350
0.350 divided by 0.196 = 1.79
8.5 divided by 3.6 = 2.36
```

In logs the upper side is 1.79 times the lower; in percentage points, 2.36 times. The gap narrows. It does not close. Taking logs moves the men's body fat towards balance. It does not make it balanced.

Now the women.

```
34.0 minus 28.1 = 5.9
39.2 minus 34.0 = 5.2
34.0 divided by 28.1 = 1.210
39.2 divided by 34.0 = 1.153
ln 1.210 = 0.191
ln 1.153 = 0.142
0.191 divided by 0.142 = 1.35
5.9 divided by 5.2 = 1.13
```

The women's quartiles are nearly balanced as they stand, 5.9 below and 5.2 above. Their mean, 33.8, sits beside their median, 34.0. In logs the wider side is 1.35 times the narrower; in percentage points, 1.13 times. For the women, logs would make the shape worse.

So the log scale is not a cure to apply to every positive quantity. It is a claim that the effects behind a quantity multiply. Check that claim on the quartiles before you rely on it.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. Three numbers per group cannot show a whole shape. The quartile check flags the direction of a tail; it cannot prove a quantity is log-normal. The figures are for one college's medical students. They say nothing about body fat in other groups, and nothing about what body fat means for health.

Illustration 4. Here is why the choice matters wherever a range is reported. Limpert and colleagues quote a paper whose data were log-normal but were summarised as a mean and standard deviation, 4.1 ± 3.7. Read ± as "plus or minus".

Work out the usual "mean plus or minus two standard deviations".

```
3.7 times 2 = 7.4
4.1 minus 7.4 = -3.3
4.1 plus 7.4 = 11.5
```

A range from minus 3.3 to 11.5. The quantity can only be positive, so nearly half of the range below the mean (3.3 of its 7.4) is impossible. Even the plus-or-minus-one band misleads. In their words, $\bar{x} \pm s$ is the mean plus or minus one standard deviation.

> *the probability of the corresponding $\bar{x} \pm s$ interval (0.4 to 7.8) turns out to be 88.4% instead of 68.3%.*

The band that should hold about two thirds of values holds 88.4 per cent. Limpert and colleagues give the same data on the log scale as a geometric mean of 3.0 and a multiplicative standard deviation of 2.2. Divide and multiply.

```
3.0 divided by 2.2 = 1.36
3.0 times 2.2 = 6.6
```

About 68 per cent of the values lie between 1.36 and 6.6. For about 95.5 per cent, use the square of 2.2.

```
2.2 times 2.2 = 4.84
3.0 divided by 4.84 = 0.62
3.0 times 4.84 = 14.5
```

From 0.62 to 14.5. The range is lopsided around 3.0, as the data are, and it never goes below zero.

Figures quoted in this illustration are from [153].

WHERE THIS PICTURE BREAKS. The 88.4 and 68 per cent figures are Limpert's, for one data set that fitted a log-normal shape well. On a quantity that is not log-normal, the times-or-divided-by range is as wrong as the plus-or-minus one. A range holding 95.5 per cent of values is a description of a group. It is not, by itself, a line between healthy and unwell.

MUST KNOW POINTS FOR YOU.

- "A long right tail means something is wrong with the data, so drop the big values" is the error. When effects multiply, a long right tail is the expected shape. Keep them, and analyse on the log scale.
- Never set a range as mean plus or minus two standard deviations on a skewed positive quantity.
- To average a quantity whose effects multiply, take logs, average, and turn the answer back with e. The result is the geometric mean.
- Before you log a variable, take each quartile's distance from the median twice: in its own units, and as the logarithm of the ratio to the median. If the two log distances are no closer to equal than the two plain distances, logging will not help.
- When a journalist asks for the typical value of a skewed quantity, give the median and say it is the middle person.
- Teach the difference with two dice in a resident session. Add them, then multiply them, and tally both.

COMMON MISREADING. "Turn the mean of the logs back, and you have the mean." You have the geometric mean. For a log-normal quantity that is the median, the middle value, and it lies below the ordinary mean, sometimes far below. A paper that reports a back-transformed mean as

"the mean" has reported the middle and called it the average. Ask which one a table means before you compare it with another study.

REPORTING IT. Body fat was skewed to the right in men. We therefore report the median, 20.3 per cent, with first and third quartiles of 16.7 and 28.8, rather than the mean and standard deviation.

EXERCISE 1 (retrieval). From memory: what kind of effects give a normal shape, and what kind give a log-normal one? How do you work out a geometric mean, and which number of a log-normal quantity does it estimate?

EXERCISE 2 (interpretation). Table 1 of Kiran and colleagues gives the men's body fat as a mean of 22.1 per cent with a standard deviation of 7.3. Their median is 20.3 per cent, with quartiles 16.7 and 28.8. The women's mean is 33.8, with a standard deviation of 7.8. Their median is 34.0, with quartiles 28.1 and 39.2.

For each sex, say which summary you would quote, and why. Say what the table does not show.

EXERCISE 3 (teaching). A first-year resident has ten minutes and a whiteboard. Explain why a skewed quantity can become balanced on the log scale. Use the doubling-and-halving culture, and end with what the log scale cannot do.

PRACTICE. Eight problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Work out ln 8 and ln 0.125. Then find the number whose natural logarithm is 1.5.

2. Work out the geometric mean and the ordinary mean of 2, 8 and 32.

3. Seven values are 1, 2, 4, 8, 16, 32 and 64. Work out their median, their ordinary mean and their geometric mean.

4. Limpert and colleagues report the age at onset of Alzheimer's disease from one study. It can be described by a geometric mean of 60 years and a multiplicative standard deviation of 1.16.

Work out the range of ages expected to hold about 68 per cent of onsets, and the range for about 95.5 per cent.

5. Kiran and colleagues' Table 1 gives the men's waist circumference as a median of 83.5 cm with quartiles 77.0 and 92.0 cm.

Work out how far each quartile sits from the median in centimetres, and as a ratio. Say whether the log scale would balance the waist figures.

6. Here is a worked answer. Find the step that broke.

> three counts: 1, 10 and 100
> $\log_{10}$ of each: 0, 1, 2
> mean of the logs: (0 plus 1 plus 2) divided by 3 = 1
> geometric mean = e^{1} = 2.72

7. Here is a worked answer. Find the step that broke.

```
a skewed positive quantity: mean 4.1, standard deviation 3.7
logs taken, averaged, and turned back: 3.0
so the mean of the quantity is 3.0, not 4.1
the earlier 4.1 was an error
```

8. A colleague prepares a reference range for a skewed positive quantity from a paper that reports only "4.1 ± 3.7". She writes: "The normal range is the mean plus or minus two standard deviations. Values above 11.5 are abnormal, and no value can be abnormally low."

Decide what to compute, compute it, and say what your answer does not establish.

17 · Covariance and correlation of two random variables

DEFINITION. Let X and Y be two discrete random variables on the same chance process. Their joint distribution lists every pair of values (x, y) the two can take together, with the probability P(X = x and Y = y) of each pair. Adding a row or a column gives the distribution of X or of Y alone.

The covariance of X and Y, written Cov(X, Y) and read "the covariance of X and Y", is the expected value of the product of their deviations from their own expected values.

```
Cov(X, Y) = E[(X - E[X])(Y - E[Y])]
```

It carries the unit of X times the unit of Y. The covariance of X with itself is Var(X).

Multiplying out the bracket gives a shorter form. E[XY] is the expected value of the product X times Y. E[X] and E[Y] are fixed numbers, so they come outside an expected value, as in Book 2, section 14.

```
(X - E[X])(Y - E[Y]) = XY - X E[Y] - E[X] Y + E[X] E[Y]
Cov(X, Y) = E[XY] - E[X] E[Y] - E[X] E[Y] + E[X] E[Y]
Cov(X, Y) = E[XY] - E[X] × E[Y]
```

The variance of a sum or a difference of any two random variables, independent or not, is:

```
Var(X + Y) = Var(X) + Var(Y) + 2 Cov(X, Y)
Var(X - Y) = Var(X) + Var(Y) - 2 Cov(X, Y)
```

The correlation of X and Y, written with the Greek letter rho, ρ, is the covariance divided by both standard deviations.

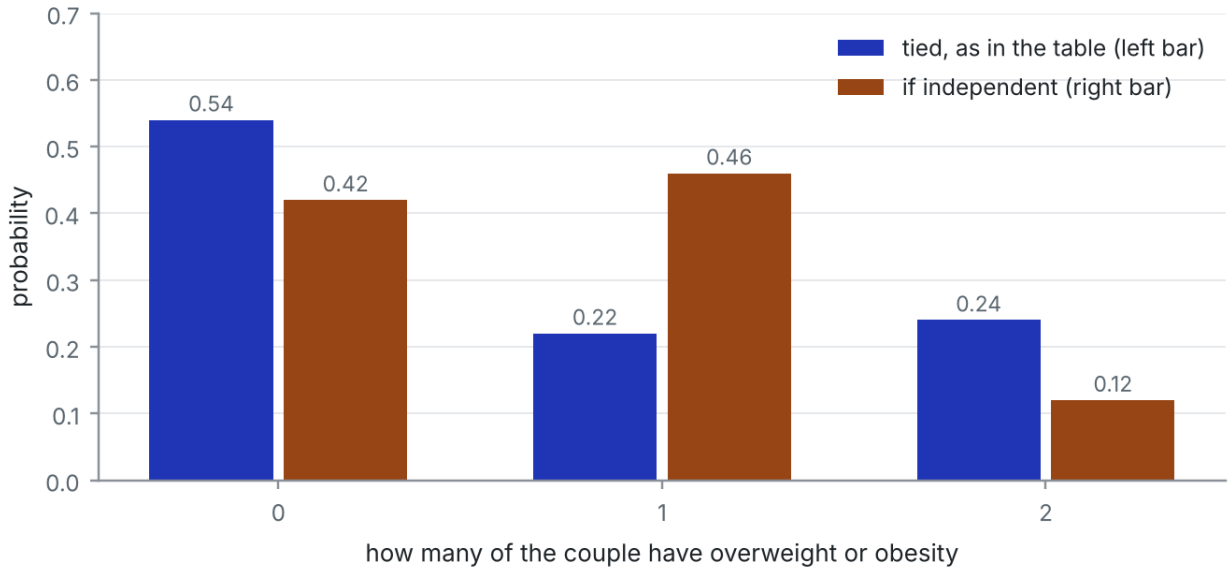
```
ρ = Cov(X, Y) ÷ [SD(X) × SD(Y)]
```

It has no unit. It always lies between −1 and 1 (stated here, not proved; the proof is beyond this book). It measures only how closely the two follow a straight line.

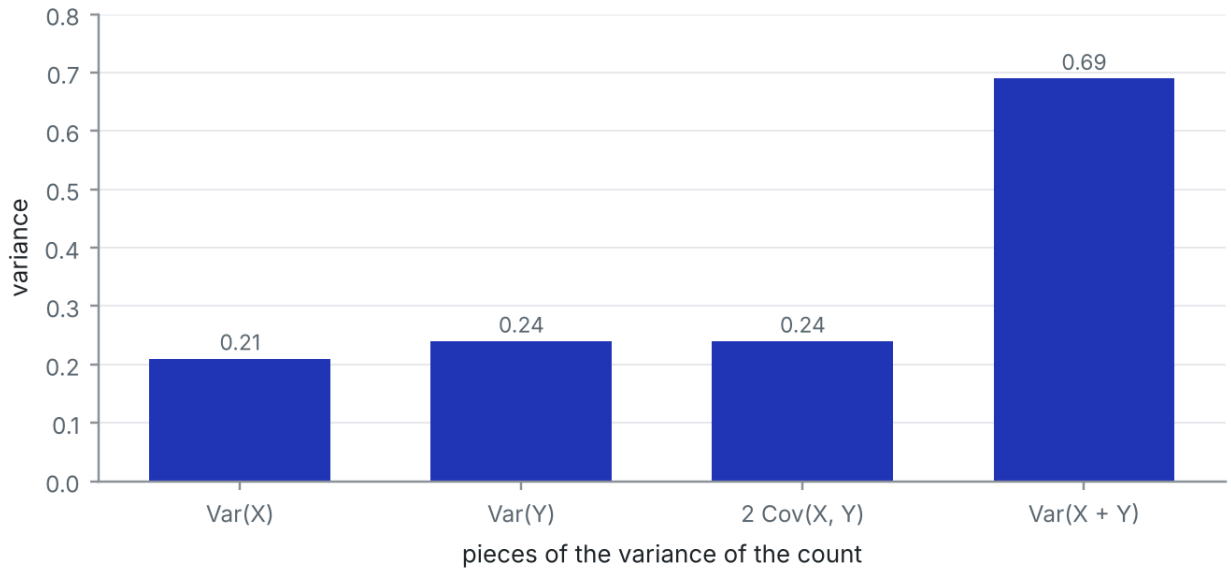
If X and Y are independent, then E[XY] = E[X] × E[Y], so their covariance and correlation are 0. The reverse is false: a covariance of 0 does not make X and Y independent. [154, 155, 156, 157, 158, 159, 160, 161]

IN PLAIN TERMS. The covariance measures how two quantities move together. When one is above its average, is the other usually above its average too? Then the covariance is positive, and the sum spreads more than the adding rule says.

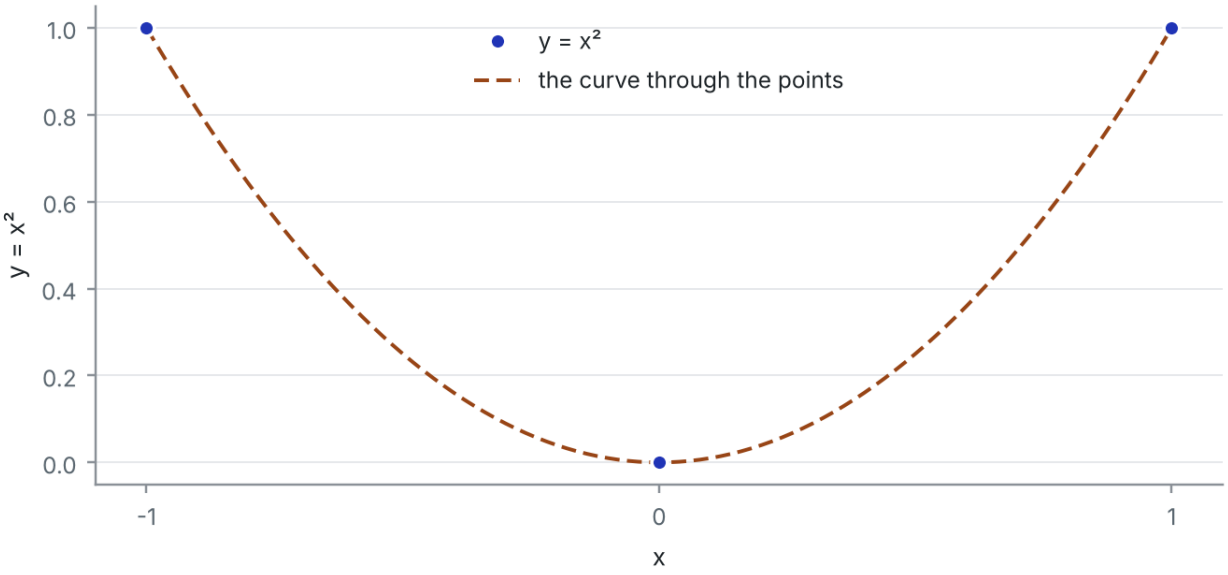
The correlation is the same measure with the units taken out.



How many of a made-up couple have overweight or obesity. With the tie between spouses (covariance 0.12), the count is 0 with probability 0.54, 1 with 0.22 and 2 with 0.24: variance 0.69. If the two were independent with the same 0.30 and 0.40, it would be 0.42, 0.46 and 0.12: variance 0.45. The tie moves probability from one to both or neither.



The variance of the couple's count, 0.69, is Var(X), 0.21, plus Var(Y), 0.24, plus 2 Cov(X, Y), 0.24. The adding rule keeps only the first two and gives 0.45.



X is minus 1, 0 or 1, each with probability one third, and Y is X squared. The three points lie on a U. The covariance and correlation are 0, yet X fixes Y: a correlation sees only a straight-line tilt.

Illustration 1. Start where the adding rule breaks. In Book 2, section 14 a made-up day-1 energy surplus X was minus 300, 0, 200 or 500 kcal a day, with probabilities 0.2, 0.3, 0.4 and 0.1, and E[X] was 70. Day 2's surplus Y was always 140 minus day 1's. So E[Y] was 70 too, and X + Y was 140 on every row.

First find Var(X), as in Book 2, section 15. Take each value's deviation from 70, square it, weight it by its probability and add.

x (kcal/day)	x − 70	$(x - 70)^2$	probability	$(x - 70)^2 \times$ probability
-300	-370	136900	0.2	27380
0	-70	4900	0.3	1470
200	130	16900	0.4	6760
500	430	184900	0.1	18490

```
27380 plus 1470 plus 6760 plus 18490 = 54100
```

Var(X) is 54,100, in the unit (kcal a day) squared. Y's deviations are X's with the sign turned round, so its squared deviations are the same, and Var(Y) is 54,100 too.

Now try the adding rule on X + Y.

```
54100 plus 54100 = 108200
```

That says X + Y spreads a great deal. But X + Y is 140 on every row. It never varies. Its variance is 0. The adding rule is wrong here, because the two days are not independent.

The missing piece is the covariance. On each row, multiply X's deviation by Y's, weight the product by the row's probability, and add. Y's deviation is minus X's: on the first row, Y is 440 and 440 minus 70 is 370.

```
-370 times 370 times 0.2 = -27380
-70 times 70 times 0.3 = -1470
130 times -130 times 0.4 = -6760
430 times -430 times 0.1 = -18490
-27380 plus -1470 plus -6760 plus -18490 = -54100
```

Cov(X, Y) is minus 54,100. Every row has one day above 70 and the other below, so every product is negative. Put it into the full rule.

```
54100 plus 54100 plus 2 times -54100 = 0
```

Var(X + Y) is 0, as it must be. The correlation is the covariance divided by both standard deviations. Each standard deviation is the square root of 54,100, so their product is 54,100.

```
-54100 divided by 54100 = -1
```

A correlation of minus 1: Y is an exact falling straight line in X, Y = 140 − X.

Where does the 2 Cov(X, Y) come from? Write the deviation of X + Y from its expected value as (X − E[X]) plus (Y − E[Y]). Square that sum. (a + b) times (a + b) is a(a + b) plus b(a + b), the multiplying out of Book 0, C1 done twice. That is $a^2 + ab + ba + b^2$, or $a^2 + b^2 + 2ab$. Take the expected value of each piece. The squares give Var(X) and Var(Y). The cross term gives 2 times E[(X − E[X])(Y − E[Y])], which is 2 Cov(X, Y). Nothing in those steps asked whether X and Y were independent.

And why does independence make the covariance 0? Multiply out (X − E[X])(Y − E[Y]) and take expected values: Cov(X, Y) = E[XY] − E[X] × E[Y]. For independent X and Y, in the sense of Book 2, section 15, each cell of the joint table is P(X = x) times P(Y = y), the product rule of section 3 applied to each pair of values. E[XY] is the sum over cells of x times y times that product. For each x, the inner sum over y is the same E[Y].

```
Σ over x and y of x y P(X = x) P(Y = y) = Σ over x of x P(X = x) × E[Y] = E[X] ×
E[Y]
```

The two terms cancel, and the covariance is 0.

WHERE THIS PICTURE BREAKS. The two days are made up and as tied together as two days can be, to show the full size of the error. Real days are tied more loosely, with a correlation somewhere between −1 and 1, and the covariance then corrects the adding rule by less. A variance in (kcal a day) squared is an awkward unit; that awkwardness is why the correlation, which has none, is the number people quote.

Illustration 2. Now a sum whose parts move together. Take a married couple aged 15 to 49 and count how many of the two have overweight or obesity. Let X be 1 if the husband does and 0 if not, and Y the same for the wife. The count is X + Y. Couples share meals, a kitchen and a neighbourhood, so there is no reason to expect X and Y to be independent.

Here is a joint table. Its numbers are made up.

	wife: Y = 1	wife: Y = 0	husband total
husband: X = 1	0.24	0.06	0.30
husband: X = 0	0.16	0.54	0.70
wife total	0.40	0.60	1

The row totals give X alone: P(X = 1) is 0.30. The column totals give Y: P(Y = 1) is 0.40. Each is a Bernoulli variable (section 12), so its expected value is p and its variance is p(1 − p).

```
0.30 times 0.70 = 0.21
0.40 times 0.60 = 0.24
```

Var(X) is 0.21 and Var(Y) is 0.24. The quick route makes the mistake first: treat the two as independent, and add.

```
0.21 plus 0.24 = 0.45
```

Now the covariance, by the short form. XY is 1 only when both have overweight or obesity, so E[XY] is the top-left cell, 0.24.

```
0.30 times 0.40 = 0.12
0.24 minus 0.12 = 0.12
```

Cov(X, Y) is 0.12. Check it by the definition, cell by cell: the deviations are 1 − 0.30 or 0 − 0.30 for X, and 1 − 0.40 or 0 − 0.40 for Y.

```
0.70 times 0.60 times 0.24 = 0.1008
0.70 times -0.40 times 0.06 = -0.0168
-0.30 times 0.60 times 0.16 = -0.0288
-0.30 times -0.40 times 0.54 = 0.0648
0.1008 plus -0.0168 plus -0.0288 plus 0.0648 = 0.12
```

The same 0.12. Now the variance of the count.

```
0.21 plus 0.24 plus 2 times 0.12 = 0.69
```

Var(X + Y) is 0.69, not 0.45. Check it straight from the count. The count is 2 with probability 0.24, 1 with probability 0.06 plus 0.16, which is 0.22, and 0 with probability 0.54. Its expected value is 0.70.

```
2 times 0.24 plus 1 times 0.22 = 0.70
4 times 0.24 plus 1 times 0.22 = 1.18
1.18 minus 0.70 times 0.70 = 0.69
```

The second line is E[(X + Y)^2], and the third uses Var = E[square] minus the square of E, the short form with Y set equal to X. It agrees.

If the two really were independent, with the same 0.30 and 0.40, the count would be 2 with probability 0.30 times 0.40, which is 0.12, 0 with probability 0.70 times 0.60, which is 0.42, and 1 with the rest, 0.46. Its variance would be 0.45. The tie between spouses moves probability from "one of the two" to "both" and "neither", and that widens the spread.

The correlation:

```
0.21 times 0.24 = 0.0504
square root of 0.0504 = 0.2245
0.12 divided by 0.2245 = 0.53
```

About 0.53: a moderate positive tie, with no unit.

This is why a survey that samples whole households, as the National Family Health Survey (NFHS) does, cannot treat its people as independent draws. Adults in one household are positively tied, so a count from households spreads more than the binomial of section 12 would say.

WHERE THIS PICTURE BREAKS. The table is made up, with round margins; they are not survey figures. It says nothing about how strongly real spouses' weights are tied. The covariance describes a tie; it does not say what causes it, whether shared food, shared surroundings, or people choosing partners like themselves.

Illustration 3. Now the trap the other way. Let X take the values minus 1, 0 and 1, each with probability one third, and let Y be X squared. Y is decided completely by X.

```
x    y = x²  probability  x times y
-1 1    1/3     -1
0    0    1/3     0
1    1    1/3     1
```

```
E[X] = -1 times 1/3 plus 0 times 1/3 plus 1 times 1/3 = 0
E[XY] = -1 times 1/3 plus 0 times 1/3 plus 1 times 1/3 = 0
```

E[Y] is 2/3. The covariance is E[XY] − E[X] × E[Y].

```
0 minus 0 times 2/3 = 0
```

The covariance is 0, so the correlation is 0. Yet X and Y are as dependent as can be: knowing X fixes Y. Check it with the rule of section 3. P(X = 1) is 1/3 and P(Y = 0) is 1/3, so independence would need P(X = 1 and Y = 0) to be 1/9. It is 0, because X = 1 forces Y = 1.

The three points lie on a U, not on a line. The two arms of the U cancel. The covariance sees only a straight-line tilt, and a U has none.

WHERE THIS PICTURE BREAKS. The numbers are bare and chosen for a perfect U. Real U-shaped patterns are noisier, and their correlation comes out near 0 rather than exactly 0. The lesson holds either way: a correlation near 0 rules out a straight-line tie and nothing else.

MUST KNOW POINTS FOR YOU.

- "The correlation is 0, so the two are unrelated" is the error. A U-shaped pattern can have a correlation of exactly 0 while one variable fixes the other. Before you accept "no relation", look at the scatter plot.
- Before you add variances, ask whether the parts are independent. If they share a household, a clinic, a measurer or a scale, add 2 Cov(X, Y) as well.
- When a paper argues "uncorrelated, so independent", ask for the joint table or the scatter plot. A covariance of 0 cannot show independence; only those can.
- Quote a correlation, not a covariance, when you compare the strength of two ties. A covariance changes with the units; the same tie in centimetres has a covariance 100 times the one in metres. The correlation does not change.
- A difference between two positively tied measurements spreads less than the adding rule says, because 2 Cov(X, Y) is taken away.
- When a journalist reports that two things are "strongly correlated", say what the number does: it measures how closely the two follow a straight line together.

COMMON MISREADING. "The correlation is near zero, so the two have nothing to do with each other." A correlation measures only how closely two quantities follow a straight line. A U-shaped tie can give a correlation of exactly 0 while one quantity fixes the other. Near 0 rules out a straight-line tie, and nothing more.

EXERCISE 1 (retrieval). From memory: write the variance of X + Y for any two random variables, and say when the last term is 0. What does a correlation of 0 tell you, and what does it not?

EXERCISE 2 (critique). A thesis protocol reads: "Each household contributes two adults. Each adult has a 0.3 probability of hypertension, so the number with hypertension in 200 households (400 adults) has variance 400 times 0.3 times 0.7, which is 84."

Find the flaw, and say which way it pushes the answer.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. X and Y each take the values 0 and 1. Here are their joint probabilities.

	Y = 0	Y = 1
X = 0	0.4	0.1
X = 1	0.2	0.3

Work out E[X], E[Y], E[XY] and Cov(X, Y).

2. Var(X) = 4, Var(Y) = 9 and Cov(X, Y) = 3. Work out Var(X + Y) and Var(X − Y).

3. First: Cov(X, Y) = −6, SD(X) = 3 and SD(Y) = 4. Work out the correlation.

Second, the other way round: the correlation of U and V is 0.8, SD(U) = 2 and SD(V) = 5. Work out Cov(U, V).

4. You weigh dal by difference, as in Book 2, section 15: the katori with dal, minus the empty katori. Each reading has a variance of 50 square grams. This scale's errors are tied: a reading that is high tends to be followed by another high one, with a covariance of 30 square grams between the two readings. These numbers are made up.

Work out the variance and the standard deviation of the dal's weight error, with the tie and without it.

5. In a made-up group, weight has an SD of 10 kg, height has an SD of 0.1 m, and the covariance of weight and height is 0.5 kg·m (kilogram-metres).

Work out the correlation. Then restate height in centimetres and work out the new SD, the new covariance and the new correlation.

6. In a made-up diet record, breakfast energy has an SD of 150 kcal. Lunch energy has an SD of 200 kcal. The correlation between them is −0.3: people who eat a larger breakfast tend to eat a smaller lunch.

Work out the SD of breakfast plus lunch. Compare it with the SD you would get if the two were independent.

7. Here is a worked answer. Find the step that broke.

```
X and Y each 1 or 0; P(X = 1) = 0.5, P(Y = 1) = 0.2, P(X = 1 and Y = 1) = 0.15
Var(X) = 0.5 times 0.5 = 0.25
Var(Y) = 0.2 times 0.8 = 0.16
Var(X + Y) = 0.25 plus 0.16 = 0.41
```

8. Here is a worked answer. Find the step that broke.

```
X is -1, 0 or 1 with probabilities 0.25, 0.5 and 0.25; Y = X squared
E[X] = 0, E[XY] = E[X cubed] = 0
Cov(X, Y) = 0 minus 0 times E[Y] = 0
so X and Y are independent, and P(X = 1 and Y = 1) = 0.25 times 0.5 = 0.125
```

9. A made-up pilot study of three people gives sleep of 5, 7 and 9 hours a night and BMI of 27, 23 and 27 kg/m^2 in the same order. A note says: "The correlation between sleep and BMI was zero, so sleep has nothing to do with BMI in this group."

Decide what to compute, compute it, and say what your answer does not establish.

10. A thesis builds a total score by adding two subscores. The protocol says: "Subscore A has an SD of 10 and subscore B an SD of 15, so the total has an SD of 25."

Decide what to compute, compute it, and say what your answer does not establish.

18 · Sums and averages of many independent draws, simulated by hand

DEFINITION. Let X1, X2, ..., Xn be n independent random variables, each with the same distribution, with expected value μ, the Greek letter mu, and variance σ^2, where σ, the Greek letter sigma, is the standard deviation (SD). Call each one a draw.

Their sum, T = X1 + X2 + ... + Xn, has:

$$E[T] = n\mu$$
$$\mathrm{Var}(T) = n\sigma^2$$
$$\mathrm{SD}(T) = \sigma \times \sqrt{n}$$

Their average, written $\bar{X}$ and read "X-bar", is T divided by n.

$$E[\bar{X}] = \mu$$
$$\mathrm{Var}(\bar{X}) = \sigma^2 \div n$$
$$\mathrm{SD}(\bar{X}) = \sigma \div \sqrt{n}$$

A binomial count is the sum of n Bernoulli draws (section 12). Divide it by n, and the share of successes has expected value p and standard deviation $\sqrt{[p(1-p) \div n]}$.

The law of large numbers: as n grows, the chance that $\bar{X}$ lies further than any fixed distance from μ tends to 0: it can be made as small as you like by taking n large enough.

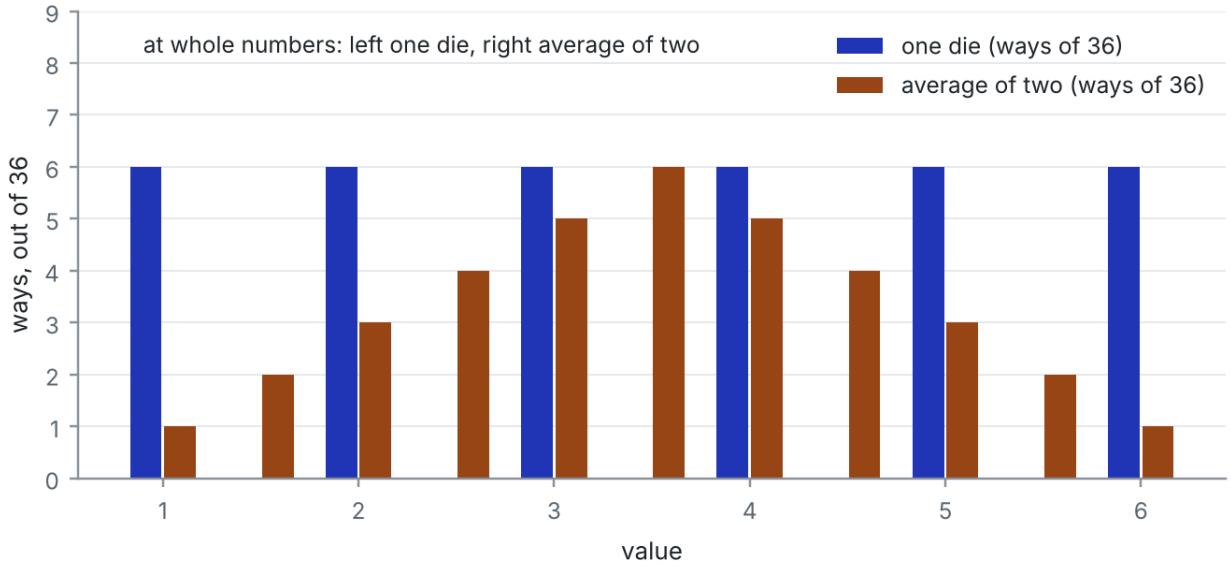
The central limit theorem: as n grows, the distribution of T, and of $\bar{X}$, comes closer and closer to a normal distribution (section 15), whatever the shape of one draw. It needs the draws to be independent and σ to be finite: the spread of one draw must be a fixed number, not one that grows without limit. A strongly skewed draw needs a much larger n.

The distribution of $\bar{X}$ over all the samples of size n that could be drawn is called the sampling distribution of the mean. [162, 163, 164, 165, 166, 167, 168, 169, 170]

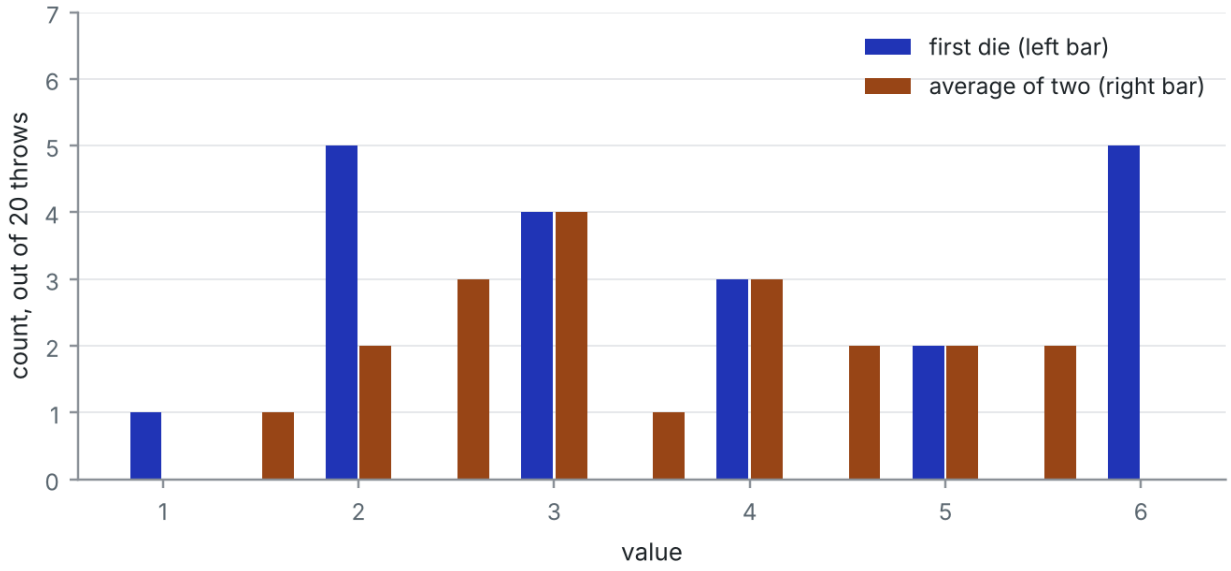
IN PLAIN TERMS. One person's value can land almost anywhere. The average of many people's values is much steadier, because their highs and lows partly cancel.

Four times as many people halve the spread of the average.

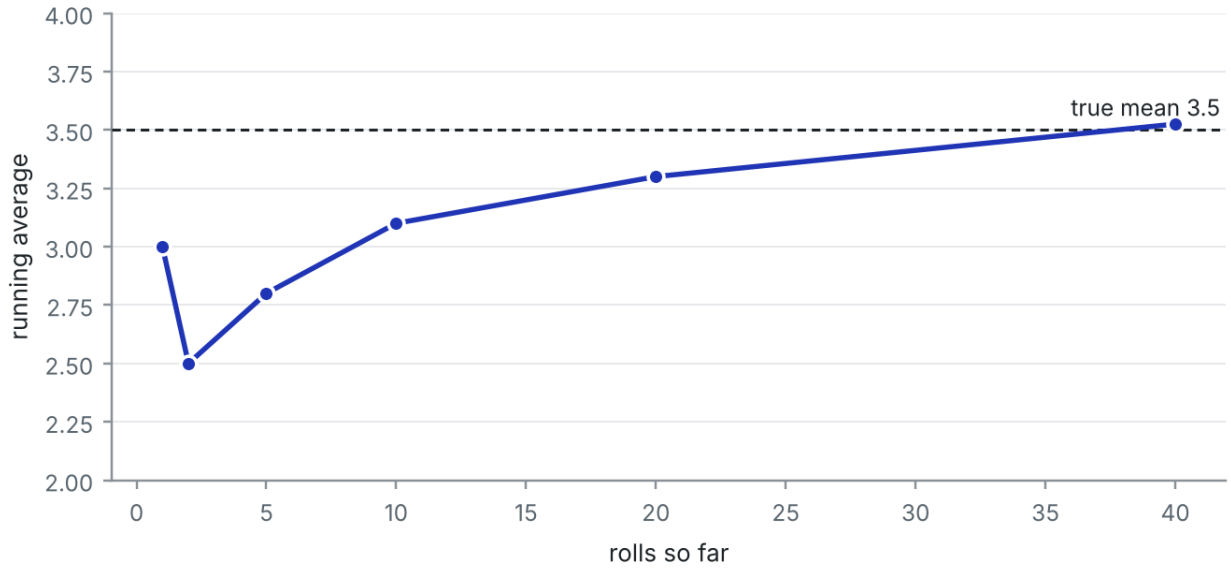
Roll dice, write the results down, and watch it happen on paper.



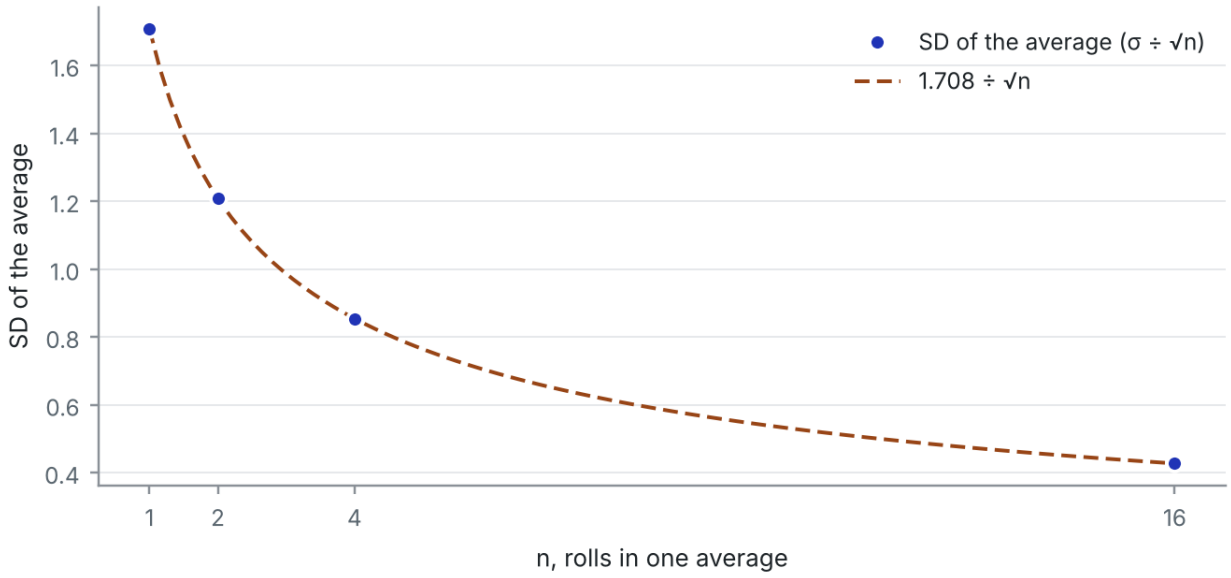
All 36 ways two dice can land. One die puts 6 ways on each face from 1 to 6, a flat spread. The average of two puts 1 way on 1.0, rising to 6 ways on 3.5 and falling back to 1 way on 6.0. The extremes are six times rarer for the average, and the shape is already peaked.



The record of 20 throws, tabulated. The first die (left bar of each pair) lands on all six faces, 5 times each on 2 and on 6. The 20 averages of two dice (right bars) crowd towards the middle, and none is 1 or 6. Their SD is about 1.21, against about 1.68 for one die.



The running average of the 40 single rolls in the record. It swings while the rolls are few, from 3.0 after one roll to 2.5 after two, and settles as they pile up, reaching 3.525 after 40. The dashed line is the true mean, 3.5.



The SD of the average of n rolls of a fair die is 1.708 divided by the square root of n: 1.708 for one roll, 1.208 for two, 0.854 for four and 0.427 for sixteen. Each fourfold increase in n halves the spread.

Illustration 1. Here is the mistake to watch for first. A resident says: “One die can land anywhere from 1 to 6. So the average of two dice can land anywhere from 1 to 6 just as easily.” Both ranges are 1 to 6. The resident’s mistake is in “just as easily”. Check it with dice and a pen.

Get two ordinary dice. Throw both. Write down the first die, the second die, and their average. Do it 20 times. That is the whole method of simulation by hand: draw, record, repeat, tabulate.

Here is one record of 20 throws, made with a random-number generator in place of real dice. Yours will differ, and that is part of the lesson.

throw	first die	second die	average
1	3	2	2.5
2	1	5	3.0
3	3	3	3.0
4	2	1	1.5
5	5	6	5.5
6	3	4	3.5
7	2	3	2.5
8	3	5	4.0
9	2	2	2.0
10	6	5	5.5
11	6	4	5.0
12	4	2	3.0
13	6	2	4.0
14	6	4	5.0
15	4	2	3.0
16	2	6	4.0
17	2	2	2.0
18	4	1	2.5
19	5	4	4.5
20	6	3	4.5

Tabulate. Count how often each value turns up in the first-die column, and in the average column.

first die	1	2	3	4	5	6
count	1	5	4	3	2	5

average	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
count	1	2	3	4	1	3	2	2	2

The first die is spread across all six faces, with plenty at 2 and at 6. The averages crowd towards the middle. None of the 20 averages is 1 or 6.

Measure the spread, as in Book 0, D5. For the first-die column the mean is 75 divided by 20, which is 3.75. The squared deviations from it add to 53.75. Divide by one less than the count.

```
53.75 divided by 19 = 2.829
square root of 2.829 = 1.682
```

For the average column the mean is 70.5 divided by 20, which is 3.525. Its squared deviations add to 27.7375.

```
27.7375 divided by 19 = 1.460
square root of 1.460 = 1.208
```

The averages spread with an SD of about 1.21, against about 1.68 for one die.

Now the theory. One fair die has μ = 3.5. Its variance is the average of the squared deviations from 3.5: 6.25, 2.25, 0.25, 0.25, 2.25 and 6.25.

```
6.25 plus 2.25 plus 0.25 plus 0.25 plus 2.25 plus 6.25 = 17.5
17.5 divided by 6 = 2.917
square root of 2.917 = 1.708
1.708 divided by square root of 2 = 1.208
```

So σ is about 1.708, and the average of two dice should have an SD of σ ÷ √2, about 1.208. The record gave 1.682 and 1.208. Another record would give other numbers near these.

The exact shapes come from listing all 36 ways two dice can land, as in section 1. One die puts 6 of the 36 ways on each face. The average of two puts 1 way on 1.0, 2 on 1.5, and so on, rising to 6 ways on 3.5 and falling back to 1 way on 6.0.

value	one die (ways of 36)	average of two (ways of 36)
1.0	6	1
1.5	0	2
2.0	6	3
2.5	0	4
3.0	6	5
3.5	0	6
4.0	6	5
4.5	0	4
5.0	6	3
5.5	0	2
6.0	6	1

An average of 1 needs both dice to show 1: one way in 36. One die shows 1 in 6 ways of 36. So the resident was wrong: the extremes are six times rarer for the average. With more dice the peak narrows further and its sides round into the bell of the normal distribution. That is the central limit theorem, seen in miniature.

WHERE THIS PICTURE BREAKS. The 20-throw record is one record, from a random-number generator. Its SDs of 1.682 and 1.208 landed close to theory; a record of 20 can land well away, and that wobble is itself what the next rung studies. Two dice are a long way from “many”. The tri-

angle is not yet a bell, and the theorem says nothing about how close to normal an average of two is.

Illustration 2. The same record shows the law of large numbers. Read the 40 single rolls in order, first die then second die, throw by throw: 3, 2, 1, 5, 3, 3, and so on. After each roll, work out the average of all the rolls so far.

rolls so far	total	running average
1	3	3.0
2	5	2.5
5	14	2.8
10	31	3.1
20	66	3.3
40	141	3.525

After one roll the running average is 3.0. After two it is 2.5. It swings while there are few rolls, and settles as they pile up: after 40 it is 3.525, close to the true 3.5.

The law of large numbers says this settling is what tends to happen. It does not say the average of 40 rolls will be 3.5, and it does not say that later rolls "make up" for early ones. A die has no memory. The early swings are simply outweighed by the growing number of rolls.

How fast does it settle? The SD of an average of n rolls is $\sigma \div \sqrt{n}$. Work it out for a few n.

n	SD of the average ($\sigma \div \sqrt{n}$)
1	1.708
2	1.208
4	0.854
16	0.427

Going from 1 roll to 4 halves the SD. Going from 4 to 16 halves it again.

WHERE THIS PICTURE BREAKS. The table of SDs assumes independent rolls of a fair die. A loaded die, or rolls that depend on each other, breaks it. The running average of 3.525 is this record's; another record will cross 3.5 at different points.

Illustration 3. Now a real quantity. In the sixth National Family Health Survey (NFHS-6, 2023-24), 30.7 per cent of women aged 15 to 49 in India had a body-mass index (BMI) of 25 or more. Open the NFHS-6 India fact sheet at https://www.nfhsiips.in and find indicator 76. Treat 0.307 as p, the chance that one woman picked at random from that population is in this group.

One woman is a Bernoulli draw: 1 or 0. Her SD is $\sqrt{[p(1 - p)]}$.

```
0.307 times 0.693 = 0.2128
square root of 0.2128 = 0.4613
```

Now pick 100 women at random and take the share among them with a BMI of 25 or more. That share is the average of 100 Bernoulli draws. Its expected value is 0.307 and its SD is 0.4613 ÷ √100.

```
0.4613 divided by 10 = 0.04613
```

About 0.046, or 4.6 percentage points. With 400 women:

```
0.4613 divided by 20 = 0.0231
```

About 2.3 percentage points. Four times the women, half the spread.

The trap is to use the SD of one woman, 0.46 or 46 points, as the spread of the share.

```
31 minus 46 = -15
31 plus 46 = 77
```

Even one SD either way would run from 0 (nothing can be below it) to about 77 per cent. Picked at random, a sample of 100 gives a share within a few points of 31 per cent nearly every time.

NFHS itself does not pick women one at a time. It picks households, and women in one household are positively tied (section 17). So a real survey's share spreads more than this formula says, and the formula here is a floor, not the survey's own figure.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The 4.6 and 2.3 points describe simple random samples from a population whose true share is exactly 30.7 per cent. They do not describe NFHS's own precision, which is set by how it picks households and is wider. They also say nothing about bias: a sample of 400 from one college is precise about that college and may be far from India's figure (Book 0, D6).

MUST KNOW POINTS FOR YOU.

- The SD of an average is the SD of one person divided by the square root of n. Never use the SD of individuals as the uncertainty of a mean; divide first.
- Halving the spread of an average costs four times the people. Doubling a sample cuts the spread by only about 29 per cent.
- In a classroom simulation of 50 averages of 4 dice, the spread of each average uses n = 4.
- The σ ÷ √n rule needs independent draws. People from one household, one cluster or one clinic day are tied, and their average spreads more than the rule says. Before you quote σ ÷ √n for a survey, ask how it was sampled.
- The central limit theorem promises a normal shape only as n grows, and how fast depends on the shape of one draw. For a strongly skewed quantity, such as one where most people have 0 and a few have a very large value, an average of 30 can still be far from normal. Look at the data before you lean on the normal shape.
- After a run of high values, the next value is no more likely to be low.
- Teach the square-root law with dice before you teach the formula.

COMMON MISREADING. "The sample has a mean of 22.1 and an SD of 7.3, so the true mean could be anywhere within 7.3 of 22.1." The SD in a table of results describes how the people differ from each other. The spread of their mean is that SD divided by the square root of n. For 131 people that is about 0.64, not 7.3. And it holds only if they were drawn independently at random.

EXERCISE 1 (retrieval). From memory: for n independent draws, each with mean μ and SD σ, give the mean and SD of their sum and of their average. What does the central limit theorem say, and what does it need?

EXERCISE 2 (design). Plan a 30-minute session in which residents simulate by hand the average of four dice. Say what each resident does, what is recorded, how the records are pooled, and what the group should see at the end.

PRACTICE. Nine problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Each of 25 independent draws has mean 5 and variance 4. Work out the mean, variance and SD of their sum, and of their average.

2. One draw has an SD of 12. How many independent draws must be averaged to bring the SD of the average down to 3? To 1?

3. OpenIntro's *Statistics* (4th edition), §3.4.4, gives John's daily commute an SD of 4 minutes, and takes the five days of a week as independent. Work out the SD of his total commute time for the week, and of his average daily time over the week.

4. Take NFHS-6's 30.7 per cent of women aged 15-49 with a BMI of 25 or more as the true share. Draw a simple random sample of 400 such women. Work out the SD of the sample share. Then work out the range in which the normal distribution puts about 95 per cent of such samples' shares.

5. Kiran and colleagues measured body fat in 131 men, medical students in Mangaluru, and give an SD of 7.3 per cent. Suppose the men's body fat in the population they came from has an SD of 7.3. Work out the SD of the mean of 131 men drawn at random from it. Say what the answer assumes.

6. Here is a worked answer. Find the step that broke.

```
one man's body fat: SD 7.3 percentage points
a sample of 131 men has mean 22.1
the sample mean could be off by about 7.3 points either way
so the population mean could be anywhere from 14.8 to 29.4
```

7. Here is a worked answer. Find the step that broke.

```
a class rolls 4 dice, takes the average, and does this 50 times
one die has SD 1.708
so the averages should spread with SD 1.708 divided by square root of 50 = 0.24
```

8. A research proposal says: "The pilot had 200 participants. The main study will have 400, so its estimate will be twice as precise."

Decide what to compute, compute it, and say what your answer does not establish.

9. A health economist says: “Thirty patients is enough for the central limit theorem. So the average cost in our sample of 30 is normal, and I will use plus or minus 1.96 SD.”

Use a made-up cost. Each patient has a 1 in 100 chance of a hospital stay that costs 1,000 units. Otherwise the cost is 0. Decide what to compute, compute it, and say what your answer does not establish.

19 · Parameter, statistic, estimator, estimate

DEFINITION. A parameter is a number that describes a whole population (Book 0, D6) or a probability model, such as the binomial's p (section 12). It is fixed, and usually unknown.

An estimator is a rule that turns a sample into a number meant to stand for a parameter. Before the sample is drawn, the estimator's result is a random variable (Book 2, section 13): it depends on which members the chance process picks, so it has a probability distribution, an expected value and a spread.

An estimate is the number the rule gives on the one sample you actually drew. Once the sample is in hand, the estimate is a fixed number, not a random one.

parameter	estimator, as a rule	estimate, read as
μ, the population mean	add the n values and divide by n	x-bar
σ, the population standard deviation (SD)	the sample SD, dividing by n − 1	s
p, the population share	count the members with the feature and divide by n	p-hat

The estimates x-bar and p-hat are written $\bar{x}$ and $\hat{p}$. The sample SD divides by n − 1, as in Book 0, D5.

A bar over a letter marks a sample mean: $\bar{X}$, a capital, is the estimator before the sample is drawn; $\bar{x}$, a small letter, is the estimate you got. A hat over a letter marks an estimate of the parameter written with that letter, as $\hat{p}$ estimates p. The capital T of section 18 is a sum, not the rule for s.

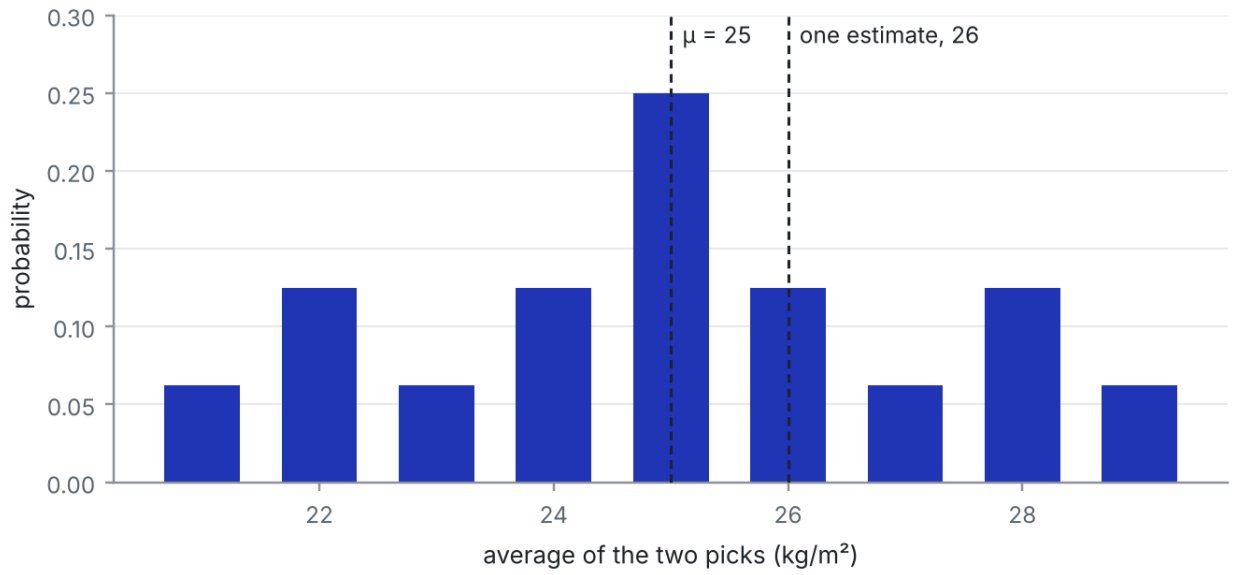
A probability belongs to the estimator, not to the estimate. "In 95 per cent of samples, the rule lands within a stated distance of μ" is a statement about the rule over all the samples it could meet. Once the sample is drawn, the estimate either is within that distance of μ or it is not. Nothing random is left to give that a probability of 95 per cent. [171, 172, 173, 174, 175]

IN PLAIN TERMS. There is a true number for the whole population. You almost never know it. That is the parameter.

You pick a rule, such as "average the people you measured". The rule is the estimator. Before you draw the sample, you cannot say what the rule will give, because you do not know who will be picked. So the rule's answer is a chance quantity.

Then you draw the sample and use the rule. You get one number. That number is the estimate. It is fixed now, and it is either close to the truth or not.

So a "95 per cent" describes how often the rule does well. It never describes the one number you got.



The rule ‘pick two of the four made-up adults with replacement and average them’ has a distribution before any sample is drawn: averages of 21 to 29 kg/m^2, with probabilities from 0.0625 to 0.25. Its expected value is 25, the population mean μ. One sample, 23 and 29, gave the estimate 26.

Illustration 1. Start with a population small enough to know completely. It is made up. Four adults in a family have a body-mass index of 21, 23, 27 and 29 kg/m^2. The population mean μ is a parameter.

```
21 plus 23 plus 27 plus 29 = 100
100 divided by 4 = 25
```

μ is 25. Now pretend you do not know it. Your rule: pick one adult at random, put them back, pick again at random, and average the two values. Putting each one back makes the two picks independent, as in section 3.

Before you pick, the result of the rule is a random variable, $\bar{X}$. There are 4 times 4 = 16 equally likely ordered pairs. Work out the average for each pair, and count how often each average turns up.

average of the two (kg/m^2)	pairs that give it	probability
21	1	0.0625
22	2	0.125
23	1	0.0625
24	2	0.125
25	4	0.25
26	2	0.125
27	1	0.0625
28	2	0.125
29	1	0.0625

That table is the distribution of the estimator. Each probability is the count divided by 16. Its expected value is the average of the 16 averages.

```
21 times 1 plus 22 times 2 plus 23 times 1 plus 24 times 2 plus 25 times 4 = 236
26 times 2 plus 27 times 1 plus 28 times 2 plus 29 times 1 = 164
236 plus 164 = 400
400 divided by 16 = 25
```

The expected value of $\bar{X}$ is 25, the same as μ. On average over all the samples it could meet, the rule lands on the parameter.

Now draw one sample. Say the picks are 23 and 29.

```
23 plus 29 = 52
52 divided by 2 = 26
```

Your estimate is $\bar{x} = 26$. It is one value from the table, and it misses μ by 1.

Ask a question about the rule. How often does it land within 2 of μ, that is, between 23 and 27? Add the probabilities for 23, 24, 25, 26 and 27.

```
0.0625 plus 0.125 plus 0.25 plus 0.125 plus 0.0625 = 0.625
```

The rule lands within 2 of μ in 62.5 per cent of samples. That is a true statement about $\bar{X}$.

Being within 2 of μ is the same as μ being within 2 of the estimate. So you might rewrite the 62.5 per cent as a range round your 26: "There is a 62.5 per cent chance that μ is between 24 and 28." It sounds the same. It is not. μ is 25, so it is between 24 and 28 for certain. Had your picks been 21 and 21, your estimate would be 21. Then μ would be outside 19 to 23 for certain. Either way the 62.5 per cent has nothing left to describe: the picks are made, and μ never moved.

WHERE THIS PICTURE BREAKS. The four adults are made up, and the population is tiny so that every sample can be listed. A real population is too large to list, and μ stays unknown. Then the rule's behaviour comes from the formulas of section 18 or from a simulation, not from a table.

Picking with replacement keeps the two picks independent. A real survey picks without replacement, which changes the spread a little when the sample is a large part of the population (section 13).

Illustration 2. Now a real sample. Kiran and colleagues measured 131 men, medical students at one college in Mangaluru, in 2019. Open their paper, "Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India", in *Heliyon* at https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/. Table 1 gives the men's mid-upper arm circumference as a mean of 30.9 cm with an SD of 3.4 cm.

Sort the three words.

```
the word    in this study
parameter    the mean arm circumference μ of the population these men stand for: un-
known
estimator    “add the 131 measurements and divide by 131”
estimate    x̄ = 30.9 cm, and s = 3.4 cm for σ
```

Table 2 gives counts by body-mass index. Among the men, 42 were in the band 25 to 29.9 kg/m^2 and 10 were 30 or more. Use the share rule for p, the share of the population with an index of 25 or more.

```
42 plus 10 = 52
52 divided by 131 is about 0.397
```

$\hat{p} = 0.397$, or 39.7 per cent.

Which population? The rule is only as good as the sample it was fed. These men came forward when the study was announced on campus. They were not picked at random from all medical students, let alone all young men in India. The estimate 0.397 stands for whatever population these volunteers resemble, and the paper cannot tell you which one that is (Book 0, D6).

The same three words fit the national figure. The India fact sheet of the sixth National Family Health Survey (NFHS-6, 2023-24) gives indicator 76: 30.7% of women aged 15 to 49 had a body-mass index of 25.0 or more. Its footnote 15 says the figure excludes pregnant women and women with a birth in the two months before. The parameter is the true share in that group of Indian women. The 30.7 is an estimate. The estimator is the survey’s own rule, which this book does not cover.

Figures quoted in this illustration are from [10, 32].

WHERE THIS PICTURE BREAKS. The table sorts words for one study. It says nothing about whether 30.9 cm or 0.397 is close to its parameter. That needs the estimator’s spread and its bias, which the next section works out.

The counts used here are Table 2’s own.

MUST KNOW POINTS FOR YOU.

- “There is a 95 per cent probability that the true value lies in our range” is the error. The 95 per cent describes the rule over all the samples it could meet; say that, never the sentence in quotes.
- Before you quote a number from a paper, name three things. Name the parameter it stands for, the rule that made it, and the sample the rule was fed.
- Never write μ or p for a number you computed from data. Write $\bar{x}$, s or $\hat{p}$.
- An estimator is only as good as the sample it is fed. Before you quote an estimate, name the population the sample was actually drawn from. For volunteers, that is whoever they resemble, and the estimate stands for nothing wider.
- A patient or a journalist may ask “so what is the real figure?” Give the estimate. Say it is an estimate from a sample, and say from whom.

- A Greek letter is a fixed unknown, and a capital with a bar is a rule. A small letter with a bar or a hat is a number from the data.

COMMON MISREADING. "We found 30.7 per cent, so the true figure is 30.7 per cent, give or take." The 30.7 is one value of a rule. It is an estimate. How far it sits from the true share depends on the rule's spread and its bias. The number on its own shows neither. Write the estimate with its symbol and its sample, never as the parameter.

REPORTING IT. Of 131 male medical students who volunteered at one college, 52 (39.7 per cent) had a body-mass index of 25 kg/m^2 or more. The sample was not drawn at random. This is an estimate of the share in a population like these volunteers.

EXERCISE 1 (retrieval). From memory: what is the difference between an estimator and an estimate? When is each a random variable? And what can a "95 per cent" ever be about?

EXERCISE 2 (interpretation). A district report says this.

> *A survey of 600 randomly chosen women aged 15 to 49 in the district found that 31.5% had a body-mass index of 25 or more. NFHS-6 reports 30.7% for India. The district's women are at higher risk than the national average.*

The district figures are made up. For each of the four numbers in the report (600, 31.5%, 30.7%, 25), say whether it is a parameter, an estimate, a sample size or a definition. Then say what the last sentence would need before you could accept it.

PRACTICE. Seven problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A sample has five values: 4, 6, 7, 9 and 9.

Work out $\bar{x}$. Then work out $\hat{p}$, the share of values that are 7 or more.

2. A sample has three values: 5, 8 and 11. Work out s, the sample standard deviation.

3. Write, as a rule in words, the estimator of the share of a population with some feature. Then look at these two: "$\hat{p} = 0.18$", and "the number in the sample with the feature, divided by the sample size". Say which is the estimator and which the estimate.

4. Kiran and colleagues' Table 2 counts 151 women. Of these, 41 had a body-mass index of 25 to 29.9 kg/m^2 and 5 had 30 or more.

Work out $\hat{p}$ for the share of women with an index of 25 or more. Name the parameter it estimates, and say what you would need to know before calling it an estimate for young Indian women.

5. Kiran's Table 1 gives, for the 131 men, a mean body-fat percentage of 22.1 with an SD of 7.3, and a median of 20.3.

For each of the three numbers, name the parameter it estimates and the rule that made it. Then write the three with the right symbols where one exists.

6. Here is a worked answer. Find the step that broke.

```
Kiran 2022: 52 of 131 men had a body-mass index of 25 or more
p = 52 divided by 131 is about 0.397
so the proportion of young Indian men with an index of 25 or more is 0.397
```

7. Go back to the four made-up adults with indices of 21, 23, 27 and 29 kg/m^2. The rule is "pick two with replacement and average them". A colleague draws one sample, gets 23 and 27, and says:

> *My average is 25. The rule lands within 2 of the true mean 62.5 per cent of the time. So there is a 62.5 per cent chance that the true mean is between 23 and 27.*

Decide what to compute, compute it, and say what the colleague's last sentence gets wrong.

20 · How good is an estimator: bias and spread, found by listing every sample

DEFINITION. The bias of an estimator is its expected value minus the parameter it stands for. An estimator is unbiased when its bias is 0: averaged over every sample it could meet, it lands on the parameter.

```
bias = E[estimator] - parameter
```

The standard error (SE) of an estimator is its standard deviation over every sample it could meet.

For the sample mean of n independent draws, each with standard deviation σ, section 18 gave the standard error.

```
E[X̄] = μ, so the bias is 0
SE of X̄ = σ ÷ √n
```

The sample variance divides the squared deviations by n − 1, as Book 0, D5 did. With independent draws, it is an unbiased estimator of σ^2. Dividing by n instead gives an estimator whose expected value is $\sigma^2 \times (n - 1) \div n$, too small by $\sigma^2 \div n$. Here is why. Write each draw's distance from μ as its distance from X̄ plus X̄'s distance from μ. Square it.

```
(Xi - μ)² = (Xi - X̄)² + 2(Xi - X̄)(X̄ - μ) + (X̄ - μ)²
```

Add over all n draws. The middle terms add to $2(\bar{X} - \mu) \times \Sigma\,(Xi - \bar{X})$, which is 0, because deviations from the sample mean always add to 0.

```
Σ (Xi - μ)² = Σ (Xi - X̄)² + n × (X̄ - μ)²
```

Take expected values. Each $(Xi - \mu)^2$ has expected value σ^2, so the left side gives $n\sigma^2$. The last term gives $n \times \mathrm{Var}(\bar{X})$, which is $n \times \sigma^2 \div n = \sigma^2$.

```
nσ² = E[Σ (Xi - X̄)²] + σ²
E[Σ (Xi - X̄)²] = (n - 1)σ²
```

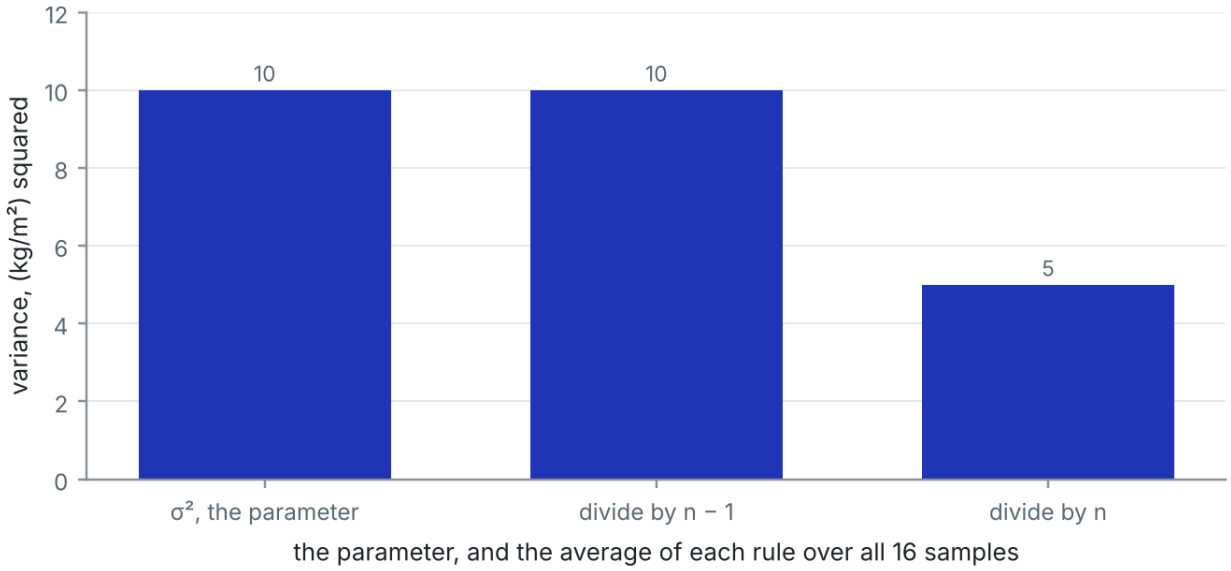
So dividing the sum of squared deviations by n − 1 gives σ^2 on average. Dividing by n gives less. The sum of squared distances is smallest when measured from the sample's own mean. So measuring from X̄ never makes that sum larger than measuring from μ, and usually smaller.

Bias is the systematic part: it points one way and a bigger sample does not shrink it. The standard error is the random part: it shrinks as n grows. [176, 177, 178, 179, 180]

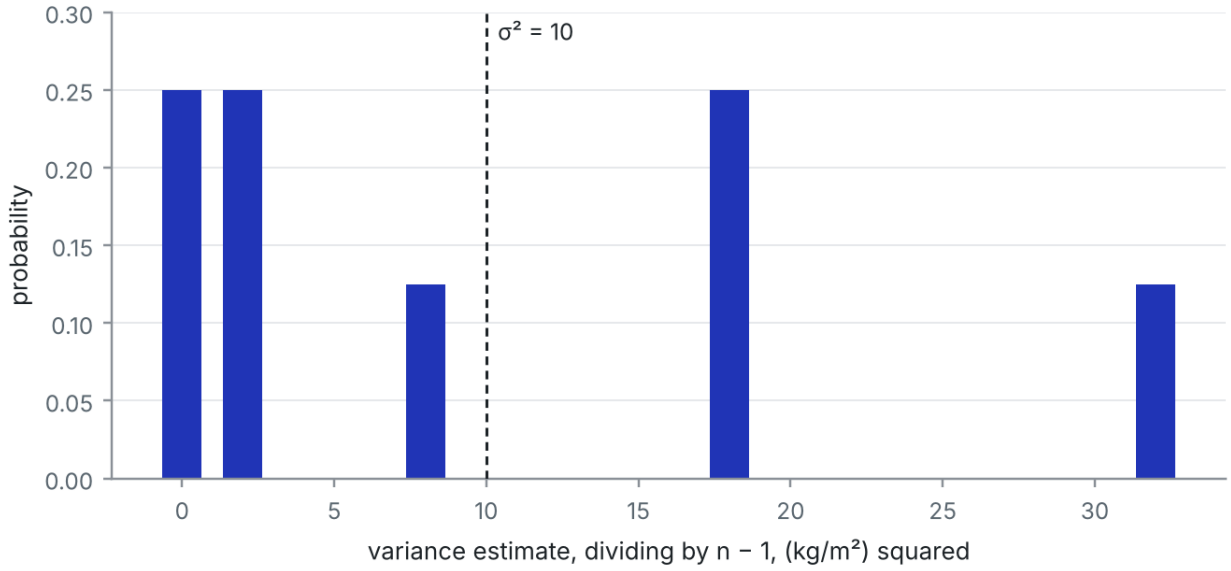
IN PLAIN TERMS. A rule for guessing a true number can go wrong in two ways.

It can aim off. Then even the average of all its guesses misses the truth. That is bias, and a bigger sample does not fix it.

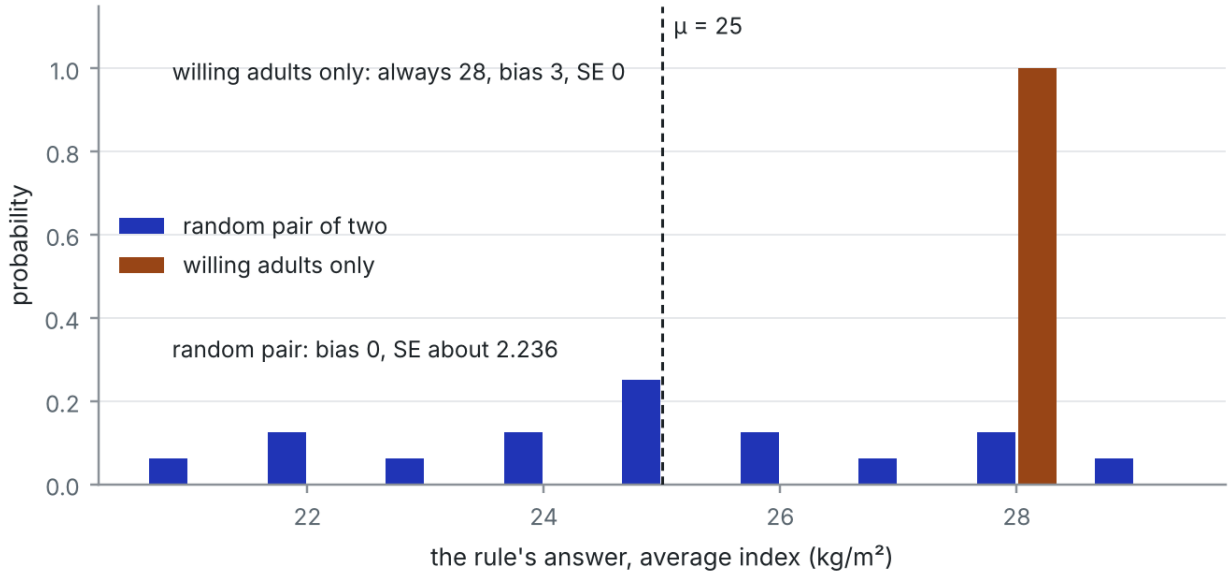
It can scatter. Then its guesses jump about from one sample to the next. That scatter is the standard error, and a bigger sample shrinks it.



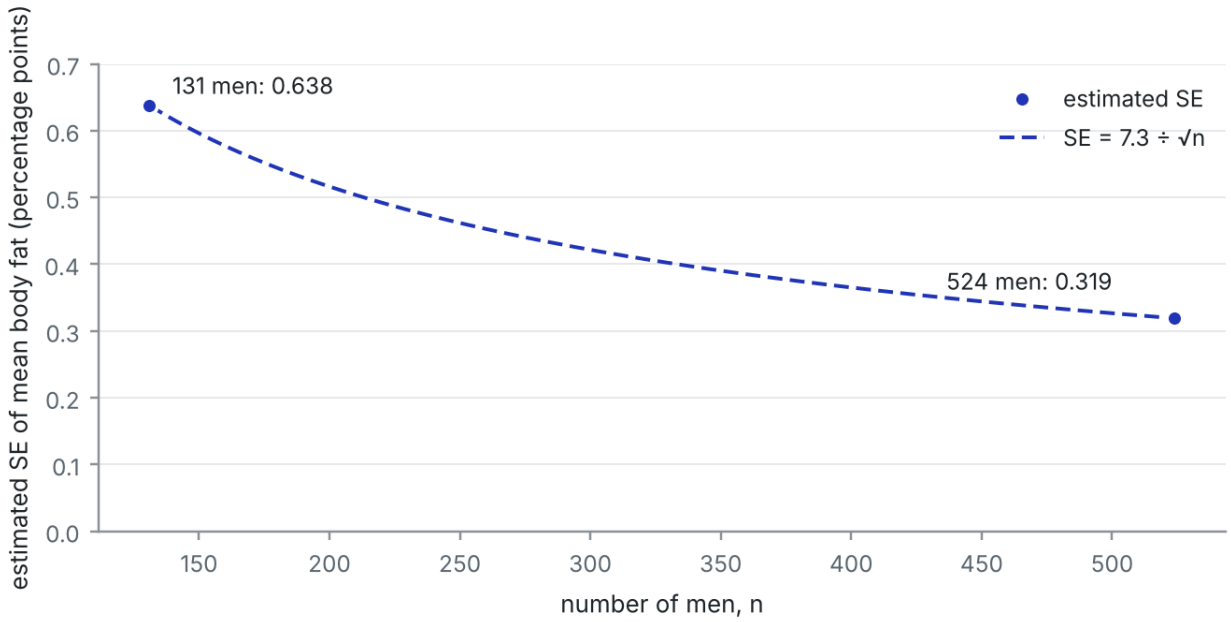
Over all 16 samples of two from the four made-up adults, dividing by n − 1 averages 10, the population variance σ²: unbiased. Dividing by n averages 5: biased low by σ² ÷ n, which is 10 divided by 2.



The ‘divide by n − 1’ rule is unbiased but scatters widely. Over the 16 samples of two it gives 0 with probability 0.25 and 32 with probability 0.125; its expected value is 10, the population variance.



Two rules for μ = 25 in the made-up family. The random pair of two scatters from 21 to 29 but averages 25: bias 0, standard error about 2.236. The rule that averages only the two willing adults, 27 and 29, gives 28 every time: standard error 0, bias 3. Precise and always wrong.



Four times the men halves the standard error. With Kiran's body-fat SD of 7.3 percentage points, 131 men give an estimated standard error of the mean of about 0.638; 524 men, four times as many, give about 0.319. The curve is 7.3 ÷ √n.

Illustration 1. Go back to the four made-up adults of the last section, with body-mass indices of 21, 23, 27 and 29 kg/m^2. Their mean μ is 25. Their variance σ^2 is the mean squared distance from 25.

```
(21 minus 25) squared = 16
(23 minus 25) squared = 4
(27 minus 25) squared = 4
(29 minus 25) squared = 16
16 plus 4 plus 4 plus 16 = 40
40 divided by 4 = 10
```

σ^2 is 10, so σ is the square root of 10, about 3.162 kg/m^2. These are the parameters.

The rule is the same: pick two with replacement. Now give it three jobs. The average estimates μ. Two rival rules estimate σ^2: add the squared deviations from the sample's own average, then divide by n, which is 2, or by n − 1, which is 1.

For a pair a and b, the average is halfway, so each value sits half the gap from it. The two squared deviations add to $(a - b)^2 \div 2$. So the "divide by n" rule gives $(a - b)^2 \div 4$, and the "divide by n − 1" rule gives $(a - b)^2 \div 2$. List all 16 ordered pairs.

first pick	second pick	average	divide by n	divide by n − 1
21	21	21	0	0
21	23	22	1	2
21	27	24	9	18
21	29	25	16	32
23	21	22	1	2
23	23	23	0	0
23	27	25	4	8
23	29	26	9	18
27	21	24	9	18
27	23	25	4	8
27	27	27	0	0
27	29	28	1	2
29	21	25	16	32
29	23	26	9	18
29	27	28	1	2
29	29	29	0	0

Every pair has probability 1 in 16, so each rule's expected value is its column's total divided by 16.

```
average column: 400 divided by 16 = 25
divide by n column: 80 divided by 16 = 5
divide by n - 1 column: 160 divided by 16 = 10
```

The average is unbiased: 25 against μ = 25. Dividing by n − 1 is unbiased: 10 against $\sigma^2 = 10$. Dividing by n is biased: 5 against 10, a bias of minus 5. That is $\sigma^2 \div n$, as the derivation said.

```
10 divided by 2 = 5
```

Now the spread of the average. Its 16 values sit at distances from 25 of −4 once, −3 twice, −2 once, −1 twice, 0 four times, 1 twice, 2 once, 3 twice and 4 once.

```
16 times 1 plus 9 times 2 plus 4 times 1 plus 1 times 2 = 40
0 times 4 plus 1 times 2 plus 4 times 1 plus 9 times 2 plus 16 times 1 = 40
40 plus 40 = 80
80 divided by 16 = 5
```

The variance of the average is 5, and its standard error is the square root of 5, about 2.236 kg/m^2. The formula gives the same.

```
σ² ÷ n = 10 divided by 2 = 5
```

You have found by counting what section 18 found by algebra.

The "divide by n − 1" rule scatters too. Its 16 values take only five different sizes.

divide by n − 1	pairs	probability
0	4	0.25
2	4	0.25
8	2	0.125
18	4	0.25
32	2	0.125

One sample in four gives a variance estimate of 0, and one in eight gives 32. The rule is right on average and wildly wrong on many single samples. Unbiased is not the same as close.

Take the square root of each "divide by n − 1" value and average: about 2.47, below σ = 3.16. The square root of an unbiased estimator is not unbiased.

WHERE THIS PICTURE BREAKS. The population is made up and tiny, so that all 16 samples fit in a table. Real populations cannot be listed; the formulas, or a simulation, take the table's place.

The listing picks with replacement, so the two picks are independent and the formulas hold exactly. Picking without replacement from four people would give twelve ordered pairs, not sixteen, and a smaller spread. The n − 1 rule is then no longer exactly unbiased for this σ^2. With a population thousands of times the sample, the difference is too small to matter.

Illustration 2. Now set a biased rule beside an unbiased one. Suppose that, in the same made-up family, only the two adults with indices 27 and 29 ever agree to be measured. A rule that averages whoever agrees gives the same answer every time.

```
27 plus 29 = 56
56 divided by 2 = 28
```

Its expected value is 28, so its bias is 28 minus 25, which is 3. Its standard error is 0, because it never varies. It is perfectly precise and always wrong, the instrument of Book 0, D7 that agrees with itself and not with the truth.

The random rule has bias 0 and a standard error of about 2.236. On a single sample it may miss by more than 3; its pair 21 and 21 misses by 4. But on average it is right, and a larger random sample shrinks its misses. A larger sample of willing people only repeats the 3.

Real studies face this choice. Kiran and colleagues measured 131 male medical students who came forward when the study was announced on campus. Open their paper in *Heliyon* at https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/. Table 1 gives the men's mean mid-upper arm circumference as 30.9 cm with an SD of 3.4 cm. Use s in place of the unknown σ, and work out the estimated standard error of their mean.

```
the square root of 131 = 11.45
3.4 divided by 11.45 = 0.297
```

About 0.3 cm. That is how far the mean would typically move between random samples of 131 from the population these men were drawn from. It says nothing about how far volunteers sit from all students, or from young men in India. That gap is bias, and no standard error shows it.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The two willing adults are made up. Real volunteers are not always the heaviest or the lightest. The point is only that whatever pulls them in pulls every sample the same way.

The 0.297 cm uses s for σ, so it is itself an estimate of the standard error. Kiran's paper has some numbers that disagree with each other; the Table 1 figures used here are its own.

MUST KNOW POINTS FOR YOU.

- "A bigger sample, so the estimate is right" is the error. A bigger sample shrinks the standard error and leaves the bias exactly where it was.
- A standard error describes only the scatter between random samples from the population the sample actually came from. It is silent on bias.
- Unbiased does not mean close. An unbiased rule can miss badly on the one sample you have.
- Divide by n − 1 when you estimate a variance from a sample; check which divisor any tool used before you copy its variance.
- The sample standard deviation s is biased low even though s^2 is unbiased, because a square root does not pass through an expected value (Book 2, section 14). For large samples the bias is small; for a sample of a handful, do not treat s as σ.

- A protocol that promises twice the precision is promising four times the participants; cost it that way.
- A trainee who already knows a precise, wrong weighing scale already knows a biased estimator with a small standard error.

COMMON MISREADING. "The standard error is small, so the estimate is close to the truth." The standard error measures only how much the estimate would scatter between random samples from the population actually sampled. A precise rule fed a biased sample gives a small standard error around the wrong value. Read a standard error as precision, never as accuracy.

REPORTING IT. The mean mid-upper arm circumference was 30.9 cm (SD 3.4 cm; estimated standard error 0.30 cm; n = 131). The standard error reflects random sampling variation only. The participants were volunteers, and the standard error does not measure how far they may differ from the wider population.

EXERCISE 1 (retrieval). From memory: define the bias and the standard error of an estimator. Which one shrinks as the sample grows? Why does the sample variance divide by n − 1?

EXERCISE 2 (critique). A thesis protocol says this.

> *We will measure waist circumference in 2,000 adults attending the medicine outpatient department. The large sample ensures that the prevalence of abdominal obesity we report will be accurate for the district.*

Name the two kinds of error the sentence runs together. Say which one the 2,000 controls. Then say what the protocol would need before it could claim "accurate for the district".

PRACTICE. Nine problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A population has three values: 2, 4 and 6. List every ordered sample of size 2 picked with replacement, and its average. Work out the expected value of the average.

2. A sample has two values, 3 and 7. Work out the variance estimate dividing by n, and the one dividing by n − 1.

3. Draws are independent with $\sigma = 12$. Work out the standard error of the mean for n = 36 and for n = 144.

4. Kiran and colleagues report, for 131 men, a mean mid-upper arm circumference of 30.9 cm with an SD of 3.4 cm. Work out the estimated standard error of the mean, with its unit, and say what it describes.

5. Kiran's Table 1 gives the 131 men's body-fat percentage with an SD of 7.3. Work out the estimated standard error of the mean. Then work out how many men a study would need to halve it, and check the new standard error.

6. Go back to the four made-up adults with indices of 21, 23, 27 and 29 kg/m^2. A new rule estimates the largest index in the population: pick two with replacement and take the larger. List how often each answer comes up over the 16 ordered pairs, and work out the rule's bias.

7. Here is a worked answer. Find the step that broke.

```
131 men, mid-upper arm circumference SD 3.4 cm
standard error of the mean = SD divided by n
= 3.4 divided by 131 = 0.026 cm
```

8. Here is a worked answer about the four made-up adults, whose σ is about 3.162 kg/m^2. Find the step that broke.

```
the sample variance, dividing by n - 1, is unbiased for σ²
so its square root, s, is unbiased for σ
so the average of s over the 16 samples is 3.162
```

9. Two made-up surveys of the same city's women aged 15 to 49 report the share with a body-mass index of 25 or more. One measured 10,000 women at shopping malls and found 35 per cent. One measured 400 women from a random sample of households and found 28 per cent. A colleague says:

> *The mall survey is 25 times larger, so its figure is the more reliable one.*

Decide what to compute, compute it, and say what the comparison does not establish.

21 · The likelihood: the same formula read as a function of the parameter

DEFINITION. The binomial formula of section 12 gives the probability of k successes in n independent trials. Each trial has the same probability of a success, p.

```
P(X = k) = (n choose k) × p^k × (1 - p)^(n - k)
```

The count k is the data, already observed, and p is the unknown parameter.

Keep k and n fixed at the data, and let p vary. Read this way, the same expression is the likelihood of p, written L(p) and read "the likelihood of p".

```
L(p) = (n choose k) × p^k × (1 - p)^(n - k), with k and n fixed at the data
```

L(p) is the probability of the data you got, if the parameter were p. It is not the probability that the parameter is p.

The term $p^k \times (1 - p)^{(n - k)}$ is the probability of one particular order of k successes and n − k failures. It is a product because the trials are independent (section 3): the probability of all the observations together is the product of their separate probabilities. Without independence, the likelihood is not a product, and this formula is wrong.

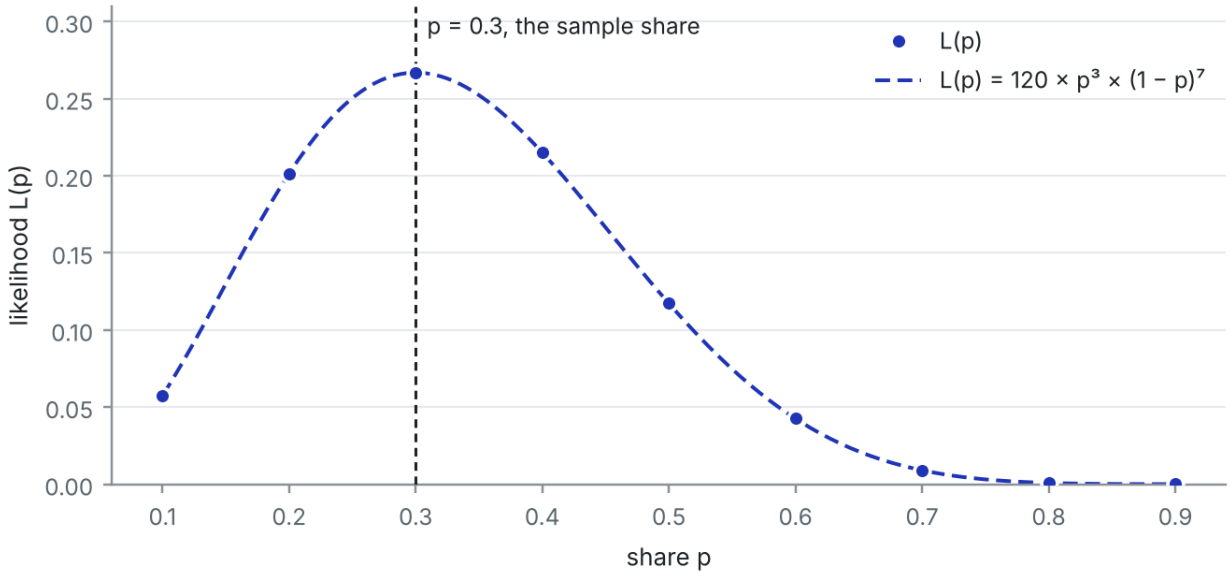
Divide the likelihoods of two values of p, L(p1) ÷ L(p2). The result is the likelihood ratio of section 8. Its two hypotheses are "p = p1" and "p = p2", and the data are the evidence. It says how many times more probable the data are under p1 than under p2. The factor n choose k is the same for every p, so it cancels from every ratio.

The value of p with the largest likelihood is the best-supported value, called the maximum likelihood estimate. For a binomial count it is k ÷ n, the sample share $\hat{p}$ of section 19. That this holds for every n and k is shown in the next rung. [181, 109, 182, 183, 184]

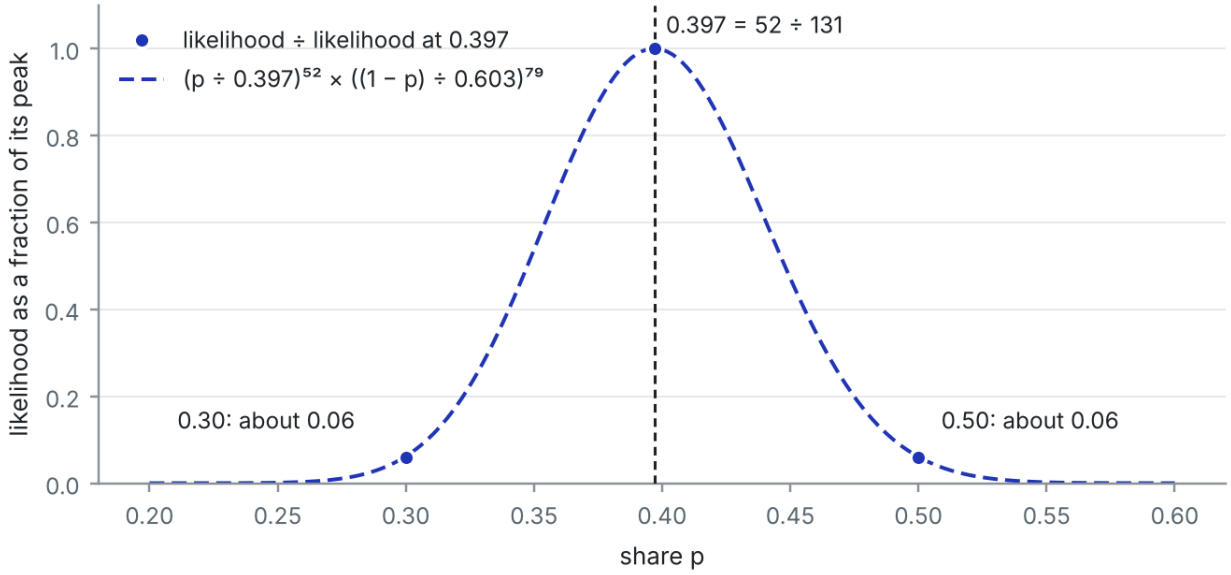
IN PLAIN TERMS. You know how to ask: if the true share were 0.3, how likely is it that 3 of 10 people have the feature?

Now turn it round. You have seen 3 of 10. Try one possible share after another, and ask the same question each time. The answers form the likelihood.

The likelihood is a probability of the data. It is never the probability that a share is true.



The likelihood of the share p after 3 of 10 women had a body-mass index of 25 or more (made-up data), $L(p) = 120 \times p^3 \times (1 - p)^7$. The curve is highest at p = 0.3, the sample share. Each point is the probability of the same data under a different p, so the points are not a distribution for p.



The likelihood of the share p after 52 of 131 men had a body-mass index of 25 or more, as a fraction of its peak at 0.397, the sample share. At 0.30 and at 0.50 it is about 0.06 of the peak, so their ratio is about 1.02 and the data barely tell them apart, though both are far below the peak. At 0.20 it is about 1 in 609,000 of the peak.

Illustration 1. Here is a made-up result. At a screening camp, 3 of 10 women measured have a body-mass index of 25 kg/m^2 or more. Treat the 10 as independent trials with an unknown share p. You want to know which values of p the data support.

First the fixed part. The number of orders of 3 successes among 10 is 10 choose 3 (section 11).

```
10 choose 3 = 10! ÷ (3! × 7!) = (10 × 9 × 8) ÷ (3 × 2 × 1) = 720 ÷ 6 = 120
```

So $L(p) = 120 \times p^3 \times (1 - p)^7$. Work it out at p = 0.3.

```
0.3^3 = 0.027
0.7^7 = 0.0823543
120 times 0.027 times 0.0823543 = 0.2668
```

Do the same at each tenth from 0.1 to 0.9.

p	L(p)
0.1	0.0574
0.2	0.2013
0.3	0.2668
0.4	0.2150
0.5	0.1172
0.6	0.0425
0.7	0.0090
0.8	0.0008
0.9	0.00001

The largest likelihood in the table is at p = 0.3, which is 3 ÷ 10, the sample share.

Now add the column.

```
0.0574 plus 0.2013 plus 0.2668 plus 0.2150 plus 0.1172 = 0.8577
0.0425 plus 0.0090 plus 0.0008 plus 0.00001 = 0.05231
0.8577 plus 0.05231 = 0.91001
```

About 0.91, not 1. Try p = 0.25 and p = 0.35 as well and the total grows past 1. These are not the probabilities of nine possible values of p. Each one is the probability of the same data under a different p.

Compare two values by dividing. The 120 cancels.

```
0.2668 divided by 0.1172 = 2.28
0.2668 divided by 0.0574 = 4.65
```

The data are about 2.3 times as probable if p is 0.3 as if p is 0.5. They are about 4.6 times as probable if p is 0.3 as if p is 0.1. That is all the likelihood says. It does not say how probable 0.3 is.

WHERE THIS PICTURE BREAKS. The camp and its count are made up. The likelihood is only as good as the binomial model behind it: independent women, one shared p. Women from the same household or the same lane share food, work and income, and their results are tied (section 3). Then the product is wrong, and the likelihood comes out more sharply peaked than the data justify.

The table steps in tenths. The best-supported value happens to fall on a tenth here; in general it sits between the rows, at k ÷ n.

Illustration 2. Now real counts. Kiran and colleagues measured 131 male medical students at one college. Open their paper in *Heliyon* at https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/. Table 2 puts 42 men in the body-mass index band 25 to 29.9 kg/m^2 and 10 at 30 or more. So 52 of 131 had an index of 25 or more. The best-supported p is the sample share.

```
52 divided by 131 is about 0.397
```

The authors planned their sample size on a share of 20 per cent, taken from the fourth National Family Health Survey (NFHS-4). How much better do the data support 0.397 than 0.20?

Dividing two powers with the same exponent gives the power of the quotient: $6^2 \div 3^2 = 36 \div 9 = 4 = 2^2$. So the ratio is $(0.397 \div 0.20)^{52} \times (0.603 \div 0.80)^{79}$, because 131 minus 52 is 79 and 1 minus 0.397 is 0.603. The powers are too large for a pocket calculator to show cleanly, so work in $\log_{10}$. A logarithm turns multiplication into addition (Book 0, A7). A power is repeated multiplication, so its log is the power times the log. Check: $2^3 = 8$, and $3 \times 0.30103 = 0.90309$, which is $\log_{10}$ of 8.

```
0.397 divided by 0.2 = 1.985
0.603 divided by 0.8 = 0.75375
52 times 0.29776 = 15.4835
79 times -0.12277 = -9.6988
15.4835 plus -9.6988 = 5.7847
```

Here 0.29776 is $\log_{10}$ of 1.985, and −0.12277 is $\log_{10}$ of 0.75375. The ratio is $10^{5.7847}$, about 609,000. The data are hundreds of thousands of times more probable if the share among the population these men stand for is 0.397 than if it is 0.20.

Now compare 0.30 with 0.50, two values on either side of the peak.

```
0.3 divided by 0.5 = 0.6
0.7 divided by 0.5 = 1.4
52 times -0.22185 = -11.5362
79 times 0.14613 = 11.5443
-11.5362 plus 11.5443 = 0.0081
```

The ratio is $10^{0.0081}$, about 1.02. The data barely tell 0.30 from 0.50. That does not mean the data are vague. Both values are far below the peak: each has a likelihood about 0.06 times that of 0.397 (practice problem 6 works this out). A ratio of two values tells you only about those two.

Figures quoted in this illustration are from [10, 32].

WHERE THIS PICTURE BREAKS. The 131 men volunteered at one college. A likelihood built on them describes the share in a population that resembles them, not the share in young Indian men, and it assumes the men's results are independent of one another.

The 0.20 was a planning figure for a sample size, not a measured share in this college. The huge ratio shows that the college's men did not look like 0.20. It does not show that 0.20 was wrong for the population NFHS-4 described.

The logs are rounded to five figures, which is enough for a ratio quoted to two or three.

MUST KNOW POINTS FOR YOU.

- "The likelihood of p = 0.3 is 0.27, so there is a 27 per cent chance the share is 0.3" is the error. It reads P(data | p) as P(p | data), the transposed conditional of section 9. A likelihood is the probability of the data if p were 0.3. A probability that p is 0.3 needs a prior for p, as a diagnosis did in section 5. That is a later book's subject.
- A small likelihood under one value means nothing until you set it beside the likelihood under a rival value.
- The likelihood multiplies the probabilities of the separate observations, which is right only if they are independent.
- A likelihood ratio near 1 between two values says only that the data do not separate those two.
- When a paper reports a "maximum likelihood estimate" of a simple proportion, check it against k ÷ n.
- Teach the likelihood as the binomial formula read sideways: the same expression, with the data held fixed and p moving.

COMMON MISREADING. "The likelihood is highest at 0.397, so 0.397 is probably the true share." The likelihood ranks values of p by how well they predict the data you got. It puts no probability on any value being true; that would need a prior, as a diagnosis does. Say "the data are best supported by a share of about 0.4", and give the ratio to any rival value you mean to discuss.

EXERCISE 1 (retrieval). From memory: what is the likelihood of a parameter value? Why do likelihoods across p not have to add to 1? What does the ratio of two likelihoods tell you?

EXERCISE 2 (interpretation). A results section says this.

> *The maximum likelihood estimate of the prevalence was 0.397. The likelihood at the national figure of 0.273 was less than 1 in 100 of the maximum, so the national prevalence is excluded.*

The data are Kiran's 52 of 131 men. Check the ratio, and say what the sentence shows and what "excluded" overstates.

PRACTICE. Six problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. In 3 independent trials there were 2 successes. Work out L(p) at p = 0.5 and at p = 0.6. Which value do the data support better?

2. In 10 independent trials there were 6 successes. Work out the likelihood ratio L(0.6) ÷ L(0.4) without working out 10 choose 6.

3. In 4 independent trials there was 1 success. Work out L(p) at p = 0.2, 0.4, 0.6 and 0.8, and say which is best supported. Then work out L at the sample share.

4. In Kiran and colleagues' Table 2, 46 of 151 women had a body-mass index of 25 or more. Of them, 41 were in the band 25 to 29.9 and 5 were at 30 or more. NFHS-6 gives 30.7 per cent for women aged 15 to 49 in India.

Work out the best-supported share for the population these women stand for. Then work out how many times more probable the counts are at that share than at 0.307, using $\log_{10}$. Say what the ratio does and does not show.

5. Here is a worked answer about 3 successes in 10 trials. Find the step that broke.

> `L(0.3) = 120 times 0.3^3 times 0.7^7 = 0.2668`
> `so the probability that p = 0.3 is 0.2668`
> `so there is a 73 per cent chance that p is something other than 0.3`

6. A colleague looks at Kiran's 52 of 131 men and says:

> *The likelihood ratio between a share of 0.30 and a share of 0.50 is about 1. So these data cannot tell us whether the share is nearer a third or a half.*

Decide what to compute, compute it, and say what the claim gets wrong and what it gets right.

22 · Parameter, estimate or prediction: what a reported number is a number of

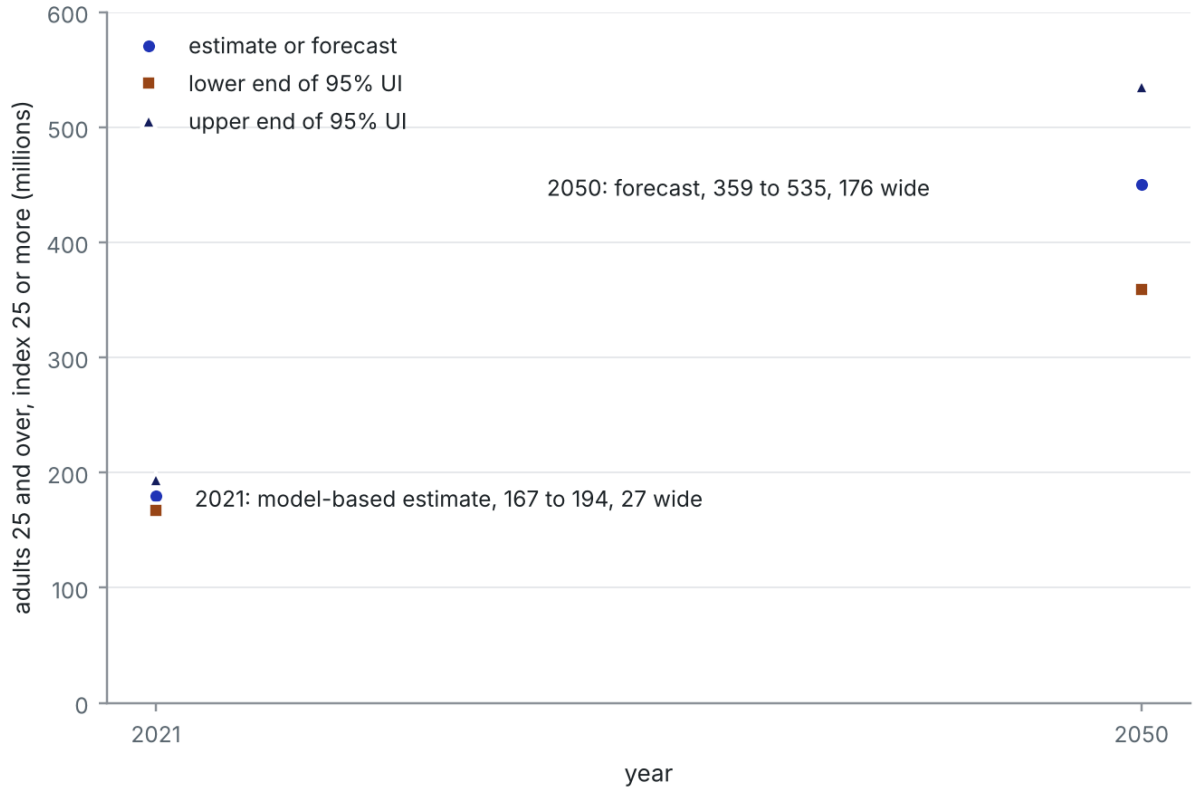
DEFINITION. A reported number about health is one of four kinds.

1. A count of everyone. The number is the group's parameter, in the sense of section 19, with no sampling error.
2. A survey estimate. A sample was drawn from a population and measured, and a rule turned the sample into a number meant to stand for a parameter of that population. It is conditional on the sample and the survey's design.
3. A model-based estimate. A model combines many sources of data, of different kinds and quality, and fills the gaps between them with assumptions. It is conditional on the data that went in and on the model.
4. A prediction. The number is about something not yet observed: a future value, or one person. A forecast is a prediction about a population in the future, conditional on a scenario for how the future goes. A risk score's output for one person, such as a positive predictive value (section 5), is a share of a group like that person. It is then applied to that person. It is conditional on the group the score was built on.

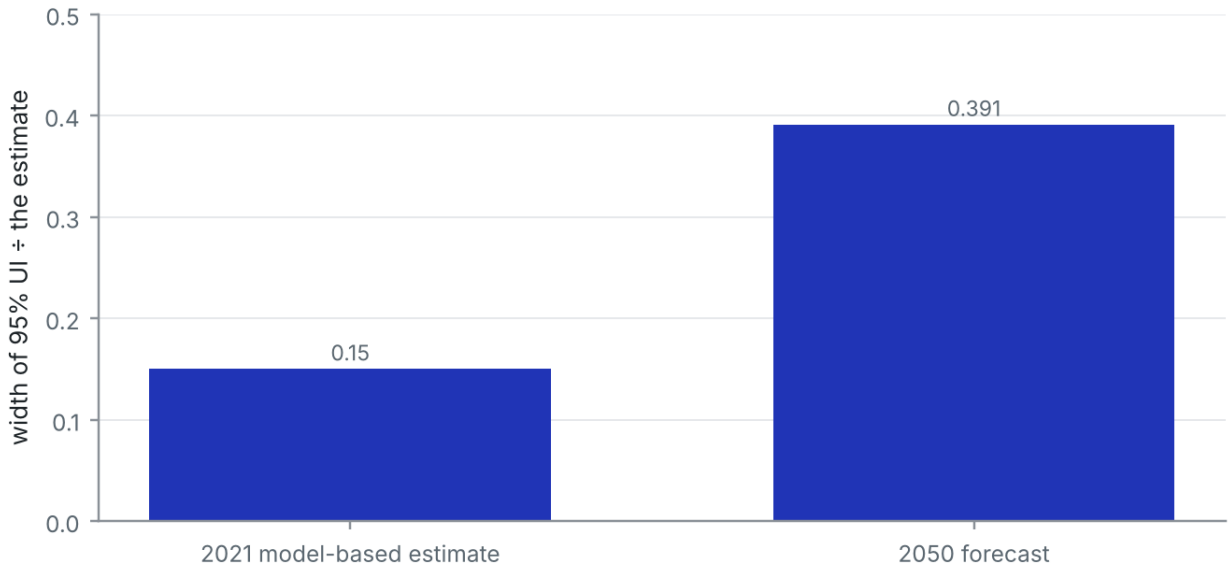
An uncertainty interval printed beside a model-based estimate or a forecast describes the spread of the model's own output. It is conditional on the model being right, and says nothing about errors the model does not contain. [185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198]

IN PLAIN TERMS. Before you use a number, ask what it is a number of.

The four can look the same on a slide: one figure, a per cent sign, perhaps a range in brackets. They are not the same, and each one can be wrong in its own way.



GBD 2021's numbers for India, adults aged 25 and over with overweight and obesity, in millions. In 2021 a model-based estimate of 180 (95% UI 167 to 194, 27 wide). In 2050 a forecast of 450 under the reference scenario (359 to 535, 176 wide). The intervals are ranges of the model's own draws; neither is a measurement.



Each interval's width as a share of its own number. The 2021 model-based estimate's interval is 27 million wide, 0.15 of 180 million; the 2050 forecast's is 176 million wide, about 0.391 of 450 million. Twenty-nine years of forecasting forward widen the range, and even the wider range holds only the model's spread under one scenario.

Illustration 1. Here is a made-up slide of the kind a district meeting sees. Its numbers are real, quoted from the sources below; its last line is the slide-maker's.

> *Overweight and obesity in India: 30.7% of women (NFHS-6). 180 million adults (Lancet, 2025). 450 million by 2050. So the problem will more than double.*

Start with the last line, because it is the one people will repeat. Work out the ratio it rests on.

```
450 divided by 180 = 2.5
```

The arithmetic is right. Now find out what each number is a number of. The last line will not survive it.

The first number. Open the *National Family Health Survey (NFHS-6), 2023-24: India and State/UT Fact Sheets*, from https://www.nfhsiips.in. Find indicator 76.

> *76. Women who are overweight or obese (BMI ≥25.0 kg/m2)15 (%) 42.8 25.5 30.7 24.0*

The four columns are NFHS-6 urban, rural and total, and NFHS-5's total; the 15 printed after the unit is a footnote mark, not part of the value. The third column, 30.7 per cent, is the India total. The survey weighed and measured a sample of women aged 15 to 49. A rule turned that sample into one number. So it is a survey estimate. The parameter it stands for is the share of all women aged 15 to 49 in India with a body mass index of 25 or more, in 2023-24. How the survey's rule weights each woman is beyond this book (section 19). A second sample drawn the same way would have given a slightly different number.

The second number. Open "Global, regional, and national prevalence of adult overweight and obesity, 1990-2021, with forecasts to 2050", by the GBD 2021 Adult BMI Collaborators, in *The Lancet* (2025), at https://doi.org/10.1016/S0140-6736(25)00355-1. Search for "followed by India".

> *followed by India (180 million [167–194])*

That is 180 million adults in 2021, with an interval from 167 to 194 million. Now find out how it was made. Search for "1350".

> *Retrospective and current prevalence trends were derived based on both self-reported and measured anthropometric data extracted from 1350 unique sources*

Nobody counted 180 million people, and no single survey measured them. A model took 1,350 sources from many countries, some of them people's own reports of their weight, and produced a number for every country and year. So this is a model-based estimate. It is about the present, like the survey's, but it rests on the model as well as the data.

Read who it covers. The paper counts adults aged 25 years and older, in 2021. NFHS-6 counts women and men aged 15 to 49, in 2023-24. The two numbers are about different people. Neither can check the other.

The interval. Search for "posterior distribution".

> *95% uncertainty intervals (UIs) for the final estimates were derived based on the 2·5th and the 97·5th percentiles of 1000 draws from the posterior distribution of ST-GPR.*

The journal prints a raised dot as its decimal point, so 2·5 is 2.5. The 2.5th percentile is the value with 2.5 per cent of the draws below it. ST-GPR (spatiotemporal Gaussian process regression) is the paper's model; how it works is beyond this book. A posterior is a model's own probability for the unknown number, built from the data and the model's assumptions; a later book teaches it.

The model produced 1,000 plausible values for India. Put them in order and cut off the lowest 25 and the highest 25. The 95 per cent uncertainty interval (UI) is the range that holds the middle 950. Within the model, 95 per cent of its draws fall inside the interval.

```
1000 times 0.025 = 25
1000 minus 25 minus 25 = 950
194 minus 167 = 27
```

The interval is 27 million wide. It shows how much the model's own output varies. It cannot show an error the model has built in, such as a wrong assumption about how people's own reports of their weight differ from measurements.

Look back at the slide. Its first two numbers sit side by side as if one confirmed the other. One is a survey estimate for women aged 15 to 49 in 2023-24. The other is a model-based estimate for adults aged 25 and over in 2021.

Figures quoted in this illustration are from [32, 199].

WHERE THIS PICTURE BREAKS. The kinds shade into each other. NFHS-6 is a survey, but a survey's rule is itself a small model, with weights and adjustments. A model-based estimate for a country with many good surveys sits close to them. The test is not which word the paper uses. Ask how much of the number came from measuring the people it describes, and how much was filled in.

The slide is made up. The four numbers on it are real.

Illustration 2. The third number. Go back to the GBD paper and search for "India (450".

> *India (450 million [359–535])*

Read the start of the same sentence: "Assuming a reference scenario in which historical trends and patterns continue into the future". This number is about 2050. Nobody has measured it, because 2050 has not happened. It is a forecast, a prediction about a population in the future. Search for "reference scenario, which" to see what the scenario is.

> *a reference scenario, which is a probabilistic forecast of the most likely future based on past trends and relationships*

Then search for "GLP1".

> *we have not built mass scale-up of GLP1 agonists or other rapid advancements in obesity prevention or treatment into our forecast*

So the forecast assumes the past goes on. It leaves out new medicines for obesity and anything else that changes the trend. If a programme works, the forecast was not wrong. Its condition stopped holding.

Compare its interval with the 2021 one.

```
535 minus 359 = 176
27 divided by 180 = 0.15
176 divided by 450 ≈ 0.391
```

The 2021 interval is 27 million wide, 0.15 of its estimate. The 2050 interval is 176 million wide, about 0.39 of its estimate. Twenty-nine years of forecasting forward widen the range. Even that range holds only the model's spread under one scenario.

Now the slide's last line. "Will more than double" turns a forecast under one scenario into a fact about the future. The paper says how many people would have overweight and obesity if the trend goes on. It counts adults aged 25 and over, and it counts overweight and obesity together. Search for "classified as obese".

> *A BMI of 25·0 or higher but less than 30·0 is classified as overweight, while a BMI of 30·0 or higher is classified as obese.*

The 180 and the 450 are not counts of people with obesity. They count everyone with a body mass index of 25 or more.

The fourth kind: a prediction for one person. A man of 52 at your clinic scores 60 or more on the Indian Diabetes Risk Score (IDRS) of section 10. You tell him there is about a one in six chance that he has diabetes that has not been found. Where did that come from? Mohan and colleagues, at https://repository.ias.ac.in/80170/, report it.

> *An IDRS value ≥ 60 had the optimum sensitivity (72.5%) and specificity (60.1%) for determining diabetes. The positive predictive value was 17.0%*

The 17.0 per cent is the share with undiagnosed diabetes among people scoring 60 or more in one Chennai study. Said to one man, it becomes a prediction about him. It holds for him only as far as he is like the people in that study, and as far as your clinic's prior is like theirs, which section 10 showed was about 10 per cent. He either has diabetes or he does not. The 17 per cent describes how often people like him do.

Figures quoted in this illustration are from [199, 98].

WHERE THIS PICTURE BREAKS. A forecast's interval is not the range the future must fall in. It is the range of the model's draws, under a scenario the authors chose. The paper itself says its es-

timates for some countries "rely more heavily on extrapolations influenced by the covariates in the model", and that they "do not reflect subnational variations". So no state or district can read its own number off the national one.

The man of 52 is made up. The IDRS figures are the study's.

Illustration 3. The first kind: a count of everyone. Your urban health centre weighed and measured every man aged 18 to 29 who came to its outpatient clinic from April to September. Suppose, as in the made-up register of section 10, that 96 of the 480 men were in the 25 to 29.9 band.

```
96 divided by 480 = 0.2
```

For those 480 men, in those six months, 0.2 is not an estimate. Every one of them was counted. It is the parameter of that group, with no sampling error at all.

It can still be wrong. A scale that reads 1 kg heavy moves men across the line at 25. A man whose height was copied wrongly lands in the wrong band. A man who came but was never measured is missing from the 480. These are the errors of counting and measuring, the systematic and random errors of Book 0, D7. Counting everyone removes only the error of sampling.

Now use the 0.2 for next month's patients, as section 10 does. It has changed kind. Next month's men were not in the register. The 0.2 is now an estimate of the share among the men who will come, and it is only as good as the guess that they resemble the last six months'.

WHERE THIS PICTURE BREAKS. A count of everyone is a parameter only for the group and period it counted. The moment it is used for anyone else, including the same clinic next month, it is an estimate.

The register is made up.

MUST KNOW POINTS FOR YOU.

- "It has a 95 per cent interval in brackets, so somebody measured it" is the error. Before you quote a number, find the sentence in the methods that says how it was made.
- Ask one question of each kind before you use it. A count: who was missed, and how was each person measured? A survey estimate: which population, in which years, from what sample? A model-based estimate: which data went in, and what did the model fill in? A prediction: under what scenario, and for people like whom?
- Do not set a survey estimate beside a model-based estimate as if one confirmed the other. Check the age range, the cut-off and the method first.
- An uncertainty interval shows the spread of the model's own output, under the model's own assumptions. Do not say there is a 95 per cent chance the true number lies inside it. The 95 per cent belongs to the model's draws and holds only if the model is right. The interval cannot show an error the model has built in.
- A forecast says what would happen if its scenario held. Quote the scenario with the number. Never present a forecast as what will happen, or use it to judge a programme designed to change the trend.

- A risk score's output for one person is a share of a group like that person. Tell the man with an IDRS of 60 or more "about 17 in 100 people like you, in the study, had diabetes not yet found". Do not say "you are 17 per cent diabetic".
- A national figure, whether survey or model, is not a figure for your state or district. Find the state or district fact sheet, or your own register, before you plan for a district.

COMMON MISREADING. "GBD 2021 gives India 180 million adults with overweight and obesity, 95 per cent UI 167 to 194 million. So there is a 95 per cent chance that the true number lies between 167 and 194 million." The interval is the range of the middle 950 of 1,000 values drawn from the model. It is a statement about the model's output, made on the assumption that the model is right. Suppose the model corrects people's own reports of their weight wrongly. Then the true number can lie outside the interval, and the interval would not show it.

REPORTING IT. The Global Burden of Disease 2021 study used a model built on 1,350 data sources. It estimated that 180 million adults aged 25 and over in India had overweight or obesity in 2021 (95% uncertainty interval 167 to 194 million). This is a model's estimate, not a count. It cannot be compared with NFHS-6 (2023-24), which measured women and men aged 15 to 49.

EXERCISE 1 (retrieval). From memory: name the four kinds of reported number, say what each is conditional on, and give the one question you ask of each.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (interpretation). A district health report, made up for this exercise, contains these five sentences. For each number, say which of the four kinds it is, and what it is conditional on.

1. "Of the 1,212 schoolchildren weighed in our district's school health camp this year, 97 had a body mass index above the cut-off."
2. "NFHS-6 found 27.3% of men aged 15-49 in India overweight or obese."
3. "By 2050, 450 million Indian adults will have overweight or obesity (Lancet)."
4. "A woman with an IDRS of 60 or more has a 17% chance of undiagnosed diabetes."
5. "In 2021, India had 180 million adults with overweight or obesity (95% UI 167-194 million)."

EXERCISE 3 (critique). A newspaper headline reads: "India will have 450 million obese adults by 2050, says Lancet study". Say everything that is wrong with it, using only what the paper itself says.

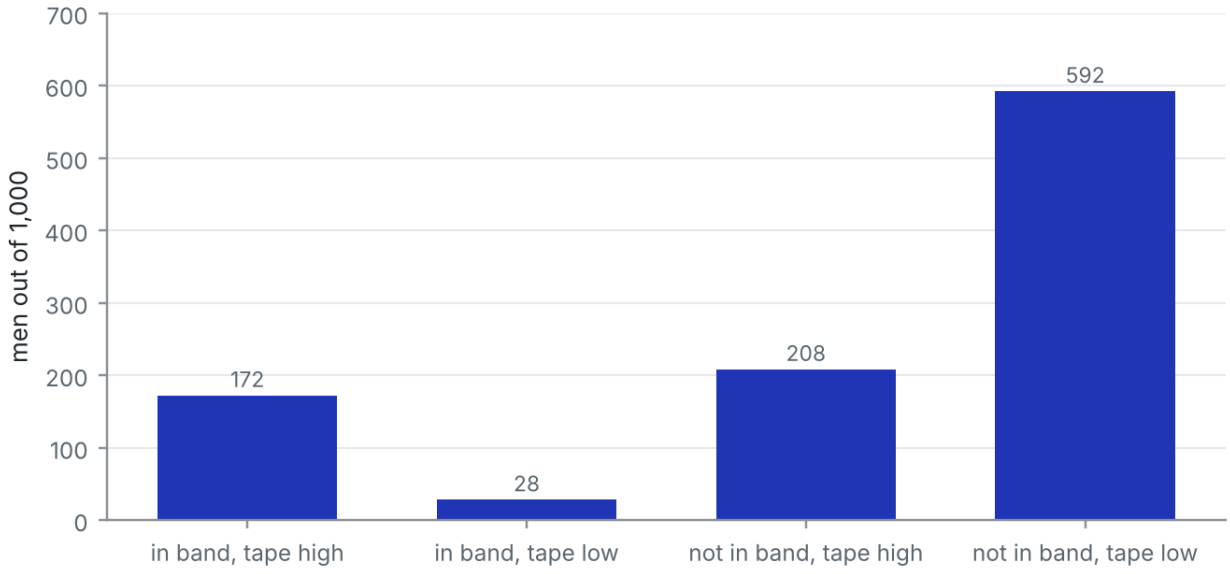
EXERCISE 4 (teaching). A journalist rings. "One survey says three in ten Indian women are overweight. A Lancet paper says 180 million Indians. Which is right?" You have three minutes. Give her one sentence she can quote, and say how sure you are.

23 · Worked journey: One clinic test, from the question to a handout a patient can read

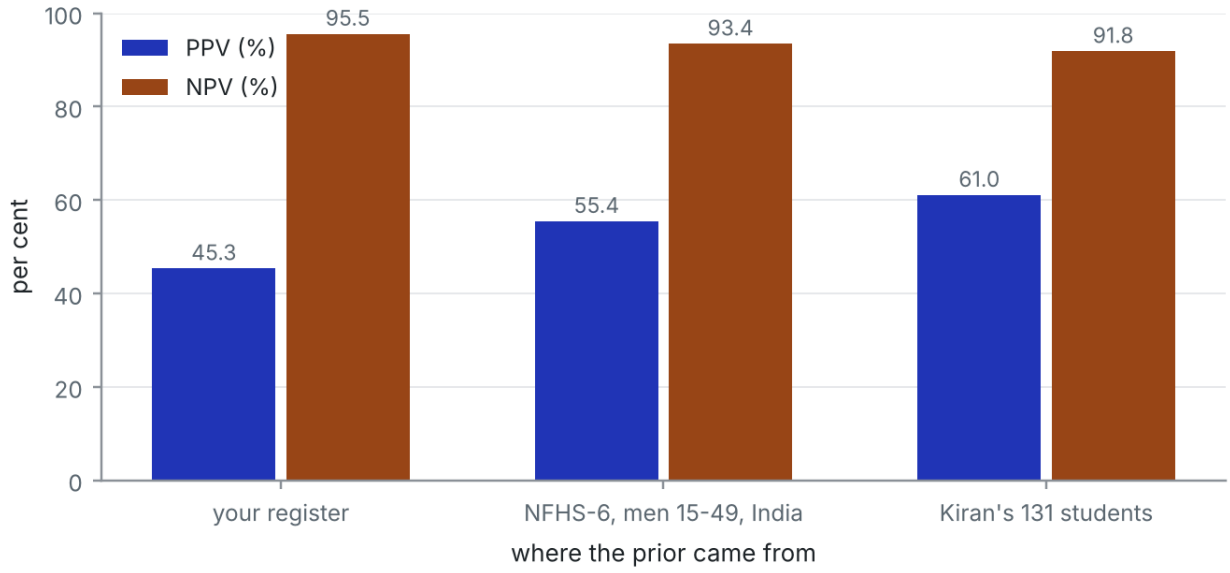
DEFINITION. In this book, a test's journey is the order of work that takes one clinic test from the question a patient asks to a one-page handout that answers it.

1. The question: what a positive or a negative result tells a person like this patient, in this clinic, about one stated condition.
2. The test's figures: its sensitivity and specificity, read from the paper with the condition, the cut-off and the people they were measured on.
3. The prior: the prevalence of the same condition among the people the test will be used on, counted from your own register and passed through the four checks of section 10.
4. The calculation, three ways that must agree: a frequency tree of 1,000 people (section 7), Bayes' theorem as a formula (section 5), and the odds form with the likelihood ratios (section 8).
5. The prior moved: the same calculation at the other priors someone will offer, and the reason each one does or does not belong.
6. The sentence a colleague will say, which transposes the conditional (section 9), and its correction.
7. The checks: counts that add up, routes that agree, and no more digits than the inputs hold.
8. The handout, and a statement of what it does not establish. [185, 200, 201, 202, 203, 204, 205, 206, 207]

IN PLAIN TERMS. Each earlier section taught one tool. Here the test is the arm tape of Book 0. The patient is a young man at your clinic whose arm measures over the cut-off. The work ends in a handout he can read.



1,000 young men at the clinic, with a prior of 0.2. The tape flags 380: 172 in the 25 to 29.9 band and 208 not in it, so a high reading is right about 45 times in 100. Of the 620 with a low reading, 592 are not in the band.



The same tape at three priors. From your register (0.2), a high reading is right 45.3 times in 100 and a low one 95.5. NFHS-6's men (0.273) would give 55.4 and 93.4, and the study's own students (about 0.32) 61.0 and 91.8. Only the register passes the four checks for the man's question.

Illustration 1. The question. A man of 24 at your urban health centre asks why the nurse measured his arm, and what it means that it came out "high". Choose the test you will write the handout for. Here it is the arm tape of Book 0, D3: a mid-upper arm circumference (MUAC) of 31.3 cm or more counts as positive. The Indian Diabetes Risk Score of section 10 would do as well. For your own handout, take a test your clinic actually uses.

Write the question down in one sentence before you look up any number.

> *Of young men at this clinic whose arm measures 31.3 cm or more, how many have a body mass index in the band 25 to 29.9? And of those whose arm measures less?*

The condition is named exactly. The question is about men like him, in this clinic. It asks for P(in the band | tape positive), the probability of being in the band given a positive tape, and the same for a negative tape.

The test's figures. Open "Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India" by Kiran and colleagues, at https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/. Search for "31.3cm".

> *The MUAC cut-off was 31.3cm for men (sensitivity: 86%; specificity: 74%) to detect overweight*

The label says "overweight", a body mass index of 25 or more. Book 0 divided the percentages back and found they fit the 42 men in the 25 to 29.9 band, against the 89 outside it. So the condition this test was measured against is the band. Your question already says so.

Now say what kind of number each figure is (section 22). The 86 per cent is 36 of 42 men, and the 74 per cent is 66 of 89, in one sample of medical students in one college. Each is an estimate of a parameter, the tape's sensitivity or specificity among young men like them (section 19). A different sample of students would have given slightly different counts. As section 10 showed, one man more or less among the 42 moves the final answer by about a percentage point.

Write down three more limits now, before you calculate.

- The men were medical students. Search for "single centre".

  > *The study use single centre measurements consisting of only young adults and this is the most important limitation.*

- The cut-off was chosen in the same sample. Search for "maximum values of Youden".

  > *After defining the Youden index for all points of the ROC curve, the maximum values of Youden index were utilized to guide the best cut-off of MUAC and NC.*

  ROC means receiver operating characteristic. The ROC curve is the list of sensitivity and specificity pairs the tape gives at every possible cut-off; the series' book on screening teaches it. The Youden index is sensitivity plus specificity minus 1. The authors tried every cut-off and kept the one that scored best on these same men. Some of that score was luck in this sample. In a new sample the same cut-off will tend to do a little worse. How much worse is a question for the series' book on screening.

- The paper's label and its numbers disagree, as Book 0 found. Any handout built on it must say it uses the 25 to 29.9 band.

Figures quoted in this illustration are from [10].

WHERE THIS PICTURE BREAKS. The arm tape is used here because Book 0 rebuilt its table and this book's Bayes sections used it. It is not a recommendation to screen with it. Whether arm circumference is a good screening test for anything is a question this book does not answer.

The man of 24 is made up.

Illustration 2. The prior. Count it from your own register, as section 10 set out. Here, as there, a made-up register stands in for yours. Your centre weighed and measured every man aged 18 to 29 who came to the outpatient clinic from April to September: 480 men, 96 of them in the 25 to 29.9 band.

```
96 divided by 480 = 0.2
```

Run the four checks before you use it. The condition is the band, as in the study. The denominator is the men the tape will be used on. The men are young outpatients, like the man of 24. Every man in the period was measured, not only those who looked heavy. All four pass, so the prior is 0.2. Altman and Bland give the prior its name in their note "Diagnostic tests 2: predictive values", at https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/.

> *The prevalence can be interpreted as the probability before the test is carried out that the subject has the disease, known as the prior probability of disease.*

First route: count 1,000 men (section 7). Split by the prior, then each part by the test.

```
0.2 times 1000 = 200
1000 minus 200 = 800
0.86 times 200 = 172
200 minus 172 = 28
0.26 times 800 = 208
800 minus 208 = 592
172 plus 208 = 380
28 plus 592 = 620
```

1,000 young men at your clinic	in the band	not in the band	total
tape 31.3 cm or more	172	208	380
tape under 31.3 cm	28	592	620
total	200	800	1,000

Read the two answers off the rows.

```
172 divided by 380 ≈ 0.4526
592 divided by 620 ≈ 0.9548
```

Of 380 men with a high tape, 172 are in the band: a positive predictive value (PPV) of about 45 per cent. Of 620 men with a low tape, 592 are not: a negative predictive value (NPV) of about 95.5 per cent.

Second route: the formula (section 5).

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
0.172 divided by 0.38 = 0.4526
0.74 times 0.8 = 0.592
0.14 times 0.2 = 0.028
0.592 plus 0.028 = 0.62
0.592 divided by 0.62 = 0.9548
```

The same two numbers. The formula is the frequency tree with 1,000 divided out.

Third route: odds and likelihood ratios (section 8). Deeks and Altman, in "Diagnostic tests 4: likelihood ratios", at https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/, give the rule.

> *The post-test odds that the patient has the disease are estimated by multiplying the pretest odds by the likelihood ratio.*

The positive likelihood ratio, LR+, is sensitivity ÷ (1 − specificity). The negative one, LR−, is (1 − sensitivity) ÷ specificity.

```
0.86 divided by 0.26 = 3.3077
0.14 divided by 0.74 = 0.1892
0.2 divided by 0.8 = 0.25
0.25 times 3.3077 = 0.8269
0.8269 divided by 1.8269 = 0.4526
0.25 times 0.1892 = 0.0473
0.0473 divided by 1.0473 = 0.0452
1 minus 0.0452 = 0.9548
```

After a positive tape the odds are 0.8269, a probability of 0.4526. After a negative tape the probability of being in the band is 0.0452, so the NPV is 0.9548. Three routes, one answer. If one of them had disagreed, you would have a slip to find before you wrote a word of the handout.

WHERE THIS PICTURE BREAKS. The three routes agree because they are one theorem written three ways. Their agreement checks your arithmetic. It does not check the inputs: a wrong prior or a sensitivity from the wrong group gives the same wrong answer by all three.

The register is made up. Use your own.

Illustration 3. The prior moved. Someone at the meeting will offer a different prior. Run the same formula at each, then decide which belongs.

The national survey. The National Family Health Survey (NFHS-6, 2023-24) India fact sheet gives this for men aged 15 to 49.

> *77. Men who are overweight or obese (BMI ≥25.0 kg/m2) (%) 36.3 23.0 27.3 22.9*

The columns are NFHS-6 urban, rural and total, then NFHS-5's total (2019-21). The India total is 27.3 per cent, a prior of 0.273.

```
0.86 times 0.273 = 0.23478
1 minus 0.273 = 0.727
0.26 times 0.727 = 0.18902
0.23478 plus 0.18902 = 0.4238
0.23478 divided by 0.4238 = 0.554
0.74 times 0.727 = 0.53798
0.14 times 0.273 = 0.03822
0.53798 plus 0.03822 = 0.5762
0.53798 divided by 0.5762 = 0.9337
```

The study's own men: 42 of 131, a prior of about 0.32.

```
42 divided by 131 ≈ 0.3206
0.86 times 0.3206 = 0.275716
1 minus 0.3206 = 0.6794
0.26 times 0.6794 = 0.176644
0.275716 plus 0.176644 = 0.45236
0.275716 divided by 0.45236 = 0.6095
0.74 times 0.6794 = 0.502756
0.14 times 0.3206 = 0.044884
0.502756 plus 0.044884 = 0.54764
0.502756 divided by 0.54764 = 0.918
```

where the prior came from	prior	PPV (%)	NPV (%)	passes the four checks?
your register	0.2	45.3	95.5	yes
NFHS-6, men 15-49, India	0.273	55.4	93.4	no: a different condition and different men
Kiran's 131 students	about 0.32	61.0	91.8	no: students in one college

The NFHS-6 row counts a body mass index of 25 or more, not the band, and all of India's men aged 15 to 49, not your clinic's. It is a survey estimate of a different parameter (section 22). The study's 0.32 belongs to medical students in Mangaluru. Only your register answers the man's question. The table also shows the size of the mistake: the wrong prior would print a PPV up to 16 percentage points too high. Altman and Bland put the rule in one line: "Predictive values observed in one study do not apply universally."

The sentence a colleague will say. A senior resident reads the handout draft and says:

> *The tape is 86 per cent sensitive. So a man with a high tape is 86 per cent likely to be overweight.*

That is the transposed conditional of section 9. Correct it on the spot, in the four steps of that section.

1. Restate both in words. 86 per cent is P(tape positive | in the band): of men in the band, how many the tape catches. The man's question is P(in the band | tape positive).
2. Ask "out of whom?" The 86 per cent is out of men in the band. His answer is out of men with a high tape.
3. Count 1,000 men: 200 in the band, 172 caught. But 208 men outside the band are caught too.
4. Say which number answers the question: 172 of 380, about 45 per cent.

And "overweight" has slipped from the band to the word on the paper's label. Keep the band.

The checks. The four cells add to 1,000. The three routes agree to four decimal places. The inputs are whole per cents from 42 and 89 men, and one man moves the PPV by about a point. So print "about 45 in 100" and "about 95 in 100", and nothing after a decimal point.

Figures quoted in this illustration are from [32].

WHERE THIS PICTURE BREAKS. The table's PPVs and NPVs are arithmetic on the stated priors. Only the first row is a prior for your clinic, and even it is an estimate of next month's share from the last six months' (section 22).

Illustration 4. The handout. Put it together. The prior line follows the template of section 10. The counts are the frequency tree, because counts are what patients read best (section 7).

> ***Your arm measurement: what a high reading means***
>
> *The nurse measured round your upper arm. For young men, 31.3 cm or more counts as a high reading. It is a quick check, not a diagnosis.*
>
> *Think of 1,000 young men who come to this clinic. About 200 of them have a body mass index (weight for height) between 25 and 29.9.*
>
> - *About 380 of the 1,000 have a high arm reading. About 172 of those 380 have a body mass index in that range: roughly 45 in 100.*
> - *About 620 have a reading under 31.3 cm. About 592 of those 620 do not have a body mass index in that range: roughly 95 in 100.*
>
> *So a high reading means the next step is to check your weight and height. On its own, it is not the answer.*
>
> *Where these numbers come from: 96 of 480 young men measured at this clinic from April to September were in the 25 to 29.9 range. The arm check's accuracy comes from one study of 131 male medical students in Mangaluru (Kiran and colleagues, 2022). Both figures are estimates; treat them as approximate.*

What it does not establish. Write these in a box at the foot of the staff copy.

- It answers a question about the 25 to 29.9 band only. A man with a body mass index of 30 or more counts as outside the band here, because that is how the study's figures fit. For "25 or more" the paper's figures cannot be rebuilt.
- The sensitivity and specificity come from young medical students in one college, measured by one observer, with the cut-off chosen on those same men. Your patients are not that sample.
- Both inputs and the prior are estimates. The 45 is "about 45", give or take a few points.
- It says nothing about whether a high body mass index in this man carries any risk to his health, or whether screening with an arm tape does any good. Those are questions for the books on screening and on risk.
- It describes men like the clinic's last six months of patients. If the clinic changes, such as a new outreach camp in a richer ward, count the prior again.

WHERE THIS PICTURE BREAKS. A handout is tested by a reader, not by its author. Give it to two patients who have not seen it and ask each to say what a high reading means. If either says "it means I am overweight", the counts are not doing their work, and the first two lines need rewriting.

MUST KNOW POINTS FOR YOU.

- Write the question before you look up a number: what a result tells a person like this patient, in this clinic, about one named condition.
- Take the prior from your own register, passed through the four checks. Never take it from a national survey or the study's own sample. At this clinic's prior of 0.2, a high reading means about 45 in 100.
- Work the PPV and NPV three ways: frequency tree, formula, and likelihood ratios. If one route disagrees, find the slip before you write the handout. Agreement checks the arithmetic, never the inputs.
- "The tape is 86 per cent sensitive, so a man with a high tape is 86 per cent likely to be overweight" is the transposed conditional. Correct it with counts: of 380 men with a high tape, 172 are in the 25 to 29.9 band, and the band, not "overweight", is what the 86 per cent was measured against.
- A cut-off chosen as the best one in a sample will tend to do a little worse in the next sample. Check whether the paper picked its cut-off on the same people it reports.
- Print "about 45 in 100", not "45.26 per cent".

COMMON MISREADING. "Three routes gave the same 45 per cent, so the handout's number is right." The three routes are one theorem written three ways, so they always agree when the arithmetic is right. They check the arithmetic. A prior from the wrong group, or a sensitivity

measured on the wrong condition, gives the same wrong answer by all three. Check the inputs against the question separately.

REPORTING IT. Pre-test probability: 96 of 480 men aged 18 to 29, examined at [clinic] from April to September, had a body mass index of 25 to 29.9 (20 per cent). Kiran and colleagues (2022) report a sensitivity of 86 per cent and a specificity of 74 per cent for an arm circumference of 31.3 cm or more. With these, a high reading raises the probability to about 45 per cent. A low reading lowers it to about 5 per cent. All three inputs are estimates; treat the results as approximate.

EXERCISE 1 (retrieval). From memory, list the steps of a test's journey in order, from the question to what the handout does not establish. For each, name the earlier section whose tool it uses, or the check you make before moving on.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A colleague's draft handout for the Indian Diabetes Risk Score (IDRS) says this.

> *A score of 60 or more means you have a 72.5% chance of having diabetes. We used the national prevalence of diabetes as our starting point.*

Find every fault, and say what the line should say instead. Use only figures you can trace to section 10.

EXERCISE 3 (teaching). A first-year resident has ten minutes and a whiteboard. Explain why a high arm reading at your clinic means "about 45 in 100", when the paper says the tape is 86 per cent sensitive.

EXERCISE 4 (build). This book's build: a one-page handout on Bayes' theorem for one diagnostic test used in your own clinic.

1. Write the question, naming the condition exactly as the test's paper defines it.
2. Read the sensitivity and specificity from the paper, with their sample and cut-off, and write down every limit the paper states.
3. Count the prior from your clinic's register, and pass it through the four checks.
4. Work the PPV and NPV by a 1,000-person frequency tree, by formula and by likelihood ratios.
5. Work them again at the national figure and at the study's own prevalence, and say why your register's prior is the one used.
6. Write the patient's page in counts, with a sources line.
7. Write the transposed sentence a colleague is most likely to say, and your correction, in the four steps of section 9.
8. List what the handout does not establish.
9. Give the page to two patients who have not seen it, and revise it if either misreads it.

References

1. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.1 Terminology, second paragraph (outcome and sample space). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).
2. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.1 Terminology, paragraph defining an event. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).
3. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.1.4 Probabilities when events are not disjoint, the TIP box "or" is inclusive. https://www.openintro.org/book/os/ (accessed 2026-10-09).
4. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.5 Tree and Venn Diagrams, subsection Venn Diagram, first paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).
5. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.1.5 Probability distributions, the box RULES FOR PROBABILITY DISTRIBUTIONS. https://www.openintro.org/book/os/ (accessed 2026-09-24).
6. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §1.2 Discrete Probability Distributions, Theorem 1.1 (Properties), CHANCE version p. 22. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).
7. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.1.3 Disjoint or mutually exclusive outcomes, first sentence. https://www.openintro.org/book/os/ (accessed 2026-10-09).
8. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.1.4, the box GENERAL ADDITION RULE. https://www.openintro.org/book/os/ (accessed 2026-10-09).
9. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §1.2, subsection Odds, first paragraph, CHANCE version pp. 27-28. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).
10. Kiran, R and Harshitha and Bhargava, Madhavi. Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India. Open access, PMC9791811. Used in Book 0 Part D as real material to compute with, not as evidence about screening. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/ (accessed 2026-09-23).
11. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.3 Defining conditional probability, the box CONDITIONAL PROBABILITY. https://www.openintro.org/book/os/ (accessed 2026-10-09).
12. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.3, the paragraph after the vertical bar is introduced. https://www.openintro.org/book/os/ (accessed 2026-10-09).
13. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.1 Terminology, the paragraph on conditional probability (the formula and its condition). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).
14. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.2 Marginal and joint probabilities, the box MARGINAL AND JOINT PROBABILITIES. https://www.openintro.org/book/os/ (accessed 2026-10-09).
15. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.5 General multiplication rule, the sentence after the box. https://www.openintro.org/book/os/ (accessed 2026-10-09).
16. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.7 Tree diagrams, the paragraph after Figure 3.17. https://www.openintro.org/book/os/ (accessed 2026-10-09).

17. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.5 Tree and Venn Diagrams, subsection Tree Diagrams, first paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

18. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.2 Independent and Mutually Exclusive Events, first line. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

19. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.2, subsection Independent Events, the paragraph after the three conditions. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

20. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1 Discrete Conditional Probability, subsection Independent Events, Definition 4.1 and Theorem 4.1, CHANCE version pp. 139-140. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

21. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.2, subsection Mutually Exclusive Events, first paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

22. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §3.2, the bullet Without replacement. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

23. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.3 Sampling from a small population, the paragraph after Guided Practice 3.51. https://www.openintro.org/book/os/ (accessed 2026-10-09).

24. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.3, the last paragraph. https://www.openintro.org/book/os/ (accessed 2026-10-09).

25. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.6 Independence considerations in conditional probability, first sentence. https://www.openintro.org/book/os/ (accessed 2026-10-09).

26. International Institute for Population Sciences. National Family Health Survey (NFHS-6), 2023-24: India and State/UT Fact Sheets. 2026. India Key Indicators table, indicators 76 and 77, and footnote 15. https://www.nfhsiips.in (accessed 2026-10-10).

27. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. figures quoted in the text. https://www.openintro.org/book/os/ (accessed 2026-10-09).

28. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1 Discrete Conditional Probability, subsection Bayes' Formula, the sentence leading to equation (4.3). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

29. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, subsection Bayes' Formula, the opening of the general case. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

30. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.8 Bayes' Theorem, Example 3.42 (mammogram), the step before P(mammogram+) is worked out. https://www.openintro.org/book/os/ (accessed 2026-10-09).

31. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.7 Tree diagrams, Example 3.40, the sentence after P(final = A) is worked out. https://www.openintro.org/book/os/ (accessed 2026-10-09).

32. International Institute for Population Sciences. National Family Health Survey (NFHS-6), 2023-24: India and State/UT Fact Sheets. 2026. figures quoted in the text. https://www.nfhsiips.in (accessed 2026-10-10).

33. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1 Discrete Conditional Probability, subsection Bayes' Formula, the paragraph defining prior and posterior probabilities. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

34. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, subsection Bayes' Formula, the paragraph after the formula. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

35. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.2.8 Bayes' Theorem, the paragraph after the box BAYES' THEOREM: INVERTING PROBABILITIES. https://www.openintro.org/book/os/ (accessed 2026-10-09).

36. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. BMJ 1994;309:102, Statistics Notes, Diagnostic tests 2: predictive values, the definitions. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

37. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. BMJ 1994;309:102, the paragraph on prior and posterior probability. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

38. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. BMJ 1994;309:102, the paragraph after the liver-scan predictive values. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

39. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B. Interpretation by physicians of clinical laboratory results. 1978. Methods, first paragraph (the question), p. 999. https://doi.org/10.1056/NEJM197811022991808 (accessed 2026-10-09).

40. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B. Interpretation by physicians of clinical laboratory results. 1978. Methods, first paragraph (the 60 participants), p. 999. https://doi.org/10.1056/NEJM197811022991808 (accessed 2026-10-09).

41. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B. Interpretation by physicians of clinical laboratory results. 1978. Methods, second paragraph (the authors' worked answer), p. 999. https://doi.org/10.1056/NEJM197811022991808 (accessed 2026-10-09).

42. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B. Interpretation by physicians of clinical laboratory results. 1978. Results, first paragraph, p. 999. https://doi.org/10.1056/NEJM197811022991808 (accessed 2026-10-09).

43. Manrai, Arjun K and Bhatia, Gaurav and Strymish, Judith and Kohane, Isaac S and Jain, Sachin H. Medicine's uncomfortable relationship with math: calculating positive predictive value. 2014. Results, first paragraph (NIH author manuscript). https://pmc.ncbi.nlm.nih.gov/articles/PMC4955674/ (accessed 2026-10-09).

44. Manrai, Arjun K and Bhatia, Gaurav and Strymish, Judith and Kohane, Isaac S and Jain, Sachin H. Medicine's uncomfortable relationship with math: calculating positive predictive value. 2014. Methods (NIH author manuscript). https://pmc.ncbi.nlm.nih.gov/articles/PMC4955674/ (accessed 2026-10-09).

45. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B. Interpretation by physicians of clinical laboratory results. 1978. figures quoted in the text. https://doi.org/10.1056/NEJM197811022991808 (accessed 2026-10-09).

46. Manrai, Arjun K and Bhatia, Gaurav and Strymish, Judith and Kohane, Isaac S and Jain, Sachin H. Medicine's uncomfortable relationship with math: calculating positive predictive value. 2014. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC4955674/ (accessed 2026-10-09).

47. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. Abstract, Method. https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

48. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. Abstract, Results. https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

49. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. Introduction, p. 538, last paragraph. https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

50. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. p. 539, first paragraph (PPV for natural frequencies). https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

51. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. Results, p. 539, non-Bayesian strategies. https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

52. Hoffrage, Ulrich and Gigerenzer, Gerd. Using natural frequencies to improve diagnostic inferences. 1998. figures quoted in the text. https://pure.mpg.de/rest/items/item_2103015_6/component/file_2103014/content (accessed 2026-10-09).

53. Gigerenzer, Gerd and Hoffrage, Ulrich. How to improve Bayesian reasoning without instruction: frequency formats. 1995. figures quoted in the text. https://pure.mpg.de/rest/items/item_2543495_7/component/file_2562163/content (accessed 2026-10-09).

54. Hoffrage, Ulrich and Lindsey, Samuel and Hertwig, Ralph and Gigerenzer, Gerd. Communicating statistical information. 2000. figures quoted in the text. https://pure.mpg.de/rest/items/item_2102628_7/component/file_2102627/content (accessed 2026-10-09).

55. Kirk, Donna and OpenStax. Contemporary Mathematics. 2023. §7.7 What Are the Odds?, subsection Odds as a Ratio of Probabilities, first sentence. https://openstax.org/books/contemporary-mathematics/pages/7-7-what-are-the-odds (accessed 2026-09-23).

56. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §1.2, subsection Odds, the paragraph before Definition 1.4. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

57. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, subsection Bayes' Formula, second paragraph. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

58. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. Second paragraph, the definition of a likelihood ratio. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

59. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. Paragraph after the box, on Bayes' theorem. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

60. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. Paragraph after the box, the two likelihood ratios from sensitivity and specificity. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

61. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. Paragraph on the liver-scan odds, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

62. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

63. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, subsection Bayes Probabilities, first paragraph. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

64. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, subsection Bayes' Formula, the paragraph on medical diagnosis before Example 4.16. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

65. Royal Statistical Society. Royal Statistical Society concerned by issues raised in Sally Clark case [news release; statement regarding statistical issues in the Sally Clark case]. 2001. Statement of 23 October 2001, fourth para-

graph. http://web.archive.org/web/20011211220114/http://www.rss.org.uk/archive/reports/sclark.html (accessed 2026-10-09).

66. Royal Statistical Society. Royal Statistical Society concerned by issues raised in Sally Clark case [news release; statement regarding statistical issues in the Sally Clark case]. 2001. Statement of 23 October 2001, fourth paragraph, last sentence. http://web.archive.org/web/20011211220114/http://www.rss.org.uk/archive/reports/sclark.html (accessed 2026-10-09).

67. Royal Statistical Society. Royal Statistical Society concerned by issues raised in Sally Clark case [news release; statement regarding statistical issues in the Sally Clark case]. 2001. Statement of 23 October 2001, second paragraph. http://web.archive.org/web/20011211220114/http://www.rss.org.uk/archive/reports/sclark.html (accessed 2026-10-09).

68. Hill, Ray. Multiple sudden infant deaths – coincidence or beyond coincidence?. 2004. p. 325, Concluding remarks, Objection 3. https://doi.org/10.1111/j.1365-3016.2004.00560.x (accessed 2026-10-09).

69. Hill, Ray. Multiple sudden infant deaths – coincidence or beyond coincidence?. 2004. p. 325, Objection 2. https://doi.org/10.1111/j.1365-3016.2004.00560.x (accessed 2026-10-09).

70. Hill, Ray. Multiple sudden infant deaths – coincidence or beyond coincidence?. 2004. p. 321, Definition of SIDS (Beckwith's 1969 definition, as Hill quotes it). https://doi.org/10.1111/j.1365-3016.2004.00560.x (accessed 2026-10-09).

71. Hill, Ray. Multiple sudden infant deaths – coincidence or beyond coincidence?. 2004. p. 325, Objection 1. https://doi.org/10.1111/j.1365-3016.2004.00560.x (accessed 2026-10-09).

72. Court of Appeal (Criminal Division). R v Clark [2003] EWCA Crim 1020. 2003. Para 178. https://case-law.nationalarchives.gov.uk/ewca/crim/2003/1020 (accessed 2026-10-09).

73. Court of Appeal (Criminal Division). R v Clark [2003] EWCA Crim 1020. 2003. Para 177. https://case-law.nationalarchives.gov.uk/ewca/crim/2003/1020 (accessed 2026-10-09).

74. Court of Appeal (Criminal Division). R v Clark [2003] EWCA Crim 1020. 2003. Para 6, the grounds of appeal. https://caselaw.nationalarchives.gov.uk/ewca/crim/2003/1020 (accessed 2026-10-09).

75. Court of Appeal (Criminal Division). R v Clark [2003] EWCA Crim 1020. 2003. Para 180. https://case-law.nationalarchives.gov.uk/ewca/crim/2003/1020 (accessed 2026-10-09).

76. Government of India. The Bharatiya Sakshya Adhiniyam, 2023. 2023. s.2(1)(j), definition of "proved". Gazette of India, via the Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf (accessed 2026-10-09).

77. Government of India. The Bharatiya Sakshya Adhiniyam, 2023. 2023. s.39(1), opinions of experts. Gazette of India, via the Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf (accessed 2026-10-09).

78. Government of India. The Bharatiya Sakshya Adhiniyam, 2023. 2023. s.45, grounds of opinion. Gazette of India, via the Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf (accessed 2026-10-09).

79. Government of India. The Bharatiya Sakshya Adhiniyam, 2023. 2023. s.170(1), repeal. Gazette of India, via the Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf (accessed 2026-10-09).

80. Press Information Bureau, Ministry of Home Affairs. Highlights of New Criminal Laws [press release; commencement of the Bharatiya Sakshya Adhiniyam, 2023 on 1 July 2024]. 2024. PIB release 2039055, 30 July 2024, first paragraph. Press Information Bureau. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2039055 (accessed 2026-10-09).

81. Supreme Court of India. Manoj and Others v. State of Madhya Pradesh. 2022. Para 134, quoting the CFSL Kolkata article (2007), Statistical Analysis. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

82. Supreme Court of India. Manoj and Others v. State of Madhya Pradesh. 2022. Para 136, quoting the Law Commission of India's 185th Report. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

83. Supreme Court of India. Manoj and Others v. State of Madhya Pradesh. 2022. Para 142. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

84. Supreme Court of India. Manoj and Others v. State of Madhya Pradesh. 2022. Para 143. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

85. Supreme Court of India. Pattu Rajan v. State of Tamil Nadu. 2019. Para 32, last paragraph, second sentence. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

86. Supreme Court of India. Pattu Rajan v. State of Tamil Nadu. 2019. Para 32, last paragraph, first sentence. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

87. Supreme Court of India. Manoj and Others v. State of Madhya Pradesh. 2022. figures quoted in the text. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

88. Watkins, Stephen J. Conviction by mathematical error? Doctors and lawyers should get probability theory right. Editorial. PMID 10617504, PMC1117305. BMJ copyright; no open licence stated on the PMC page. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC1117305/ (accessed 2026-10-09).

89. Court of Appeal (Criminal Division). R v Clark [2003] EWCA Crim 1020. 2003. figures quoted in the text. https://caselaw.nationalarchives.gov.uk/ewca/crim/2003/1020 (accessed 2026-10-09).

90. Hill, Ray. Multiple sudden infant deaths – coincidence or beyond coincidence?. 2004. figures quoted in the text. https://doi.org/10.1111/j.1365-3016.2004.00560.x (accessed 2026-10-09).

91. Press Information Bureau, Ministry of Home Affairs. Highlights of New Criminal Laws [press release; commencement of the Bharatiya Sakshya Adhiniyam, 2023 on 1 July 2024]. 2024. figures quoted in the text. Press Information Bureau. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2039055 (accessed 2026-10-09).

92. Government of India. The Bharatiya Sakshya Adhiniyam, 2023. 2023. figures quoted in the text. Gazette of India, via the Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/2024-04/250882_english_01042024_0.pdf (accessed 2026-10-09).

93. Supreme Court of India. Pattu Rajan v. State of Tamil Nadu. 2019. figures quoted in the text. Supreme Court of India. https://registry.opendata.aws/indian-supreme-court-judgments/ (accessed 2026-10-09).

94. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. p. 102, paragraph on prior and posterior probabilities. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

95. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. p. 102, third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

96. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. p. 102, end of the fourth paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

97. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1, Bayes Probabilities, p. 145 of the CHANCE version. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

98. Mohan, V and Deepa, R and Deepa, M and Somannavar, S and Datta, M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. 2005. figures quoted in the text. https://repository.ias.ac.in/80170/ (accessed 2026-10-09).

99. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §3.1, A Counting Technique, p. 76 of the CHANCE version. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

100. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §3.1, Theorem 3.1 and the paragraph Factorials, p. 80. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

101. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §3.1, Factorials, p. 80. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

102. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1 The binomial distribution, the factorial list. https://www.openintro.org/book/os/ (accessed 2026-10-09).

103. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §3.2, Binomial Coefficients, p. 93. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

104. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §3.2, proof of Theorem 3.5, p. 94. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

105. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, the general formula and its reading. https://www.openintro.org/book/os/ (accessed 2026-10-09).

106. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, footnote 21. https://www.openintro.org/book/os/ (accessed 2026-10-09).

107. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.2.1 Bernoulli distribution, the paragraph defining a Bernoulli random variable. https://www.openintro.org/book/os/ (accessed 2026-10-09).

108. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, the box IS IT BINOMIAL? FOUR CONDITIONS TO CHECK. https://www.openintro.org/book/os/ (accessed 2026-10-09).

109. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, the paragraph after Guided Practice 4.29. https://www.openintro.org/book/os/ (accessed 2026-10-09).

110. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, the paragraph before the formula for a single scenario. https://www.openintro.org/book/os/ (accessed 2026-10-09).

111. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2 Variance of Discrete Random Variables, subsection Bernoulli Trials (this version p. 261). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

112. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §4.3 Binomial Distribution, the ABC College committee example (not binomial). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

113. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1, footnote 23 (answer to Guided Practice 4.32, four friends who smoke). https://www.openintro.org/book/os/ (accessed 2026-10-09).

114. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1 Important Distributions, subsection Geometric Distribution (this version p. 184). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

115. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, Geometric Distribution, the paragraph after Figure 5.1 (this version p. 185). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

116. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.2.2 Geometric distribution, the paragraph after the box GEOMETRIC DISTRIBUTION. https://www.openintro.org/book/os/ (accessed 2026-10-09).

117. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.2.2, the last paragraph of the subsection. https://www.openintro.org/book/os/ (accessed 2026-10-09).

118. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §4.4 Geometric Distribution, attribute 5 (two ways to define the random variable). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

119. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, subsection Hypergeometric Distribution (this version p. 193). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

120. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, Hypergeometric Distribution, the derivation of h(N, k, n, x). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

121. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, Hypergeometric Distribution, the paragraph on the binomial limit. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

122. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.3 Sampling from a small population, the last paragraph. https://www.openintro.org/book/os/ (accessed 2026-10-09).

123. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §4.6 Poisson Distribution, opening list, characteristic 1. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

124. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1 Important Distributions, Poisson Distribution, second paragraph (the constant rate). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

125. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, Poisson Distribution, third paragraph (independence in non-overlapping intervals). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

126. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §5.1, Poisson Distribution, the subintervals paragraph (at most one occurrence in a short subinterval). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

127. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.5 Poisson distribution, the box POISSON DISTRIBUTION. https://www.openintro.org/book/os/ (accessed 2026-10-09).

128. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §4.6, opening list, characteristic 2 (approximation to the binomial). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

129. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2 Variance of Discrete Random Variables, Poisson Distribution paragraph. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

130. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.5, the paragraph after the box (guidelines; independence; rates that change). https://www.openintro.org/book/os/ (accessed 2026-10-09).

131. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1.1 Normal distribution model, the paragraph defining N(μ, σ) and the parameters. https://www.openintro.org/book/os/ (accessed 2026-10-09).

132. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §9.1 Central Limit Theorem for Bernoulli Trials, opening paragraph (the normal density function). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

133. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §9.1, second paragraph (what the central limit theorem tells us). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

134. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1.2 Standardizing with Z-scores, the box THE Z-SCORE. https://www.openintro.org/book/os/ (accessed 2026-10-09).

135. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §6.2 Using the Normal Distribution, opening paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

136. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1.5 68-95-99.7 rule, footnote 15 (the three areas to four places). https://www.openintro.org/book/os/ (accessed 2026-10-09).

137. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1.4, footnote 11 (the 97.5th percentile of the standard normal). https://www.openintro.org/book/os/ (accessed 2026-10-09).

138. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.2 Normal approximation to the binomial distribution, the box NORMAL APPROXIMATION OF THE BINOMIAL DISTRIBUTION. https://www.openintro.org/book/os/ (accessed 2026-10-09).

139. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.3, the box IMPROVING THE NORMAL APPROXIMATION FOR THE BINOMIAL DISTRIBUTION. https://www.openintro.org/book/os/ (accessed 2026-10-09).

140. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1, the box NORMAL DISTRIBUTION FACTS. https://www.openintro.org/book/os/ (accessed 2026-10-09).

141. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.1 Normal distribution, opening paragraph (adult male heights). https://www.openintro.org/book/os/ (accessed 2026-10-09).

142. National Institute of Standards and Technology. NIST/SEMATECH e-Handbook of Statistical Methods, section 1.3.3.21: Normal Probability Plot. §1.3.3.21 Normal Probability Plot, Purpose row. Web pages. https://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm (accessed 2026-10-09).

143. National Institute of Standards and Technology. NIST/SEMATECH e-Handbook of Statistical Methods, section 1.3.3.21: Normal Probability Plot. §1.3.3.24 Quantile-Quantile Plot, Purpose row, last sentence. Web pages. https://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm (accessed 2026-10-09).

144. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Basic properties of log-normal distributions, first paragraph. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

145. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Box: Definition and properties of the log-normal distribution, first paragraph. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

146. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Opening section, last paragraph before the title block. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

147. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Section after the dice illustration, the multiplicative central limit theorem. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

148. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Basic properties, the paragraph after Figure 4 (the 68.3% interval). https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

149. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Basic properties, the paragraph after Figure 4 (the 95.5% interval). https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

150. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Basic properties, the estimation paragraph. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

151. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Basic properties, the median paragraph. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

152. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. Box: Definition and properties of the log-normal distribution, the estimation from $\bar{x} \pm s$. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

153. Limpert, Eckhard and Stahel, Werner A and Abbt, Markus. Log-normal distributions across the sciences: keys and clues. 2001. figures quoted in the text. https://stat.ethz.ch/~stahel/lognormal/bioscience.pdf (accessed 2026-10-09).

154. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2, exercise 23 (covariance). https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

155. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2, exercise 23 (covariance), second sentence. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

156. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2, Properties of Variance, proof of Theorem 6.8. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

157. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2, Properties of Variance, the paragraph before Theorem 6.8. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

158. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §3.4.4 Variability in linear combinations of random variables, the paragraph after the formula. https://www.openintro.org/book/os/ (accessed 2026-09-24).

159. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §12.3, The Correlation Coefficient r, What the VALUE of r tells us, first bullet. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

160. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §12.3, What the VALUE of r tells us, fourth bullet. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

161. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §12.3, What the VALUE of r tells us, third bullet. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

162. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2, Properties of Variance, the paragraph before Theorem 6.9. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

163. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §7.1 The Central Limit Theorem for Sample Means, second paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

164. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §7.1, second paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

165. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §7.1, second paragraph, last sentence. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

166. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §7.1, third paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

167. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §7.2 The Central Limit Theorem for Sums, second paragraph (bold). https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-10-09).

168. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §9.1 Central Limit Theorem for Bernoulli Trials, second paragraph. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

169. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §8.1 Law of Large Numbers for Discrete Random Variables, opening paragraph. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

170. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §7.1.2, the paragraph after Example 7.1 (extremely skewed data). https://www.openintro.org/book/os/ (accessed 2026-10-09).

171. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.1 Definitions of Statistics, Probability, and Key Terms, the paragraph defining statistic and parameter. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

172. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.5 Measures of the Center of the Data, the paragraph before 'Calculating the Mean of Grouped Frequency Tables'. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

173. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.7 Measures of the Spread of the Data, the paragraph on s and σ. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

174. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §9.1 Central Limit Theorem for Bernoulli Trials, Example 9.4 (the poll), the paragraph naming the 95 percent confidence interval, this version p. 334. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

175. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §9.1, Example 9.4, the paragraph defining p. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

176. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.7 Measures of the Spread of the Data, the paragraph on dividing by n − 1. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

177. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.7, subsection Sampling Variability of a Statistic. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

178. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.7, subsection Sampling Variability of a Statistic, the sentence after. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

179. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §6.2 Variance of Discrete Random Variables, Example 6.18, this version p. 261. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

180. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.1 Definitions of Statistics, Probability, and Key Terms, the paragraph after the definitions. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

181. Diez, David M and -Rundel, Mine and Barr, Christopher D. OpenIntro Statistics. 2019. §4.3.1 The binomial distribution, the box BINOMIAL DISTRIBUTION. https://www.openintro.org/book/os/ (accessed 2026-10-09).

182. Grinstead, Charles M. and Snell, J. Laurie. Grinstead and Snell's Introduction to Probability. 2006. §4.1 Discrete Conditional Probability, Historical Remarks, on Bernoulli and Bayes, this version p. 150. https://math.dartmouth.edu/~prob/prob/prob.pdf (accessed 2026-10-09).

183. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. Second paragraph, the formal definition of a likelihood ratio. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

184. Royal Statistical Society. Royal Statistical Society concerned by issues raised in Sally Clark case [news release; statement regarding statistical issues in the Sally Clark case]. 2001. News release of 23 October 2001, the paragraph on the Prosecutor's Fallacy. http://web.archive.org/web/20011211220114/http://www.rss.org.uk/archive/reports/sclark.html (accessed 2026-10-09).

185. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.1 Definitions of Statistics, Probability, and Key Terms, the paragraph on statistics and parameters. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

186. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.2 Data, Sampling, and Variation in Data and Sampling, subsection Sampling, first sentence. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

187. International Institute for Population Sciences. National Family Health Survey (NFHS-6), 2023-24: India and State/UT Fact Sheets. 2026. India Key Indicators table, indicator 76, and footnote 15. https://www.nfhsiips.in (accessed 2026-10-10).

188. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Summary, Methods, first two sentences. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

189. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Summary, Findings, third sentence. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

190. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Methods, Estimation of overweight and obesity prevalence from 1990 to 2021, last two sentences. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

191. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Results, Forecasts of adult overweight and obesity in 2050, first two sentences. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

192. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Results, Forecasts of adult overweight and obesity in 2050, opening words of the first sentence. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

193. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Methods, Forecasts of overweight and obesity prevalence from 2022 to 2050, first sentence. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

194. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Methods, Forecasts of overweight and obesity prevalence from 2022 to 2050, the sentence after the assumptions of the ARC submodels. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

195. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Discussion, limitations paragraph, fourth limitation. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

196. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Discussion, limitations paragraph, third limitation. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

197. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. Methods, Overview, the definition of overweight and obese. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

198. Mohan, V and Deepa, R and Deepa, M and Somannavar, S and Datta, M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. 2005. Results, the paragraph on Table 2. https://repository.ias.ac.in/80170/ (accessed 2026-10-09).

199. GBD 2021 Adult BMI Collaborators. Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. 2025. figures quoted in the text. https://doi.org/10.1016/S0140-6736(25)00355-1 (accessed 2026-10-09).

200. Kiran, R and Harshitha and Bhargava, Madhavi. Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India. Open access, PMC9791811. Used in Book 0 Part D as real material to compute with, not as evidence about screening. Abstract, Results, second sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/ (accessed 2026-09-23).

201. Kiran, R and Harshitha and Bhargava, Madhavi. Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India. Open access, PMC9791811. Used in Book 0 Part D as real material to compute with, not as evidence about screening. Table 2, Men column, the overweight and obese rows. https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/ (accessed 2026-09-23).

202. Kiran, R and Harshitha and Bhargava, Madhavi. Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India. Open access, PMC9791811. Used in Book 0 Part D as real material to compute with, not as evidence about screening. Material and methods, Statistical analysis, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/ (accessed 2026-09-23).

203. Kiran, R and Harshitha and Bhargava, Madhavi. Mid-upper arm circumference and neck circumference to screen for overweight-obesity in young adults in South India. Open access, PMC9791811. Used in Book 0 Part D as real material to compute with, not as evidence about screening. Discussion, Strengths and limitations. https://pmc.ncbi.nlm.nih.gov/articles/PMC9791811/ (accessed 2026-09-23).

204. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. BMJ 1994;309:102, the paragraph on prior and posterior probabilities. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

205. Altman, Douglas G. and Bland, J. Martin. Diagnostic tests 2: predictive values. Statistics Notes. PMID 8038641, PMC2540558 (scanned page). BMJ copyright; no open licence stated. BMJ 1994;309:102, end of the third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC2540558/ (accessed 2026-10-09).

206. Deeks, Jonathan J and Altman, Douglas G. Diagnostic tests 4: likelihood ratios. 2004. BMJ 2004;329:168, the paragraph on adapting a study's results to your patients. https://pmc.ncbi.nlm.nih.gov/articles/PMC478236/ (accessed 2026-10-09).

207. International Institute for Population Sciences. National Family Health Survey (NFHS-6), 2023-24: India and State/UT Fact Sheets. 2026. India Key Indicators table, indicator 77. https://www.nfhsiips.in (accessed 2026-10-10).

Appendix · Worked answers

1 · Sample space, events and the three rules every probability obeys

EXERCISE 1. No probability is negative. The whole sample space has probability 1. If two events cannot both happen, the probability of one or the other is the sum of their probabilities.

The general addition rule: P(A or B) = P(A) + P(B) − P(A and B).

The outcomes in both A and B are counted once in P(A) and again in P(B). Subtracting P(A and B) takes the second count away, so each outcome is counted once.

EXERCISE 2. The three shares add to 115 per cent.

```
30 plus 45 plus 40 = 115
```

A list of categories that cannot overlap must add to 100 per cent. This one adds to more, which tells you the categories overlap: people ticked more than one. The table even forces a count. The 400 people made 460 ticks, so there are at least 60 ticks beyond one each. One person can carry at most two extra ticks.

```
1.15 times 400 = 460
460 minus 400 = 60
60 divided by 2 = 30
```

So at least 30 people ticked more than one. That much the table shows.

It does not show that everyone has at least one. The share with at least one is "diabetes or hypertension or overweight". Without the overlaps it is known only as a range. It is at least 45 per cent, the largest single share, if everyone with diabetes or overweight also has hypertension. It is at most 100 per cent. The 115 is not a count of anything: people with two conditions are in it twice and people with three are in it three times.

To get the share with at least one, the survey must report how many ticked none. Then the complement rule gives it: 1 minus the share with none.

1. A is 2, 4 and 6. B is 4, 5 and 6.

"A and B" is the outcomes in both: 4 and 6.

"A or B" is the outcomes in at least one: 2, 4, 5 and 6. The 4 and the 6 are listed once.

"Not A" is the outcomes outside A: 1, 3 and 5.

2. Use the general addition rule.

```
0.3 plus 0.5 minus 0.1 = 0.7
```

P(A or B) is 0.7. "Neither A nor B" is the complement of "A or B".

```
1 minus 0.7 = 0.3
```

The probability that neither happens is 0.3.

3. Turn the general addition rule round. P(A or B) = P(A) + P(B) − P(A and B), so P(A and B) = P(A) + P(B) − P(A or B).

```
0.45 plus 0.30 = 0.75
0.75 minus 0.60 = 0.15
```

P(A and B) is 0.15. It is above zero, so some outcomes are in both. A and B are not mutually exclusive.

4. The four bands cannot overlap: each man has one BMI. Check that they cover all 131 men.

```
5 plus 74 plus 42 plus 10 = 131
```

“25 or more” is “25 to 29.9 or 30 and over”, two mutually exclusive bands. Rule 3 adds them.

```
42 divided by 131 ≈ 0.3206
10 divided by 131 ≈ 0.0763
0.3206 plus 0.0763 = 0.3969
52 divided by 131 ≈ 0.3969
```

The probability is 0.3969: 52 of the 131 men.

“18.5 or more” is the complement of “below 18.5”. The complement rule is quickest.

```
5 divided by 131 ≈ 0.0382
1 minus 0.0382 = 0.9618
```

The probability is 0.9618: 126 of the 131 men.

5. The general addition rule needs the overlap, which is missing. So work out the two extremes.

The most the overlap can be is 41, if every man with the large waist also has the high BMI. Then “either” is only the 52.

```
52 minus 41 = 11
52 divided by 131 ≈ 0.3969
```

The least the overlap can be is 0, if no man is in both. Then “either” is every man in either group.

```
52 plus 41 = 93
93 divided by 131 ≈ 0.7099
```

The probability lies between 0.3969 and 0.7099: between 52 and 93 of the 131 men. In the first case, 11 of the 52 men with the high BMI have a smaller waist. To narrow the range you need the count of men in both groups.

6. The second line broke. It adds as if “flagged” and “in the band” could not both happen. They can: 36 men are both. Subtract the overlap once.

```
36 divided by 131 ≈ 0.2748
0.4504 plus 0.3206 minus 0.2748 = 0.4962
```

About 50 per cent, 65 men, not 77 per cent.

Why the wrong answer looks reasonable: "or means add" is the rule Book 0 taught for mutually exclusive events, and the arithmetic in each line is right. The working never asked whether a man could be in both groups.

7. The second line broke. A and B cannot both happen, so "A and B" holds no outcome and P(A and B) is 0, not 0.06. Multiplying the two probabilities gives P(A and B) only for independent events, and these are not independent: learning that B happened tells you A did not.

```
0.2 plus 0.3 minus 0 = 0.5
```

P(A or B) is 0.5.

Why the wrong answer looks reasonable: the general addition rule is used correctly, and "multiply for and" is a rule the reader knows from Book 0. It was used where its condition fails.

8. "One or the other" is P(overweight or obese, or anaemic). The general addition rule needs the share of women with both, and the fact sheet does not give it. So compute the range.

```
24.0 plus 57.0 = 81.0
```

The 81.0 per cent is the most it could be. It is reached only if no woman with overweight or obesity is also anaemic. The least it could be is 57.0 per cent, if every woman with overweight or obesity is also anaemic. So the share is somewhere from 57.0 to 81.0 per cent, and the presentation quoted the top of the range as the answer.

There is a second problem. The two figures are not shares of the same women. The fact sheet's footnote 21, on indicator 88, says it "Excludes pregnant women and women with a birth in the preceding 2 months." Indicator 95 is for all women aged 15-49. Before combining them at all, say that the two groups differ.

What the numbers do not establish: how many women have both, and so how many have at least one. Nor do they say that the two conditions are linked or unlinked. A survey table of the two together would answer both questions.

2 · Conditional probability as a definition, and the multiplication rule

EXERCISE 1. P(A | B) = P(A and B) ÷ P(B), for P(B) above 0.

P(A and B) = P(B) × P(A | B), and equally P(A) × P(B | A).

The two conditionals share the top, P(A and B), and divide it by different bottoms, P(B) and P(A). They are equal only when P(A) equals P(B), or when A and B never happen together, so that both are 0.

EXERCISE 2.

```
120 divided by 400 = 0.3
160 divided by 400 = 0.4
90 divided by 400 = 0.225
90 divided by 120 = 0.75
```

P(raised sugar) is 0.3 and P(overweight) is 0.4: marginal. P(raised sugar and overweight) is 0.225: joint. P(overweight | raised sugar) is 0.75: conditional, on the 120 with a raised sugar.

The question asks for the other conditional: the condition is now "overweight".

```
0.225 divided by 0.4 = 0.5625
90 divided by 160 = 0.5625
```

About 56 per cent. 90 divided by 120 answers “of those with a raised sugar, how many were overweight”. The question asked “of those overweight, how many had a raised sugar”, so the bottom is 160.

What the report does not show: anything about adults who were not screened, or why the two go together.

1. The people who are B: 20 plus 10. The people who are A: 20 plus 30.

```
20 plus 10 = 30
20 plus 30 = 50
20 divided by 30 ≈ 0.6667
20 divided by 50 = 0.4
```

P(A | B) is 0.6667 and P(B | A) is 0.4. The top, 20, is the same.

2.

```
0.12 divided by 0.4 = 0.3
```

P(A | B) is 0.3.

3. Use the multiplication rule, P(A and B) = P(B) × P(A | B).

```
0.25 times 0.6 = 0.15
```

P(A and B) is 0.15.

4. The multiplication rule says 0.18 = P(B) × 0.45. Divide both sides by 0.45.

```
0.18 divided by 0.45 = 0.4
```

P(B) is 0.4. Check it: 0.4 times 0.45 is 0.18.

5. On the first split, P(B) is 0.3 and P(not B) is 1 minus 0.3. On each second split, the two conditionals add to 1.

```
1 minus 0.3 = 0.7
1 minus 0.5 = 0.5
1 minus 0.2 = 0.8
0.3 times 0.5 = 0.15
0.3 times 0.5 = 0.15
0.7 times 0.2 = 0.14
0.7 times 0.8 = 0.56
0.15 plus 0.15 plus 0.14 plus 0.56 = 1
```

P(B and A) is 0.15, P(B and not A) is 0.15, P(not B and A) is 0.14 and P(not B and not A) is 0.56. They add to 1.

6.

```
36 divided by 131 ≈ 0.2748
42 divided by 131 ≈ 0.3206
0.2748 divided by 0.3206 = 0.8571
```

P(flagged | in the band) is about 0.857, or 86 per cent of the men in the band. It is the sensitivity, as the paper prints it.

7. The women not in the band:

```
151 minus 41 = 110
```

The women in the band whom the tape flags: 88 per cent of 41, to the nearest woman, checked by dividing back.

```
0.88 times 41 = 36.08
36 divided by 41 ≈ 0.878
```

36 rounds back to 88 per cent. The women outside the band whom the tape clears: 83 per cent of 110.

```
0.83 times 110 = 91.3
91 divided by 110 ≈ 0.827
```

91 rounds back to 83 per cent. The rest follow by subtraction.

```
41 minus 36 = 5
110 minus 91 = 19
36 plus 19 = 55
```

	in the band	not in the band	total
arm 28.5 cm or more	36	19	55
arm under 28.5 cm	5	91	96
total	41	110	151

The chance that a flagged woman is in the band divides by the flagged row.

```
36 divided by 55 ≈ 0.6545
```

About 65 per cent of the women the tape flags are in the band. The table is rebuilt from rounded percentages, as the men's was.

8. Outside the band: 1 minus 0.321. Wrongly flagged, given outside the band: 1 minus the specificity.

```
1 minus 0.321 = 0.679
1 minus 0.74 = 0.26
0.679 times 0.26 = 0.1765
0.1765 times 131 = 23.1
```

The probability is about 0.177, which is about 23 men of the 131, as the rebuilt table has. The 0.1 of a man comes from the rounded percentages.

9. The third line broke. The condition is "flagged", so the bottom is the 59 flagged men, not the 42 in the band.

```
36 divided by 59 ≈ 0.6102
```

About 61 per cent. The working computed P(flagged | in the band), the sensitivity.

Why the wrong answer looks reasonable: 86 per cent is the number the paper prints, and the top line, 36, is right for both questions. Only the bottom changes, and the working took the bottom from the wrong total.

10. The third line broke. It multiplies the chance of being outside the band by the sensitivity, which is P(flagged | in the band). The second split needs the conditional for men outside the band: P(flagged | not in the band), which is 1 minus the specificity.

```
1 minus 0.74 = 0.26
0.6794 times 0.26 = 0.1766
0.1766 times 131 = 23.1
```

About 23 men, as in the rebuilt table, not 77. A check would have caught it: 77 men flagged outside the band is more than the 59 the tape flags in all.

Why the wrong answer looks reasonable: "flagged" appears in both, so the sensitivity looks like the chance of being flagged. On a tree diagram, each second split needs the conditional for its own first-stage group.

11. The second line broke. Dividing by the 500 screened gives P(overweight and raised sugar), a joint probability. The condition is "raised sugar", so the bottom is the 100 with a raised sugar.

```
60 divided by 100 = 0.6
```

60 per cent of adults with a raised sugar have overweight. The figures are made up.

Why the wrong answer looks reasonable: dividing by the number screened is the habit of every prevalence. It gives a true number, 12 per cent of all adults are both. It answers a different question.

12. The paper's 86 per cent is the sensitivity, P(flagged | in the 25-29.9 band). The news item states P(in the band | flagged). Compute that from the rebuilt table.

```
36 divided by 59 ≈ 0.6102
```

About 61 per cent of the men the tape flagged were in the band, not 86 per cent. "Accurate" names neither conditional, so the sentence should name the one it means.

What the 61 per cent does not establish. It comes from a table rebuilt from rounded percentages, for 131 men at one college. "In the band" means 25 to 29.9. It leaves out the 10 men at 30 or more. So for the question "is his index 25 or more?", 61 per cent is the least the share could be. And it holds only where about a third of the men are in the band; where fewer are, it is lower.

13. The first sentence gives P(overweight | diabetes), 0.667. The second claims P(diabetes | overweight). Build the joint probability with the multiplication rule, then divide by the other condition.

```
2 divided by 3 ≈ 0.6667
0.1 times 0.6667 = 0.06667
0.06667 divided by 0.3 = 0.2222
```

In this clinic a patient with overweight has about a 22 per cent chance of diabetes, not 67 per cent. The two conditionals share the top, 0.0667, and divide it by 0.1 and by 0.3.

What this does not establish. The 10 and 30 per cent are made up; with the clinic's own figures the answer changes, and the speaker gave none. Neither conditional says whether being overweight raises the chance of diabetes. That needs a comparison with patients who are not overweight, and even then an association is not a cause.

3 · Independence given a condition, and where it fails

EXERCISE 1. A and B are independent when P(A | B) = P(A): knowing B leaves the chance of A unchanged. Then the multiplication rule gives the check P(A and B) = P(A) × P(B).

If A and B cannot both happen, P(A and B) is 0. But P(A) × P(B) is above 0 when each is possible, so the check fails: knowing B happened makes A impossible.

EXERCISE 2. The flaw is the independence assumption. Two people from one household share meals, income, habits and often genes. Knowing one has overweight raises the chance the other does, so P(second has overweight | first has overweight) is likely above 0.25, and the product understates the chance that both are.

```
0.25 times 0.25 = 0.0625
```

The 6 per cent is right only if household members are independent, which is exactly what is in doubt.

The protocol should not assume it. It should count, from its own data, how many households have both members with overweight, and compare that with 0.0625 of the households. It should also treat the household, not the person, as the unit when it works out how sure it can be. Two people from one house carry less information than two strangers.

1.

```
0.4 times 0.5 = 0.2
0.2 divided by 0.5 = 0.4
```

P(A) × P(B) is 0.2, which equals P(A and B), so A and B are independent. P(A | B) is 0.4, the same as P(A).

2. Mutually exclusive means no outcome is in both, so P(A and B) is 0.

```
0.3 times 0.2 = 0.06
0 divided by 0.2 = 0
```

P(A) × P(B) is 0.06, not 0. P(A | B) is 0, not 0.3. A and B are dependent.

3. Without replacement: after one marked slip is drawn, 1 marked slip is left among 4.

```
2 divided by 5 = 0.4
1 divided by 4 = 0.25
0.4 times 0.25 = 0.1
```

With replacement, each draw is independent.

```
0.4 times 0.4 = 0.16
```

0.1 without replacement and 0.16 with it.

4. The conditional form: compare P(flagged | in the band) with P(flagged).

```
36 divided by 42 ≈ 0.8571
59 divided by 131 ≈ 0.4504
```

The product form: compare P(flagged and in the band) with P(flagged) × P(in the band).

```
36 divided by 131 ≈ 0.2748
42 divided by 131 ≈ 0.3206
0.4504 times 0.3206 = 0.1444
```

0.857 against 0.450, and 0.275 against 0.144. Both forms say the two are dependent. A man in the band is about twice as likely to be flagged as a man picked at random from the 131.

5. Without replacement, the second pick depends on the first.

```
3 divided by 12 = 0.25
2 divided by 11 ≈ 0.1818
0.25 times 0.1818 = 0.0455
```

As if independent:

```
0.25 times 0.25 = 0.0625
```

0.045 against 0.063. Once one file of a patient with diabetes is picked, only 2 of the 11 files left belong to such patients. So the chance for the second pick falls from 0.25 to 0.18. Two files are 17 per cent of the ward, well above OpenIntro's 10 per cent rule of thumb, so the difference shows.

6. Count within each group, then add.

```
1000 minus 200 = 800
0.9 times 200 = 180
0.9 times 180 = 162
0.2 times 800 = 160
0.2 times 160 = 32
180 plus 160 = 340
162 plus 32 = 194
194 divided by 340 ≈ 0.5706
340 divided by 1000 = 0.34
```

The first test flags 340 people and both tests flag 194. Given a positive first test, the second is positive with probability about 0.57. Overall it is positive with probability 0.34, by the same count as the first. The two tests are dependent across everyone, though independent inside each group.

7. The third line broke. It treats the two measurements as independent. They are of the same man, with the same arm, by the same observer. A man missed the first time was missed mostly because his arm is below 31.3 cm, and it will still be below 31.3 cm the next day. So P(missed second time | missed first time) is close to 1, not 0.14, and the chance of being missed both times is close to 0.14.

```
0.14 times 1 = 0.14
```

Measuring twice catches only the misses that came from measurement scatter, not the ones that came from the man's own body.

Why the wrong answer looks reasonable: the 14 per cent is a share of different men. Read as a chance that each measurement fails at random, it invites multiplying. It describes which men the tape misses, and the same men are missed again.

8. The second line broke. Independent events do overlap: their overlap is P(A) × P(B), not 0. It is mutually exclusive events whose overlap is 0.

```
0.3 times 0.4 = 0.12
0.3 plus 0.4 minus 0.12 = 0.58
```

P(A or B) is 0.58, not 0.7.

Why the wrong answer looks reasonable: "independent" sounds like "separate", and separate sounds like "no overlap". The two words name different ideas. Independent is about whether one tells you about the other; mutually exclusive is about whether both can happen.

9. The note multiplies the two shares, which is the check for independent events used as if independence were known.

```
0.307 times 0.273 = 0.083811
```

About 8.4 per cent. That is the share of couples in which both partners have overweight or obesity only if a wife's status tells you nothing about her husband's.

What the number does not establish. Partners share a kitchen, an income and a neighbourhood, so their weights are likely to go together, and the true share could be well above 8.4 per cent. The two figures are shares of all women and all men aged 15-49, not of married couples. The women's figure also leaves out pregnant women and women who gave birth in the last two months. To know the share of couples, you need data on couples: the survey's couple records, not its two separate rows.

4 · The law of total probability: a probability as a weighted average over groups

EXERCISE 1. The probability of an event is the sum, over the groups, of the probability of the event in that group times the probability of being in that group. It needs groups that do not overlap and that take in everyone.

For a test:

```
P(positive) = sensitivity × prevalence + (1 - specificity) × (1 - prevalence)
```

The two weights, the prevalence and 1 minus the prevalence, are never negative and add to 1. So the positive share is a weighted average of the sensitivity and the false-positive rate, 1 minus the specificity. It lies between them, nearer the rate of the larger group.

EXERCISE 2. Urban and rural women do not overlap, and together they take in every woman. So the district figure is a weighted average of 31 and 18 per cent, and it must lie between them. 33 per cent is above both, which a weighted average can never be.

So one of three things is true. One figure is a misprint. Or the three were measured on different groups of women, for example different ages or years. Or the "whole district" row counts something else, such as a different body-mass index cut-off.

Ask which table each figure came from, what population and cut-off each covers, and how many women were measured in each group. Quote none of them until the three agree.

1. Group 2 holds 1 minus 0.3 of the population.

```
1 minus 0.3 = 0.7
0.9 times 0.3 = 0.27
0.2 times 0.7 = 0.14
0.27 plus 0.14 = 0.41
```

P(B) is 0.41.

2. Check the shares first.

```
0.5 plus 0.3 plus 0.2 = 1.0
```

They add to 1. Now weight each probability by its group's share.

```
0.1 times 0.5 = 0.05
0.4 times 0.3 = 0.12
0.7 times 0.2 = 0.14
0.05 plus 0.12 plus 0.14 = 0.31
```

P(B) is 0.31. The plain average is higher.

```
0.1 plus 0.4 plus 0.7 = 1.2
1.2 divided by 3 = 0.4
```

The plain average, 0.4, gives the smallest group as much say as the largest. The largest group has the lowest rate, so the weighted answer is lower.

3. Call the share in group 1 w. Group 2 then holds 1 − w.

```
0.24 = 0.8 × w + 0.1 × (1 - w)
0.24 = 0.8 × w + 0.1 - 0.1 × w
0.24 = 0.1 + (0.8 - 0.1) × w
```

Take 0.1 from both sides, then divide by the gap.

```
0.24 minus 0.1 = 0.14
0.8 minus 0.1 = 0.7
0.14 divided by 0.7 = 0.2
```

Group 1 holds 0.2 of the population. Check it.

```
0.8 times 0.2 plus 0.1 times 0.8 = 0.24
```

4. The groups are the men in the band and the men outside it.

```
131 minus 42 = 89
42 divided by 131 = 0.3206…
89 divided by 131 = 0.6794…
1 minus 0.74 = 0.26
0.86 times 0.3206 = 0.2757
0.26 times 0.6794 = 0.1766
0.2757 plus 0.1766 = 0.4523
0.4523 times 131 = 59.25
```

About 59 men, which is 45 per cent of the 131. You can also work in counts and skip the shares.

```
0.86 times 42 = 36.12
0.26 times 89 = 23.14
36.12 plus 23.14 = 59.26
```

The rebuilt table in Book 0 has 59 men in its flagged row, 36 and 23, the same counts rounded to whole men. The 59 is what the printed percentages imply, not a row the paper prints.

5. Call the urban share w.

```
27.3 = 36.3 × w + 23.0 × (1 - w)
27.3 minus 23.0 = 4.3
36.3 minus 23.0 = 13.3
4.3 divided by 13.3 = 0.323
```

About 0.32 of the men. Put it back to check.

```
36.3 times 0.323 plus 23.0 times 0.677 = 27.2959
```

That rounds to 27.3.

Move each printed figure by up to 0.05. The weight is largest when the India figure is highest, and the urban and rural figures are lowest.

```
27.35 minus 22.95 = 4.4
36.25 minus 22.95 = 13.3
4.4 divided by 13.3 = 0.331
27.25 minus 23.05 = 4.2
36.35 minus 23.05 = 13.3
4.2 divided by 13.3 = 0.316
```

So the share lies somewhere from about 0.32 to 0.33. Say “about 0.32”, not 0.3233.

6. The fourth line broke. It takes the plain average of the two rates, as if the two groups were the same size. They are 200 men and 800 men. Weight each rate by its group’s share.

```
200 divided by 1000 = 0.2
800 divided by 1000 = 0.8
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
```

About 380 of the 1,000 men will be flagged, not 560.

Why the wrong answer looks reasonable: the share flagged really is a mix of the two rates, and a plain average is the first mix anyone reaches for. It is right at one prevalence only, a half. The working was given the prevalence in the second line and never used it.

7. The third line broke. For the 0.8 without the condition, it uses the specificity, 0.74. The specificity is the share of them the test clears, so 0.74 times 0.8 counts true negatives, not positives. The share of them the test flags is 1 minus the specificity.

```
1 minus 0.74 = 0.26
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
```

P(positive) is 0.38.

Why the wrong answer looks reasonable: sensitivity and specificity are the two numbers a test is sold on, so they feel like the two numbers to weight. The check is to say each term’s event aloud: “positive and has it”, then “positive and has it not”. The 0.74 is “negative and has it not”, the wrong event.

8. What to compute: 38 per cent is the share who tested positive, not the share in the band. The law of total probability links the two, so run it backwards for the prevalence. Call the prevalence p.

```
190 divided by 500 = 0.38
0.38 = 0.86 × p + 0.26 × (1 - p)
0.38 = 0.26 + (0.86 - 0.26) × p
0.38 minus 0.26 = 0.12
0.86 minus 0.26 = 0.60
0.12 divided by 0.60 = 0.2
```

If the tape's rates hold here, about 20 per cent of these men are in the band, not 38.

What this does not establish. The sensitivity and specificity come from 131 medical students in one college. Carrying them to this camp is an assumption, and the answer moves with them. The 0.60 gap at the bottom also magnifies any error in the top line. The band is 25 to 29.9, so the answer says nothing about men at 30 or more. And it is a prevalence among the men who came to the camp, not among all young men in the area. To know the prevalence, weigh and measure a sample of the men.

9. What to compute: the district share in each survey, as a weighted average of the town and village shares.

```
1 minus 0.3 = 0.7
30 times 0.3 plus 15 times 0.7 = 19.5
1 minus 0.4 = 0.6
30 times 0.4 plus 15 times 0.6 = 21.0
```

The district share rises from 19.5 to 21.0 per cent. Neither the towns nor the villages changed. Only the weights did: more of the women now live in towns, where the share is higher.

So the press line is false as logic. A rise in the total can come from the rates, from the weights, or from both.

What this does not establish. The district is made up, so it shows only that the press line's "must" fails. It does not say what happened in any real district. To find out, ask for the town and village shares in both surveys, and for the share of women in each. Then you can split the rise into the part from the rates and the part from the weights.

5 · Bayes' theorem, from the multiplication rule, on a diagnostic test

EXERCISE 1. The general multiplication rule gives P(A and B) two ways.

```
P(A and B) = P(B) × P(A | B) = P(A) × P(B | A)
```

Divide both sides by P(B).

```
P(A | B) = P(B | A) × P(A) ÷ P(B)
```

For a test, with P(B) from the law of total probability:

```
PPV = sensitivity × prevalence ÷ [sensitivity × prevalence + (1 - specificity) × (1
- prevalence)]
```

The PPV and NPV change with the prevalence. The sensitivity and specificity do not, as long as the test behaves the same way in the new group, because each is worked out inside one group.

EXERCISE 2. A good plan has her do the work, in this order.

First, name her number. Write P(positive | in the band) = 0.86 and read the bar as "given". Ask her which way round her question runs. It is P(in the band | positive). Then say what "overweight" means here: the tape was checked against the 25-to-29.9 band.

Second, draw 1,000 men split by the band: 200 in, 800 out. Have her flag 86 per cent of the 200 and 26 per cent of the 800, which is 172 and 208.

Third, ask "out of whom?" Of the 380 flagged, 172 are in the band. Have her divide: about 0.45.

Fourth, write the formula beside the counts, and show that each count is one term of it. Then change the 200 to 50 and let her rework it: the PPV falls to about 0.15 while the 86 per cent stays.

What she should leave with: the 86 per cent describes the tape among men in the band. The number for the man in front of her needs the prevalence.

1. Use Bayes' theorem: P(A | B) = P(B | A) × P(A) ÷ P(B).

```
0.6 times 0.3 = 0.18
0.18 divided by 0.25 = 0.72
```

P(A | B) is 0.72.

2. P(B) is not given, so build it first with the law of total probability.

```
1 minus 0.4 = 0.6
0.5 times 0.4 = 0.2
0.2 times 0.6 = 0.12
0.2 plus 0.12 = 0.32
```

P(B) is 0.32. Now Bayes' theorem.

```
0.2 divided by 0.32 = 0.625
```

P(A | B) is 0.625.

3. PPV first.

```
0.9 times 0.1 = 0.09
1 minus 0.8 = 0.2
1 minus 0.1 = 0.9
0.2 times 0.9 = 0.18
0.09 plus 0.18 = 0.27
0.09 divided by 0.27 = 0.3333
```

The PPV is about 0.33. Now the NPV.

```
0.8 times 0.9 = 0.72
1 minus 0.9 = 0.1
0.1 times 0.1 = 0.01
0.72 plus 0.01 = 0.73
0.72 divided by 0.73 = 0.9863
```

The NPV is about 0.99. A positive result is right one time in three; a negative result is right 99 times in 100.

4. The prevalence is 42 of 131.

```
42 divided by 131 = 0.3206…
1 minus 0.3206 = 0.6794
0.86 times 0.3206 = 0.2757
0.26 times 0.6794 = 0.1766
0.2757 plus 0.1766 = 0.4523
0.2757 divided by 0.4523 = 0.6096
```

The PPV is about 0.61. Now the NPV.

```
0.74 times 0.6794 = 0.5028
0.14 times 0.3206 = 0.0449
0.5028 plus 0.0449 = 0.5477
0.5028 divided by 0.5477 = 0.918
```

The NPV is about 0.92. The table gives these.

```
36 divided by 59 = 0.6102…
66 divided by 72 = 0.9167…
```

The two ways agree to two decimal places. The table rounded each cell to whole men; the formula did not.

5. Call the prevalence p. The false-positive rate is 1 minus 0.9, which is 0.1. A PPV of 0.5 means the positives who have the condition are half of all positives, so the top line equals the other term of the bottom line.

```
0.8 × p = 0.1 × (1 - p)
0.8 × p = 0.1 - 0.1 × p
```

Add 0.1 × p to both sides, as in Book 0, C2.

```
0.8 plus 0.1 = 0.9
0.9 × p = 0.1
0.1 divided by 0.9 = 0.1111
```

The prevalence is about 0.11. Check it.

```
0.8 times 0.1111 = 0.0889
0.1 times 0.8889 = 0.0889
0.0889 divided by 0.1778 = 0.5
```

6. At 0.75:

```
0.895 times 0.75 = 0.67125
1 minus 0.628 = 0.372
0.372 times 0.25 = 0.093
0.67125 plus 0.093 = 0.76425
0.67125 divided by 0.76425 = 0.8783
```

The PPV is about 0.88. At 0.25:

```
0.895 times 0.25 = 0.22375
0.372 times 0.75 = 0.279
0.22375 plus 0.279 = 0.50275
0.22375 divided by 0.50275 = 0.4451
0.628 times 0.75 = 0.471
1 minus 0.895 = 0.105
0.105 times 0.25 = 0.02625
0.471 plus 0.02625 = 0.49725
0.471 divided by 0.49725 = 0.9472
```

A PPV of about 0.45 and an NPV of about 0.95, as the note prints. The same scan's positive results go from right nearly nine times in ten to right fewer than half the time.

7.

```
0.86 times 0.05 = 0.043
1 minus 0.05 = 0.95
0.26 times 0.95 = 0.247
0.043 plus 0.247 = 0.29
0.043 divided by 0.29 = 0.1483
0.74 times 0.95 = 0.703
0.14 times 0.05 = 0.007
0.703 plus 0.007 = 0.71
0.703 divided by 0.71 = 0.9901
```

The PPV is about 0.15 and the NPV about 0.99. There, about six flagged men in seven are not in the band, so a positive result means "measure his weight and height", not "he is in the band".

8. Scoring below 80 is a negative result. The probability of diabetes after it is the diabetic negatives divided by all negatives.

```
1 minus 0.151 = 0.849
0.849 times 0.101 = 0.085749
1 minus 0.101 = 0.899
0.95 times 0.899 = 0.85405
0.085749 plus 0.85405 = 0.939799
0.085749 divided by 0.939799 = 0.0912
```

About 9.1 per cent, against a prior of 10.1 per cent. At this cut-off a negative score barely moves the probability, because the score misses about 85 in every 100 people with undiagnosed diabetes. A score below 80 does not clear him.

9. The second line broke. It swapped the sensitivity and the specificity: it used 0.74 where the sensitivity belongs, and 1 minus 0.86 where 1 minus the specificity belongs. The top line must be P(positive | in the band) times the prevalence.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
0.172 divided by 0.38 = 0.4526
```

The PPV is about 0.45, not 0.57.

Why the wrong answer looks reasonable: the formula has the right shape, every number in it came from the problem, and 0.57 is a believable PPV. Only saying each number as a conditional catches it: 0.74 is "negative given not in the band", which has no place in the top line.

10. The third line broke. The tape is the same, so its sensitivity and specificity may carry over. Its PPV does not, because the PPV depends on the prevalence, and the camp's is 0.2, not the students' 0.32. Recompute at 0.2.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 divided by 0.38 = 0.4526
```

About 45 in every 100 flagged men, not 61.

Why the wrong answer looks reasonable: "the same test gives the same results" is true of the sensitivity and specificity, which describe the tape. The PPV sounds like a property of the tape too, and papers print it beside the other two. It describes the tape and the group together.

11. What to compute: the leaflet's second sentence is a claim about P(in the band | positive), the PPV. Work it out at the camp's prevalence.

```
0.86 times 0.2 = 0.172
0.26 times 0.8 = 0.208
0.172 plus 0.208 = 0.38
0.172 divided by 0.38 = 0.4526
```

About 45 in 100 positive men are in the band. The leaflet is wrong for more than half of the men it speaks to.

A sentence the leaflet could use: "If your result is positive, about 45 in 100 men like you turn out to have a body-mass index from 25 to 29.9. We will measure your weight and height to find out."

What this does not establish. The 0.86 and 0.74 come from medical students in one college and fit the 25-to-29.9 band, so the 0.45 is no more certain than they are. The one in five is an assumption about this camp; at another prevalence the answer moves. And the band leaves out men at 30 or more, so "overweight" in the leaflet does not match what the tape was checked against.

12. What to compute: the 61 per cent belongs to the students' prevalence, about 0.32. Use the survey figure as the camp's prior, and recompute the PPV from the tape's sensitivity and specificity.

```
0.86 times 0.23 = 0.1978
1 minus 0.23 = 0.77
0.26 times 0.77 = 0.2002
0.1978 plus 0.2002 = 0.398
0.1978 divided by 0.398 = 0.497
```

About 50 per cent, not 61.

What this does not establish. The survey's figure counts a body-mass index of 25 or more. The tape's 86 and 74 per cent fit the 25-to-29.9 band. So the prior and the test do not measure the same thing, and 0.50 mixes them. The survey figure is for rural men in all of India, aged 15 to 49; the men who come to one camp may differ. And the tape's rates come from medical students. Say "about half, and we will confirm every positive", and ask which prior fits the men you expect.

6 · The base-rate case that defeats most clinicians

EXERCISE 1. A test for a disease with a prevalence of 1 in 1,000 has a false-positive rate of 5 per cent. What is the chance that a person who tests positive has the disease, knowing nothing else about them?

Right answer: about 2 per cent, assuming the test catches everyone with the disease. Of 1,000 people, 1 is ill and tests positive; about 50 of the 999 well people test positive too. One in about 51.

Commonest wrong answer: 95 per cent. That is 1 minus the false-positive rate, the specificity, P(test negative | no disease). It is given as if it were P(disease | test positive).

EXERCISE 2. Draw 1,000 people as a box. Split off the one person with the disease, and the 999 without.

Ask her: of the one who is ill, how many test positive? Assume all, so one. Of the 999 who are well, how many test positive? Five per cent: about 50. Write 1 and 50 side by side under the word “positive”.

Now ask her question again, pointing at those two numbers: of the people who test positive, how many are ill? One of about 51. About 2 per cent.

Then show where her 95 per cent lives. Circle the 999 well people: 95 per cent of them test negative. That is the specificity, a share of the well. Her question was a share of the positives. Write the two conditionals one above the other: P(negative | well) and P(ill | positive). Same words, different order, different denominator.

End with the move she can use next time: before answering, ask how common the disease is in the group the patient comes from. If no one can say, no one can answer.

EXERCISE 3. Who: every resident and faculty member who attends one departmental meeting, or a list drawn at random from the department’s roster. Say which, because a convenience sample, like Manrai’s, describes only the people who were there.

The question: use the Casscells wording unchanged, so your result can be set beside the two studies. Ask each person to answer alone, on paper, in a few minutes.

Correct: decide before you look at the answers. Manrai counted “2%”, “1.96%” or “<2%”. Write your rule down first, so it cannot drift to fit the answers.

Record: the answer, and the person’s training level. Count the share correct and the share answering 95 per cent.

What it could show: how common the error is among those people, on that day, on that question. What it could not show: whether the error changes what they do with real patients, or how the whole profession in India would answer. A few dozen people give a rough share, not a precise one.

1.

```
10000 divided by 500 = 20
10000 minus 20 = 9980
0.02 times 9980 = 199.6
20 plus 199.6 = 219.6
20 divided by 219.6 = 0.0911
```

20 people have the condition and all test positive. About 200 of the 9,980 without it test positive. Of about 220 positives, 20 have the condition: a chance of about 9.1 per cent.

2. Call the prevalence p. Out of everyone tested, a share p have the condition and test positive, and a share 0.05 × (1 − p) test positive without it. The chance is 50 per cent when the two are equal.

```
p = 0.05 × (1 - p)
p = 0.05 - 0.05 × p
```

Add 0.05 × p to both sides.

```
1.05 × p = 0.05
0.05 divided by 1.05 = 0.0476
```

A prevalence of about 4.8 per cent, about 1 in 21. Check: of 1,000 people, about 48 have it; 0.05 times 952 is about 48 false positives. Equal, so a positive is right half the time.

3.

```
1000 divided by 20 = 50
1000 minus 50 = 950
0.05 times 950 = 47.5
50 plus 47.5 = 97.5
50 divided by 97.5 = 0.513
```

50 patients have the disease and all test positive. About 48 of the 950 without it test positive too. The chance is about 51 per cent, about 26 times the 2 per cent of the Casscells question, with the same test. The prevalence rose fifty-fold; the chance rose about 26-fold.

4.

```
0.20 times 1000 = 200
1000 minus 200 = 800
0.86 times 200 = 172
1 minus 0.74 = 0.26
0.26 times 800 = 208
172 plus 208 = 380
172 divided by 380 is about 0.453
```

Of 380 men flagged by the tape, 172 are in the band: about 45 per cent, the figure Book 0 reached for the same group. The colleague's 74 per cent is the specificity, P(tape negative | not in the band): a share of the 800 men outside the band. It is the Casscells error with a real test.

The 45 per cent holds for a group where one man in five is in the band. Kiran's figures come from 131 medical students in one college, printed to whole per cents.

5. The second line broke. "Right 95 per cent of the time" is true only of the people without the disease: 95 per cent of them test negative. That is P(test negative | no disease). It says nothing about how often a positive result is right.

The third line then applies the 95 per cent to the positives, a different group. Count it:

```
0.05 times 999 = 49.95
1 divided by 50.95 = 0.0196
```

About 2 per cent. Why the wrong answer looks reasonable: "accuracy" sounds like one property of the test that applies to every result. It is not. A share of the healthy and a share of the positives are different fractions, and only the prevalence connects them. The prevalence never appears in the wrong working, which is the tell.

6. The third line broke. A false-positive rate of 5 per cent means 5 per cent of the people without the disease test positive: 5 per cent of the 999. It is not 5 per cent of the positives. The worker took the right numerator, false positives, and divided by the wrong group.

```
0.05 times 999 = 49.95
1 plus 49.95 = 50.95
49.95 divided by 50.95 = 0.980
1 divided by 50.95 = 0.0196
```

About 98 per cent of the positive results are false, and about 2 per cent are true.

Why it looks reasonable: the first line set up the right group of 1,000, so the answer feels counted. But the 999 were never used. The name "false-positive rate" invites reading it as "the share of positives that are false". Say out loud what the rate is a share of before you use it.

7. What to compute: the chance of the condition after a positive result. That needs the prevalence in the camp's population, which the colleague did not give. So compute it at a few prevalences, assuming the test catches everyone who has the condition.

```
0.01 times 1000 = 10
0.05 times 990 = 49.5
10 divided by 59.5 = 0.168
0.10 times 1000 = 100
0.05 times 900 = 45
100 divided by 145 is about 0.690
```

At 1 per cent prevalence, about 17 per cent of those flagged have the condition. At 10 per cent, about 69 per cent. It reaches 95 per cent only when about half the people screened have it, which is not a screening camp.

What this does not establish. It does not say the referrals are wrong: a 17 per cent chance may be worth a specialist visit, and that is a clinical and cost judgement. It does not give the camp's real figure, which needs the camp's own prevalence and the test's real sensitivity. And it says nothing about the people the test missed.

8. Check where 18 per cent comes from and what it is a share of.

```
11 divided by 60 is about 0.183
14 divided by 61 is about 0.230
```

18 per cent is Casscells' 11 of 60 in 1978. The later study, Manrai's 14 of 61 in 2013, gave 23 per cent. Both are shares of about 60 people.

What it does not establish. "Doctors" in the story means 60 students, house officers and attending physicians met in Harvard hallways for a study published in 1978, not doctors in general, and not Indian doctors. It was one abstract question with no patient. It shows that this one question was answered wrongly by most of them. It does not show they "can't do maths". Many could compute but ignored the prevalence. It does not show how often the error changed care for a real patient: the authors themselves call that "less clear". A fair line would be: "in two small studies, at Harvard teaching hospitals in 1978 and a Boston-area hospital in 2013, about four in five clinicians gave a wrong answer, most often 95 per cent when the answer was 2."

7 · Natural frequencies: Bayes by counting people

EXERCISE 1.

1. Choose a group, say 1,000 people, and use the prevalence to find how many have the condition.
2. Use the sensitivity on those people to find how many have the condition and test positive.

3. Use the false-positive rate on everyone else to find how many test positive without the condition.
4. Divide the count from step 2 by the sum of the counts from steps 2 and 3.

Hoffrage and Gigerenzer: with probabilities the physicians reasoned correctly in 10 per cent of cases; with natural frequencies, 46 per cent.

EXERCISE 2. Start by unpacking "80 per cent accurate": it is the share of women with cancer whom the test catches. It says nothing about the women without cancer.

Then count. Picture 1,000 women of 40 at routine screening. About 10 have breast cancer, and 8 of those test positive. Of the 990 without cancer, about 99 also test positive. So 107 women walk out with a positive result, and 8 of them have cancer.

The quotable sentence: "Of every 107 women of 40 who test positive at routine screening, about 8 have breast cancer; a positive result means more tests, not a diagnosis."

Calibrate: these are rounded teaching numbers from a 1982 study, not today's Indian figures. The pattern, not the 8, is the point.

EXERCISE 3. A good panel has four parts.

The table: 1,000 patients split by the condition (from your prevalence), then each part by the test (sensitivity on those with it, the false-positive rate on those without). All four cells and the totals, rounded to whole people only at the last split.

Two sentences, for example: "Of 1,000 patients like you tested here, about X test positive. About Y of those X have the condition." And the same for a negative result.

The sources line: the paper and table the sensitivity and specificity came from, and the population they were measured in; the prevalence, with its source and period. If the prevalence is a guess, the handout says so.

Check the panel: every count is a whole person, the four cells add to 1,000, and Y ÷ X matches the PPV from the formula of section 5 to the rounding.

1.

```
0.10 times 1000 = 100
1000 minus 100 = 900
0.90 times 100 = 90
100 minus 90 = 10
0.20 times 900 = 180
900 minus 180 = 720
```

True positives 90, false negatives 10, false positives 180, true negatives 720. Check: 90 plus 10 plus 180 plus 720 is 1,000.

2.

```
90 plus 180 = 270
90 divided by 270 is about 0.333
720 plus 10 = 730
720 divided by 730 is about 0.986
```

PPV about 33 per cent: one positive in three is right. NPV about 98.6 per cent.

3.

```
40 divided by 2000 = 0.02
30 divided by 40 = 0.75
2000 minus 40 = 1960
1960 minus 98 = 1862
1862 divided by 1960 = 0.95
30 plus 98 = 128
30 divided by 128 is about 0.234
40 minus 30 = 10
1862 plus 10 = 1872
1862 divided by 1872 is about 0.995
```

Prevalence 2 per cent, sensitivity 75 per cent, specificity 95 per cent, PPV about 23 per cent, NPV about 99.5 per cent. The same counts give all five: this is the reverse of problem 1.

4.

```
0.01 times 10000 = 100
10000 minus 100 = 9900
0.80 times 100 = 80
0.096 times 9900 = 950.4
80 plus 950.4 = 1030.4
80 divided by 1030.4 is about 0.0776
```

100 women have breast cancer and 80 of them test positive. About 950 of the 9,900 without it test positive. The chance is about 7.8 per cent, the .078 the paper gives. The 10,000 avoids the 95.04 that 1,000 women would give.

5.

```
0.10 times 1000 = 100
1000 minus 100 = 900
0.86 times 100 = 86
100 minus 86 = 14
1 minus 0.74 = 0.26
0.26 times 900 = 234
900 minus 234 = 666
86 plus 234 = 320
86 divided by 320 is about 0.269
666 plus 14 = 680
666 divided by 680 is about 0.979
```

Of 320 men flagged by the tape, 86 are in the band: a PPV of about 27 per cent of flagged men. Of 680 men the tape clears, 666 are outside the band: an NPV of about 98 per cent of cleared men. At one in five, Book 0 found a PPV of about 45 per cent; halving the prevalence cut it to 27.

The 1 in 10 is stated for this problem, not measured. Kiran's figures come from 131 medical students in one college.

6. The second and third lines broke: the worker never split by the condition. Sensitivity applies only to the people who have the condition, and the false-positive rate only to those who do not.

```
0.01 times 1000 = 10
1000 minus 10 = 990
0.90 times 10 = 9
0.09 times 990 = 89.1
9 plus 89.1 = 98.1
9 divided by 98.1 = 0.0917
```

The PPV is about 9 per cent, not 91.

Why the wrong answer looks reasonable: it does count people, and the arithmetic in each line is right. But 900 true positives from a group in which only 10 have the condition is impossible. Check every count against the count above it.

7. The second line broke, at the rounding. A tenth of a person cannot be rounded to nobody without changing the question: the disease has vanished from the group. The group is too small for its smallest count. Take 1,000 people instead.

```
1000 divided by 1000 = 1
0.05 times 999 = 49.95
1 plus 49.95 = 50.95
1 divided by 50.95 = 0.0196
```

About 2 per cent, not 0. Note too that the wrong working took 5 per cent of all 100 people, not of those without the disease.

Why the wrong answer looks reasonable: rounding to whole people is usually right, and here it looks like the method's own rule. Round only when the count is large enough that rounding cannot remove a whole group.

8. Compute the PPV by counting 1,000 people.

```
0.01 times 1000 = 10
0.90 times 10 = 9
1000 minus 10 = 990
0.09 times 990 = 89.1
9 plus 89.1 = 98.1
9 divided by 98.1 = 0.0917
```

About 98 of every 1,000 people screened test positive, and about 9 of them have the condition. The rewrite: "Of every 1,000 people screened, about 98 get a positive result. About 9 of those 98 have the condition. A positive result means you need a further test; most people with a positive result do not have the condition."

What it does not establish. The 1 in 100 is for the people screened as a whole; a person with symptoms or a family history comes from a group where the condition is commoner. The test's 90 and 9 per cent were measured somewhere else and are carried over. The rewrite says nothing about whether screening is worth doing.

9. Check the ratio against Hoffrage and Gigerenzer's results, overall and by problem.

```
46 divided by 10 = 4.6
46 divided by 8 = 5.75
67 divided by 4 = 16.75
42 divided by 21 = 2
29 divided by 8 = 3.625
```

Overall, 46 per cent correct against 10 per cent is 4.6 times, so "almost five times" matches the headline result. By problem it ranged from twice to about 17 times.

What it does not establish. These were 48 German physicians, and the percentages are shares of cases, not of doctors. “Five times better” hides that 54 per cent of the answers with natural frequencies were still wrong. A ratio of two small percentages swings a lot: the colorectal 4 per cent was 1 of 24 physicians, as Hoffrage and colleagues later restated it. It was written problems, not patients, and it does not show that patients understood better.

8 · The odds form of Bayes’ theorem and the likelihood ratio

EXERCISE 1. Posterior odds = prior odds × LR, where the LR of a result is P(result | condition) ÷ P(result | no condition).

LR+ = sensitivity ÷ (1 − specificity). LR− = (1 − sensitivity) ÷ specificity.

You may multiply two likelihood ratios only when the two results are conditionally independent: independent among people with the condition, and independent among people without it. Odds go in and come out; turn a probability into odds first, with odds = p ÷ (1 − p), and back after, with p = odds ÷ (1 + odds).

EXERCISE 2. The prior odds are 0.1 against 0.9.

```
0.1 divided by 0.9 = 0.1111
0.1111 times 3.3077 = 0.3675
0.3675 divided by 1.3675 = 0.2687
0.1111 times 0.1892 = 0.0210
0.0210 divided by 1.0210 = 0.0206
```

After a positive tape, about 0.27. After a negative tape, about 0.02. The positive result does not reach one half: at a prior of one in ten, most positive tapes are in men outside the band. The ten per cent is your estimate for the exercise, not a measured figure.

EXERCISE 3. What is right: a positive answer makes the condition more likely, and the 2 says how much more often that answer turns up in people who have the condition than in people who do not.

What is wrong: the 2 multiplies the odds, not the risk. When the condition is rare, odds and risk are almost the same, so a rare risk roughly doubles: one in a hundred becomes about two in a hundred. When it is common, it does not: a half becomes two thirds, not one.

A sentence to quote: “A positive answer roughly doubles a small risk, but it cannot double a large one. And it tells you nothing about your own risk until you know how common the condition is in people like you.”

1.

```
1 minus 0.2 = 0.8
0.2 divided by 0.8 = 0.25
1 plus 1.5 = 2.5
1.5 divided by 2.5 = 0.6
```

Odds of 0.25, which is 1 to 4. A probability of 0.6.

2.

```
1 minus 0.8 = 0.2
0.9 divided by 0.2 = 4.5
1 minus 0.9 = 0.1
0.1 divided by 0.8 = 0.125
```

LR+ is 4.5 and LR− is 0.125.

3.

```
0.1 divided by 0.9 = 0.1111
0.1111 times 6 = 0.6667
0.6667 divided by 1.6667 = 0.4
```

The posterior probability is 0.4. Exactly: prior odds of 1 to 9 times 6 give 6 to 9, which is 2 to 3, and 2 out of 5 is 0.4.

4. A probability of 0.5 is odds of 1. So the LR must take the prior odds up to 1.

```
0.05 divided by 0.95 = 0.05263
1 divided by 0.05263 = 19.0
```

An LR of 19. Exactly: the prior odds are 1 to 19, and 19 times that is 1.

5.

```
1 minus 0.601 = 0.399
0.725 divided by 0.399 = 1.8170
1 minus 0.725 = 0.275
0.275 divided by 0.601 = 0.4576
0.101 divided by 0.899 = 0.11235
0.11235 times 1.8170 = 0.20414
0.20414 divided by 1.20414 = 0.1695
0.11235 times 0.4576 = 0.05141
0.05141 divided by 1.05141 = 0.0489
```

LR+ is about 1.82 and LR− about 0.46. A score of 60 or more takes the probability from about 0.10 to about 0.17, or 17 per cent of people. A lower score takes it to about 0.05. Neither moves it far: this is a score for deciding whom to test with blood glucose, not one that diagnoses.

6.

```
0.1 divided by 0.9 = 0.1111
0.1111 times 0.55 = 0.0611
0.0611 divided by 1.0611 = 0.0576
0.1111 times 0.76 = 0.0844
0.0844 divided by 1.0844 = 0.0778
```

About 0.058 and about 0.078. Neither rules the disease out: both LRs are far above the 0.1 that Deeks and Altman name for strong evidence, and the probability barely falls. Their own note says these bands “are not small enough to rule out the disease with confidence”.

7.

```
1 minus 0.628 = 0.372
0.895 divided by 0.372 = 2.406
0.75 divided by 0.25 = 3.0
3.0 times 2.406 = 7.218
7.218 divided by 8.218 = 0.8783
```

LR+ is about 2.41, the pre-test odds 3.0, the post-test odds about 7.22 and the probability about 0.878, as Altman and Bland print it. The probability after a positive scan looks high, but most of it was there before the scan: the prior was

already 0.75. An LR+ of 2.41 is a modest test. As they write, a high likelihood ratio does not by itself make a positive test a good indicator, and the reverse holds too: a high post-test probability does not make a test good.

8. The third line broke. An LR multiplies the odds, not the probability. Turn 0.2 into odds first.

```
0.2 divided by 0.8 = 0.25
0.25 times 3.3077 = 0.8269
0.8269 divided by 1.8269 = 0.4526
```

The answer is about 0.45, not 0.66. Why the wrong answer looks reasonable: at small probabilities, odds and probability are nearly equal, so multiplying the probability gives nearly the right answer. At 0.2 they have already parted. At a prior of 0.4 the same mistake would give 1.32, a probability above 1, which shows at once that the method is wrong.

9. The second line broke. LR− is P(negative | in the band) ÷ P(negative | not in the band). A man in the band tests negative with probability 1 − sensitivity; a man outside it with probability equal to the specificity.

```
1 minus 0.86 = 0.14
0.14 divided by 0.74 = 0.1892
```

LR− is about 0.19, and a negative tape lowers the odds. The worked answer turned the ratio upside down. Why it looks reasonable: both numbers in it are the right two numbers. Only their order is wrong. Say each one as a conditional aloud before you divide: “negative given in the band” goes on top. The last line should have been a warning in itself: a negative test that raises the probability is a sign the arithmetic, not the test, is wrong.

10. Compute the posterior probability from each prior, and divide it by the prior.

```
0.01 divided by 0.99 = 0.010101
0.010101 times 2 = 0.020202
0.020202 divided by 1.020202 = 0.0198
0.0198 divided by 0.01 = 1.98
0.5 divided by 0.5 = 1
1 times 2 = 2
2 divided by 3 ≈ 0.6667
0.6667 divided by 0.5 = 1.333
0.6 divided by 0.4 = 1.5
1.5 times 2 = 3
3 divided by 4 = 0.75
0.75 divided by 0.6 = 1.25
```

From 0.01 the chance almost doubles, to about 0.02. From 0.5 it rises by a third, and from 0.6 by a quarter. “Doubles” is true of the odds, and of the chance only when it is small.

What it does not establish. The 2 itself is an estimate from whatever group the app’s makers tested, and may not hold for you. A doubled small chance is still small: 0.02 means 98 people in 100 who answer yes do not have the condition. And the claim says nothing about your own prior, without which no posterior can be stated.

11. Compute the posterior the colleague has in mind, then ask whether the multiplication was allowed.

```
0.25 times 3.3077 times 4.5 = 3.7212
3.7212 divided by 4.7212 = 0.7882
```

Even on the colleague’s own terms the probability is about 0.79, not “almost certainly”. About one such man in five would still be outside the band.

And the multiplication is not allowed as it stands. Both tapes read the same thing, body size. Among men in the band, a man with a large arm is likely to have a large neck too, and the same outside it. The two results are not condi-

tionally independent, so the combined LR is less than 15 by an unknown amount. What you can say is the arm tape's result alone, about 0.45, and that the neck tape adds something, but less than the product suggests.

What it does not establish. No combined likelihood ratio can be stated without a study that measured both tapes on the same men. Neither tape measures body fat; both stand in for the body-mass index band, which is itself a stand-in.

9 · The prosecutor's fallacy: P(evidence | innocent) is not P(innocent | evidence)

EXERCISE 1. It reads P(evidence | innocent) as P(innocent | evidence). The first is out of innocent people; the second is out of everyone the evidence fits.

The four steps: say both conditionals in words; ask "out of whom?" for each; count a population, such as 1,000 or 10 lakh, through both splits; say which number answers the question being asked.

The mirror error counts everyone in a large population the evidence would also fit, and leaves out the other evidence that may have already narrowed the field. Both errors leave out the prior: one ignores it, the other uses the wrong one.

EXERCISE 2. The flaw is the "therefore". 1 in 10 lakh is P(match | not the source): out of people who are not the source, about 1 in 10 lakh would match. The sentence turns it into P(not the source | match). That needs a count of who else could have left the stain. If 50 lakh men could have, about 5 would match by chance, and a matching man is the source with probability about 1 in 6. If the other evidence leaves 20, a chance match among them is very unlikely.

A rewrite: "The DNA expert deposed that the profile is found in about 1 in 10 lakh persons who are not related to the source. How much this tells us about the accused depends on how many other persons could have left the stain, which the rest of the evidence must settle."

EXERCISE 3. Two things went wrong, and neither was the multiplication.

First, it treated two deaths in one family as if they were two unconnected events, like two coin tosses. Brothers share genes and homes, so after one sudden death a second is several times more likely. Squaring threw that away.

Second, it answered the wrong question. 1 in 73 million was a guess at how rare two natural deaths are. The jury needed to know whether two natural deaths or two murders were more likely, given that two deaths had happened. Two murders in one family are rarer still. One later estimate put double natural deaths about nine times more common than double murders.

A sentence to quote: "A rare event under one explanation is not evidence for the other explanation, unless the other is less rare."

1.

```
9 divided by 900 = 0.01
9 plus 90 = 99
9 divided by 99 ≈ 0.0909
```

P(E | H) is 0.01: out of the 900 with H. P(H | E) is about 0.091: out of the 99 with E. The same 9 people sit on top of both. The bottoms differ.

2.

```
10,000 times 0.02 = 200
10,000 minus 200 = 9,800
200 times 0.95 = 190
9,800 times 0.10 = 980
190 plus 980 = 1,170
190 divided by 1,170 ≈ 0.162
```

P(B | A) is about 0.16, not 0.95. Of the 1,170 people with A, only 190 have B.

3.

```
1,000 times 1,000 = 10,00,000
1,000 divided by 8 = 125
1,000 times 125 = 1,25,000
10,00,000 divided by 1,25,000 = 8
```

The second event has a probability of 8 in 1,000, which is 1 in 125. Both happen with probability 1 in 1,000 times 1 in 125, which is 1 in 1,25,000. Squaring gives 1 in 10 lakh, 8 times too rare.

4.

```
27,50,000 divided by 3,78,000 ≈ 7.3
7.3 divided by 5 ≈ 1.5
```

About once every 7 years if the deaths were independent; about once every year and a half if a second death is 5 times as likely. Watkins gives “once every seven years” and “about once every year and a half”. Either way, somewhere in England a family will suffer two, and its rarity in one family says nothing about guilt in that family.

5.

```
5,00,000 divided by 1,00,000 = 5
1 divided by 6 ≈ 0.167
50 divided by 1,00,000 = 0.0005
1 divided by 1.0005 ≈ 0.9995
```

With 5 lakh men, about 5 innocent men match, and a matching man is the source with probability about 1 in 6. With 50 men, about 0.9995. The RMP did not change. The pool did.

6.

```
18 divided by 2 = 9
18 plus 2 = 20
18 divided by 20 = 0.9
1,300 times 1,300 = 16,90,000
16,90,000 divided by 5 = 3,38,000
```

Odds of 9 to 1, a probability of 0.9 among the two explanations, and a probability of about 1 in 3,38,000 for two SIDS deaths in a family. The last number is very small and the first is large. That is not a contradiction: they answer different questions. The small one is P(two deaths | SIDS). The 9 to 1 compares the two explanations once two deaths have happened.

7. The fourth line broke. 5 per cent is P(positive | no disease). The line states it as P(no disease | positive). Count 1,00,000 people, assuming the test catches everyone with the disease.

```
1,00,000 divided by 1,000 = 100
1,00,000 minus 100 = 99,900
99,900 times 0.05 = 4,995
100 plus 4,995 = 5,095
100 divided by 5,095 ≈ 0.0196
```

The chance the person has the disease is about 2 per cent, and the chance the positive is false is about 98 per cent. Why the wrong answer looks reasonable: both lines use the right numbers, and "5 per cent false positives" sounds like a fact about positive results. It is a fact about people without the disease.

8. The arithmetic in the third line is right.

```
7,30,00,000 divided by 7,00,000 ≈ 104
```

Two steps broke. The first line squared 1 in 8,543 as if two deaths in one family were independent; they are not. And the third line divided by every birth in the country, though the 1 in 8,543 applied only to families with the Clarks' low-risk features. Hill calls the deduction "complete nonsense" for that reason.

Why it looks reasonable: dividing a rate into births is the right move for an event that could happen in any family at the stated rate. Here the rate belonged to a small group, and the pair was not independent. Watkins put a double cot death at about once every year and a half somewhere in England. Hill put second cot deaths at "perhaps four or five times a year in the UK".

9. Compute the probability that a matching man is the source for each pool.

```
20,00,000 divided by 10,00,000 = 2
1 divided by 3 ≈ 0.333
20 divided by 10,00,000 = 0.00002
1 divided by 1.00002 ≈ 0.99998
```

With 20 lakh men who could have left the stain, about 2 innocent men match, and the accused is the source with probability about 1 in 3. With 20, about 0.99998. The story's 99.9999 per cent is 1 minus the RMP: it has transposed the conditional and assumed no one else could have left the stain.

What it does not establish. Even a man who left the stain may not have committed the crime. The report could be wrong through contamination or a mix-up, which the RMP does not cover. Relatives match more often than unrelated men. And the right pool comes from the rest of the evidence, which the story does not give.

10. The 40 per cent is P(score 60 or more | no diabetes), 1 minus the specificity. The colleague took 1 minus it as P(diabetes | score 60 or more). Compute the positive predictive value (PPV) by counting 1,000 people.

```
1,000 times 0.101 = 101
1,000 minus 101 = 899
101 times 0.725 = 73.225
1 minus 0.601 = 0.399
899 times 0.399 = 358.701
73.225 plus 358.701 = 431.926
73.225 divided by 431.926 ≈ 0.170
```

About 17 per cent, as the paper's own PPV of 17.0 per cent says. In three sentences: "The 40 per cent is out of people without diabetes, not out of people who score 60. Out of 1,000 people like those in the study, about 432 score 60 or more and about 73 of them have diabetes. So her chance is about 17 per cent, and the score tells us to test her glucose, not that she has diabetes."

What it does not establish. The 10.1 per cent is a Chennai prevalence; your clinic's may differ, and the answer moves with it. The sensitivity and specificity come from one study. And a score is not a diagnosis: only a glucose test settles it.

10 · Where the prior comes from: pre-test probability in your own clinic

EXERCISE 1. The prior is the probability that a person like this patient, in this setting, has the condition before the test result is known.

Estimate it as a proportion: the number who had the condition, divided by the number examined, in a stated group over a stated period.

The four checks:

1. The condition is defined as in the test's study, with the same reference test and cut-off.
2. The denominator holds only people the test would be applied to. People already known to have the condition are left out.
3. The people are like your patient in age, sex, setting and why they came.
4. The count covers a stated period, and everyone in it was examined, not only the suspected.

EXERCISE 2. A good plan names each item concretely.

The condition matches the study. For IDRS it is diabetes newly found by a 2-hour plasma glucose of 200 mg/dL or more, the World Health Organization criterion Mohan and colleagues used in 2005. It is not "diabetes" from any record.

The register is one in which the reference test was done. For IDRS that is a screening register with a glucose result against every name. A clinic list of people already on treatment is not enough.

The denominator is the people the test would be used on. For IDRS, adults not already known to have diabetes. Leave out known cases, and say how many you left out.

The period is stated, such as April to September, and long enough to hold more than a handful of cases. With five cases, one more or less moves the prior by a fifth.

The check: count how many in the denominator have a result for the reference test. If some do not, the prior is uncertain, and in which direction depends on who was skipped. Say so.

The sentence, for example: "Pre-test probability: 27 of 260 adults not known to have diabetes, screened with a glucose test at this centre from April to September, had diabetes (about 10 per cent)." The counts in that sentence are made up; yours come from your register.

EXERCISE 3. The NFHS-6 figure, 27.3 per cent, is a survey estimate of a parameter. The parameter is the share of all men aged 15 to 49 in India with a body-mass index of 25 or more, in 2023 to 2024. It describes India's men, not your clinic's. Its condition, 25 or more, is not the tape's condition, the 25 to 29.9 band.

Kiran and colleagues' 32 per cent is a proportion in one sample: 42 of 131 male medical students in one college, in the 25 to 29.9 band. It estimates the prevalence among students like them, and the sample was not drawn at random from any wider group.

The register's 20 per cent is a proportion among the men measured at your clinic in one period. It estimates the prevalence among men who come to your clinic.

The register figure belongs in the handout. It has the right condition, the right people and a stated period, provided every man in it was measured. The other two are true of their own groups.

1.

```
45 divided by 600 = 0.075
0.075 times 100 = 7.5
```

The prevalence is 0.075, or 7.5 per cent.

2. The false-positive rate is 1 minus 0.9, which is 0.1.

At 0.05:

```
0.8 times 0.05 = 0.04
0.1 times 0.95 = 0.095
0.04 plus 0.095 = 0.135
0.04 divided by 0.135 = 0.2962…
```

At 0.2:

```
0.8 times 0.2 = 0.16
0.1 times 0.8 = 0.08
0.16 plus 0.08 = 0.24
0.16 divided by 0.24 = 0.6666…
```

About 0.30 at a prior of 0.05, and about 0.67 at 0.2. The test did not change.

3. Use the odds form of section 8. The positive likelihood ratio (LR+) is the sensitivity divided by 1 minus the specificity.

```
1 minus 0.8 = 0.2
0.9 divided by 0.2 = 4.5
```

A PPV of 0.5 is posterior odds of 0.5 divided by 0.5, which is 1. Prior odds times 4.5 gives 1, so the prior odds are 1 divided by 4.5.

```
1 divided by 4.5 = 0.2222…
0.2222 divided by 1.2222 = 0.1818…
```

The prior is about 0.18. Check it.

```
0.9 times 0.1818 = 0.16362
0.2 times 0.8182 = 0.16364
0.16362 divided by 0.32726 = 0.49997…
```

About 0.5, as reported.

4. The score is used to find undiagnosed diabetes, so leave out the 40 who already knew.

```
300 minus 40 = 260
27 divided by 260 = 0.10384…
```

The prior is about 10.4 per cent: 27 of the 260 adults screened from April to June who did not already know they had diabetes.

The tempting answer is 67 of 300, about 22.3 per cent. That counts people the score would never be used on.

5.

```
1 minus 0.601 = 0.399
1 minus 0.05 = 0.95
0.725 times 0.05 = 0.03625
0.399 times 0.95 = 0.37905
0.03625 plus 0.37905 = 0.4153
0.03625 divided by 0.4153 = 0.08728…
```

About 8.7 per cent. Fewer than one in eleven people with a score of 60 or more in this camp will have undiagnosed diabetes, against 17.0 per cent in the paper's Chennai sample.

6.

```
36 divided by 42 = 0.8571…
0.8571 times 0.2 = 0.17142
65 divided by 89 = 0.7303…
1 minus 0.7303 = 0.2697
0.2697 times 0.8 = 0.21576
0.17142 plus 0.21576 = 0.38718
0.17142 divided by 0.38718 = 0.44274…
67 divided by 89 = 0.7528…
1 minus 0.7528 = 0.2472
0.2472 times 0.8 = 0.19776
0.17142 plus 0.19776 = 0.36918
0.17142 divided by 0.36918 = 0.46432…
```

About 44.3 per cent and 46.4 per cent. One man either way in the specificity moves the PPV by about one percentage point each way. Print "about 45 per cent". A second decimal would claim a precision a sample of 89 men cannot give.

7. The second line broke. The 365 includes 143 people who already knew they had diabetes. The score is never applied to them, so they belong in neither the top nor the bottom.

```
365 minus 143 = 222
2350 minus 143 = 2207
222 divided by 2207 = 0.1006…
0.725 times 0.1006 = 0.072935
0.399 times 0.8994 = 0.3588606
0.072935 divided by 0.4317956 = 0.1689…
```

About 16.9 per cent, which matches the paper's printed 17.0 within rounding. The draft's 25 per cent overstates it by half again.

Why it looks reasonable: 15.5 per cent is printed in the Results as "overall prevalence", and the arithmetic after it is right. The error is in which group the prior describes.

8. Every line of arithmetic is right. The prior is the step that broke, in two ways.

First, the condition. The tape's 86 and 74 per cent fit the men in the 25 to 29.9 band, as Book 0 found. NFHS-6's 27.3 per cent counts everyone at 25 or more. The prior and the test describe different conditions. So the last line's "25 or more" is not what the formula computed.

Second, the people. 27.3 per cent is all of India's men aged 15 to 49. Your clinic's men may be older, more urban, or there for reasons linked to weight. Use your register's count for men like your patients, in the band, over a stated period.

Why it looks reasonable: NFHS-6 is the most trusted prevalence in India, and it is a real number about the right sex. It is the prevalence of a different group and a different condition.

9. What to compute: the PPV at the camp's own prior, with the paper's sensitivity and specificity.

```
1 minus 0.2 = 0.8
0.725 times 0.2 = 0.145
0.399 times 0.8 = 0.3192
0.145 plus 0.3192 = 0.4642
0.145 divided by 0.4642 = 0.31236…
```

About 31 per cent: nearer one in three than one in six. The note carried the paper's PPV to a group with twice the paper's prior.

What it does not establish. The 20 per cent is a supposition; the camp's real prior has to be counted. The sensitivity and specificity were measured in a Chennai population sample. People who come to an awareness rally may differ, so the score may behave differently in them. And no PPV says which of the people with a high score have diabetes. Each still needs the glucose test.

10. What to compute: what the handout would print with the colleague's prior, and what it would print with a different one, to show how much rides on the choice.

```
1 minus 0.83 = 0.17
1 minus 0.307 = 0.693
0.88 times 0.307 = 0.27016
0.17 times 0.693 = 0.11781
0.27016 plus 0.11781 = 0.38797
0.27016 divided by 0.38797 = 0.69634…
0.88 times 0.1 = 0.088
0.17 times 0.9 = 0.153
0.088 plus 0.153 = 0.241
0.088 divided by 0.241 = 0.36514…
```

About 70 per cent at a prior of 0.307, and about 37 per cent at 0.1. The choice of prior moves the answer by about 33 percentage points.

What to say back. "Most reliable" is about how well NFHS-6 estimates its own figure: women aged 15 to 49 across India, body-mass index 25 or more. It says nothing about whether that is the prevalence among the women at this clinic. The right prior is counted from the clinic's own register, for women like the patients, with the condition defined as the test's study defined it. And the condition differs here too. The 88 and 83 fit the 41 women in the 25 to 29.9 band. NFHS-6 counts everyone at 25 or more.

What this does not establish. The 0.1 is a supposition chosen to show the effect. The tape's figures come from medical students in one college. And a reliable national number is not a wrong one: it is the right prior for a national question.

11 · Counting arrangements: factorials and "n choose k"

EXERCISE 1. n! is the product of the whole numbers from n down to 1. 0! is defined to be 1.

```
n choose k = n! ÷ [k! (n - k)!]
```

Counting k things from n in order gives n! ÷ (n − k)! arrangements. Each set of k turns up in that count once for every order it can be put in. There are k! such orders, so dividing by k! counts each set once.

EXERCISE 2. A good answer builds it from a list.

Write the four people as four places. One way to get 2 positives is ++−−, writing + for positive and − for negative. If each person is positive with chance p, independently, the chance of exactly that order is p × p × (1 − p) × (1 − p). The resident has worked out the chance of one order.

Now ask: is ++−− the only way to get 2 positives? List them: ++−−, +−+−, +−−+, −++−, −+−+, −−++. Six orders. Each has the same two p's and two (1 − p)'s, just in different places, so each has the same chance. "Exactly 2 positives" happens if any one of them happens, and they cannot happen together, so their chances add.

So the chance of exactly 2 positives is 6 times the chance of one order. The 6 is the number of ways to choose which 2 of the 4 places are positive: 4 choose 2. Check it on the board.

```
4 times 3 = 12
12 divided by 2 = 6
```

End with the rule: one order's chance, times the number of orders. Leaving out the count answers "the first two positive and the last two negative", which nobody asked.

1.

```
5 times 4 times 3 times 2 times 1 = 120
6 times 120 = 720
```

5! is 120. 0! is 1, by definition. 6! is 6 × 5!, which is 720.

2. Keep each pair in alphabetical order: AB, AC, AD, BC, BD, CD. Six pairs.

```
4 times 3 times 2 times 1 = 24
2 times 1 = 2
24 divided by 2 = 12
12 divided by 2 = 6
```

4! ÷ (2! × 2!) is 24 ÷ 4, which is 6, as listed.

3. Every factor of 5! is also a factor of 8!, so they cancel. What is left is 8 × 7 × 6.

```
8 times 7 times 6 = 336
```

8! ÷ 5! is 336.

4.

```
9 times 8 = 72
2 times 1 = 2
72 divided by 2 = 36
```

9 choose 2 is 36. 9 choose 7 is also 36: choosing 7 of 9 to take is choosing 2 of 9 to leave. In full, 9! ÷ (7! × 2!) cancels the 7! and leaves the same 9 × 8 ÷ 2.

5. Each count for 6 is the count above it plus the one above and to its left. A set of k from 6 either leaves out the sixth thing or includes it. If it leaves it out, it is a choice of k from the first 5. If it includes it, it is a choice of k − 1 from the first 5. The two ends are 1.

```
1 plus 5 = 6
5 plus 10 = 15
10 plus 10 = 20
10 plus 5 = 15
5 plus 1 = 6
```

The counts for 6 are 1, 6, 15, 20, 15, 6, 1. Check 6 choose 3.

```
6 times 5 times 4 = 120
3 times 2 times 1 = 6
120 divided by 6 = 20
```

6. Routes first, in order.

```
12 times 11 times 10 = 1320
```

Each set of 3 has 3! orders.

```
3 times 2 times 1 = 6
1320 divided by 6 = 220
```

220 sets of 3 subcentres, and 1,320 routes.

7. Choose which 2 of the 10 places hold the positives: 10 choose 2.

```
10 times 9 = 90
2 times 1 = 2
90 divided by 2 = 45
```

45 orders. The total number of orders of positives and negatives among 10 people is 2 to the power 10, as in Book 0, D2.

```
2^10 = 1024
45 divided by 1024 = 0.0439…
```

About 4.4 per cent of the orders. That share is a share of orders, not a probability of 2 positives: the orders are equally likely only if each person is positive with chance one half, independently.

8. Choosing the 36 caught is choosing the 6 missed, so 42 choose 36 is 42 choose 6.

```
42 times 41 times 40 times 39 times 38 times 37 = 3776965920
6 times 5 times 4 times 3 times 2 times 1 = 720
3776965920 divided by 720 = 5245786
```

5,245,786 different sets of 36 men. Working out 42 choose 36 directly would need 36 factors on top and 36! below, all to cancel to the same six.

9. The last line broke. The three members of a committee have no order: Dr X, Dr Y and Dr Z chosen in any order are the same committee. The product counts each committee once for each of its 3! orders.

```
3 times 2 times 1 = 6
336 divided by 6 = 56
```

There are 56 committees.

Why it looks reasonable: the stages are real, and the counting rule is right for them. It answers “chair, then secretary, then member”, which is a different question with 336 answers.

10. The first line broke. It divides by k! but not by (n − k)!. The formula is 10! ÷ (3! × 7!). The 7! cancels the bottom seven factors of 10!.

```
10 times 9 times 8 = 720
720 divided by 6 = 120
```

There are 120 ways.

Why it looks reasonable: “divide by k! because order does not matter” is the right idea. But 10! counts all 10 things in a row, not 3 chosen from 10. A count of 604,800 sets is more than the 720 ordered choices of 3 from 10. There can never be more sets than ordered choices, so it should have looked too large.

11. What to compute: the number of sets of 3 from 6.

```
6 times 5 times 4 = 120
3 times 2 times 1 = 6
120 divided by 6 = 20
```

20 sets, as the officer says. If one set had been drawn at random, each set equally likely, the chance of any one particular set would be 1 in 20, or 0.05.

What it does not establish. The 1 in 20 is a probability only if the set was drawn at random. Suppose the district picked PHCs for being willing, near or well staffed. Then there is no chance in the choice at all, and the count of possible sets says nothing about it. And even after a random draw, a 1 in 20 chance of picking exactly the best three says nothing about how many good PHCs ended up in the pilot. Ask how the three were chosen before anyone uses the 20.

12. What to compute: the report found the chance of one order, two positives first and eight negatives after. Count the orders, and multiply.

```
0.2 times 0.2 = 0.04
0.8^8 = 0.16777…
0.04 times 0.16777 = 0.0067108…
10 times 9 = 90
90 divided by 2 = 45
45 times 0.0067108 = 0.30198…
```

About 0.30. Two positives among ten is the most common single outcome at this chance, not an unusual one. The report left out the 45.

What it does not establish. The 0.30 needs each person to have the same 0.2 chance, independently of the others, which is what the next section names as the binomial conditions. Ten people from one household or one workplace may not be independent. And 0.2 is a stated chance in the report; where it came from decides whether 0.30 applies to this camp.

12 · The Bernoulli and binomial distributions: counting successes in independent trials

EXERCISE 1. The four conditions: a fixed number of trials, n; each trial a success or a failure; the trials independent; the same probability of success, p, on every trial.

```
P(X = k) = (n choose k) × p^k × (1 - p)^(n - k)
```

n choose k is the number of orders in which k successes and n − k failures can fall. In one given order, p^k is the chance of the k successes. $(1 - p)^{(n - k)}$ is the chance of the failures. Both come from multiplying across independent trials.

The mean is np and the variance np(1 − p). Both come from adding n Bernoulli variables. The mean needs no independence, because expected values always add. The variance needs it, because variances add only for independent variables.

EXERCISE 2. Start with one trial. Write X1 as 1 for a success and 0 for a failure. Work out its expected value from the definition: 0 × (1 − p) + 1 × p = p. Then its variance: the squared distance from p, weighted by probability, which simplifies to p(1 − p). Try p = 0.307 and get 0.307 and 0.2128.

Then write the count of n trials as X1 + X2 + ... + Xn, one Bernoulli per trial. Expected values add, always, so the mean is n times p. Variances add when the trials are independent, so the variance is n times p(1 − p).

End with the warning that makes it stick. If the women come in pairs from one house, the mean is still np. The variance is not, because the step “variances add” used independence. Show the 1.46 against 2.06 from the household example.

1.

```
0 times 0.7 plus 1 times 0.3 = 0.3
0.3 times 0.7 = 0.21
√0.21 = 0.458
```

E[X] is 0.3, Var(X) is 0.21 and SD(X) is about 0.458.

2.

```
5 choose 2 = 10
0.4^2 = 0.16
0.6^3 = 0.216
10 times 0.16 times 0.216 = 0.3456
```

P(X = 2) is 0.3456.

3. Work out the chance of none, then take it from 1.

```
0.9^6 = 0.531441
1 minus 0.531441 = 0.468559
```

About 0.469.

4.

```
50 times 0.2 = 10
50 times 0.2 times 0.8 = 8
√8 = 2.828
```

Mean 10, variance 8, standard deviation about 2.83.

5. The variance is the mean times (1 − p). So divide the variance by the mean.

```
4.2 divided by 6 = 0.7
1 minus 0.7 = 0.3
6 divided by 0.3 = 20
```

1 − p is 0.7, so p is 0.3. The mean np is 6, so n is 20.

6. Take p = 0.273, so 1 − p = 0.727. The count is binomial with n = 8.

```
0.727^8 = 0.0780
8 choose 2 = 28
0.273^2 = 0.074529
0.727^6 = 0.147641
28 times 0.074529 times 0.147641 = 0.3081
```

About 0.078 for none, and about 0.308 for exactly 2 men.

7.

```
50 times 0.307 = 15.35
50 times 0.307 times 0.693 = 10.6376
√10.6376 = 3.262
25 minus 15.35 = 9.65
9.65 divided by 3.262 = 2.96
```

The expected count is 15.35 women, with a standard deviation of about 3.26 women. A count of 25 is 9.65 women above the mean, about 2.96 standard deviations. That is far out for a random sample from this population, and worth checking: was the sample random, and from where?

8. Not binomial with n = 10: condition 4 fails, because p is 0.428 for five of the women and 0.255 for the other five. But each group of five is binomial, and the two groups are independent. So add expected values, and add variances.

```
5 times 0.428 = 2.14
5 times 0.255 = 1.275
2.14 plus 1.275 = 3.415
5 times 0.428 times 0.572 = 1.22408
5 times 0.255 times 0.745 = 0.949875
1.22408 plus 0.949875 = 2.173955
√2.173955 = 1.474
```

The expected count is about 3.42 women, the variance about 2.17 and the standard deviation about 1.47 women. The single binomial formula for P(X = k) does not apply. You would have to combine the two groups' distributions count by count.

9. A miss is a score under 60 in a person with diabetes. Its chance is 1 minus the sensitivity.

```
1 minus 0.725 = 0.275
8 times 0.275 = 2.2
0.725^8 = 0.0763
```

The score is expected to miss 2.2 of the 8, and flags all eight with a chance of about 0.076. Treating the eight as independent trials with one sensitivity assumes people like those in the CURES study.

10. The second line broke. It works out the chance of one particular order: the first two women at 25 or more and the other eight under. Any two of the ten could be the two. There are 10 choose 2 such orders.

```
45 times 0.0050 = 0.225
```

The chance is about 0.23, nearly one group in four.

Why the wrong answer looks reasonable: the multiplication rule is right, and every number in it is right. It answers a narrower question, about one order, than the one asked.

11. The second line broke. 10 × 0.307 is the expected count, not a probability. Adding the ten 0.307s would be adding the chances of events that can happen together, which the addition rule forbids. Use the complement.

```
0.693^10 = 0.0255
1 minus 0.0255 = 0.9745
```

The chance of at least one is about 0.97, not certainty.

Why the wrong answer looks reasonable: for one or two rare events, adding the chances gives nearly the right answer. It fails as the chances grow, and a probability above 1 is the signal.

12. The second line broke. The 10 women come in pairs from 5 houses. Two women who share a house and a kitchen are not independent trials, so condition 3 fails. The mean, 3.07, is still right. The standard deviation and the chance of 0 are not.

How far off they are depends on how tied the pairs are. In the extreme case, where each pair always matches, the count is twice a binomial with n = 5.

```
4 times 5 times 0.307 times 0.693 = 4.2550
√4.2550 = 2.063
0.693^5 = 0.1598
```

The standard deviation could be as large as 2.06, and the chance of 0 as large as 0.16.

Why the wrong answer looks reasonable: n is fixed, each answer is yes or no, and each woman alone has the same p. Three conditions plainly hold, and the fourth is the one nobody looks at.

13. Compute the binomial chance of exactly 3 in 10, with p = 0.307 from NFHS-6.

```
120 times 0.028934 times 0.076760 = 0.2665
```

About 0.27. “In any group” is wrong: nearly three random groups of 10 in four hold some other count, and about 0.026 hold none.

What it does not establish. The 0.27 holds for 10 women picked at random from all of India’s women aged 15 to 49, taking the survey estimate as exact. A real group is a ward, a self-help group or a clinic queue. Its women share a place and a way of life. Their p may be nearer the urban 42.8% or the rural 25.5% than to 30.7%. And “overweight” on the slide means a body-mass index of 25 or more, overweight and obesity together.

A better line: “Nearly one woman in three aged 15 to 49 has a BMI of 25 or more. In a random group of 10, expect about 3 on average, and anywhere from 1 to 5 is common.”

14. Compute what 20 random women would give if p were 0.307: the mean, the standard deviation, and the chance of 11 or more.

```
20 times 0.307 = 6.14
20 times 0.307 times 0.693 = 4.2550
√4.2550 = 2.063
11 minus 6.14 = 4.86
4.86 divided by 2.063 = 2.36
```

Eleven is 4.86 women above the expected 6.14, about 2.4 standard deviations. For the chance of 11 or more, add the binomial chances of 11, 12, … up to 20: no group can have two counts.

```
167960 times 0.307^11 times 0.693^9 = 0.0141
125970 times 0.307^12 times 0.693^8 = 0.0047
77520 times 0.307^13 times 0.693^7 = 0.0013
14 to 20 together = 0.0003
0.0141 plus 0.0047 plus 0.0013 plus 0.0003 = 0.020
```

The first numbers are 20 choose 11, 20 choose 12 and 20 choose 13. So 11 or more would be uncommon for 20 random women from the national population.

What it does not establish. The 20 women were not picked at random from India. They came to one centre, and they share a place, its food and its income. The national 30.7% is the wrong comparison for them. If the centre is urban, the comparison is 42.8%. If rural, 25.5%. One count of 20 does not show that the centre differs from its own district, nor why. Get the district or block figure, and a sample whose picking you can describe, before writing “problem”.

13 · The geometric and hypergeometric distributions: waiting for the first success, and drawing without replacement

EXERCISE 1. Geometric, p the chance of success on each independent trial:

```
P(X = k) = (1 - p)^(k - 1) × p
E[X] = 1 ÷ p
```

Hypergeometric, n drawn without replacement from N of whom K have the feature:

```
P(X = k) = (K choose k) × (N - K choose n - k) ÷ (N choose n)
E[X] = n × K ÷ N
```

The binomial with n and p = K ÷ N stands in when N is large against n. A rule of thumb is a draw of under a tenth of the group.

EXERCISE 2. Start from the conditions. She draws 20 of 60 files without putting any back. That is a third of the register, well over a tenth, so the draws are not nearly independent. The binomial is the wrong model.

Draw the small ward on the board: 12 patients, 4 at 25 or more, 3 picked. The binomial says 0.296 for none; counting sets says 0.2545. Show why: after each draw, the pool changes.

Then say what stays and what changes. The mean is the same, n × K ÷ N, because expected values add with or without independence. The spread is smaller than the binomial's. So the binomial would overstate how often she sees no errors or very many.

End with the catch. The hypergeometric needs K, the number of files with errors, which is the thing she wants to learn. She can use it to ask: if K were 6, how likely is what I found? That question is where the next rung, inference, begins.

1.

```
0.8^2 times 0.2 = 0.128
1 divided by 0.2 = 5
```

P(X = 3) is 0.128 and E[X] is 5.

2. More than 4 trials are needed exactly when the first 4 all fail.

```
0.9^4 = 0.6561
1 minus 0.6561 = 0.3439
```

P(X > 4) is 0.6561 and P(X ≤ 4) is 0.3439.

3.

```
3 choose 1 = 3
7 choose 3 = 35
10 choose 4 = 210
3 times 35 divided by 210 = 0.5
4 times 3 divided by 10 = 1.2
```

The chance is 0.5, and the expected number is 1.2.

4. The mean is 1 ÷ p, so p is 1 ÷ 8.

```
1 divided by 8 = 0.125
```

p is 0.125. The first trial is a success with probability p, 0.125.

5. Take p = 0.17, each person independent.

```
1 divided by 0.17 = 5.88
0.83^5 = 0.3939
```

About 5.9 people on average, and a chance of about 0.39 that the first 5 confirm none. That holds only if the clinic's people with a score of 60 or more are like those in CURES.

6.

```
52 choose 3 = 22100
131 choose 3 = 366145
22100 divided by 366145 is about 0.0604
52 divided by 131 is about 0.3969
0.3969^3 = 0.0625
```

About 0.060 without replacement, against 0.063 from the binomial. Three is a small share of 131, so the two are close. The binomial is a little higher, as it always is for "all".

7. (a) A fixed 50, a tiny share of lakhs: binomial, n = 50, p = 0.307. (b) Stop at the first: geometric, p = 0.307. (c) A fixed 10 from a known group of 131 with 52 at 25 or more: hypergeometric.

```
50 times 0.307 = 15.35
1 divided by 0.307 = 3.26
10 times 52 divided by 131 is about 3.97
```

Expected counts: 15.35 women; 3.26 women measured; 3.97 men. In (c) the draw is about 7.6 per cent of the group, so the binomial with p = 0.397 would be close as well.

8. The second line broke. The first success on the fourth trial means three failures, then the success: $(1 - p)^{(k - 1)}$ p with k = 4.

```
0.693^3 times 0.307 = 0.1022
```

The chance is about 0.102. The working gave the chance that the first case is the fifth woman.

Why the wrong answer looks reasonable: k appears as the exponent in the binomial, so it is natural to write it there again. The trial that succeeds is not a failure, and the second form in the definition, which counts the failures, adds to the muddle.

9. The third line broke. The audit picks 3 different patients, so it draws without replacement from a small group. After one patient at 25 or more is drawn, 3 of the 11 left are at 25 or more. Count sets.

```
4 choose 3 = 4
12 choose 3 = 220
4 divided by 220 is about 0.0182
```

The chance is about 0.018, half the binomial's 0.037.

Why the wrong answer looks reasonable: p = 4 ÷ 12 is right for the first pick. And the binomial is the distribution the reader has just learned. The draw is a quarter of the ward, far over a tenth, and the binomial doubles the chance of "all three".

10. Compute the chance of no case in three, with p = 0.307 from NFHS-6 and the women as independent.

```
0.693^3 = 0.3328
1 minus 0.3328 = 0.6672
```

Under those assumptions, about a third of such searches find no case in the first three women. "At the latest" is wrong; there is no latest. The mean, 3.26 women, is not a ceiling.

What it does not establish. Houses in one lane are not random picks from India: neighbours share income, food and work, so p varies from lane to lane. A house may hold several women, or none aged 15 to 49. And 30.7% counts a body-mass index of 25 or more, overweight and obesity together. A better line: "On average she will measure about three women to find one, and some searches will take much longer."

11. Compute the chance of no man at 25 or more in 10 drawn without replacement, if 52 of the 131 really are.

```
131 minus 52 = 79
79 choose 10 = 1440680596355
131 choose 10 = 288417894029200
1440680596355 divided by 288417894029200 is about 0.0050
```

About 0.005, one draw in two hundred. So the colleague's result would be rare if the 52 were right. It is a reason to check, not a proof.

What it does not establish. A rare result happens; one draw cannot settle it. The draw may not have been random: records filed by date or by hostel can cluster. The colleague may have used a different cut-off, or read the paper's "overweight" count, which is the 25 to 29.9 band only. The way to settle it is to recount Table 2, not to argue from one sample of 10.

14 · The Poisson distribution: counts of events at a steady rate

EXERCISE 1. Events happen at a constant average rate; counts in intervals that do not overlap are independent; two events never happen at exactly the same instant.

$P(X = k) = e^{(-\lambda)} \times \lambda^k \div k!$, for $k = 0, 1, 2$ and so on. Here λ is the average count in the interval.

The mean and the variance are both λ, so the standard deviation is $\sqrt{\lambda}$.

A variance well above the mean is overdispersion. The events cluster, or the rate changes between intervals, so the single-rate Poisson model does not fit.

EXERCISE 2. What to say, in plain words. Rare events at a steady rate do bunch up by chance. With an average of one a year, a block has about a 1.9 per cent chance of four or more in a given year. That sounds small. But a district has many blocks, and each one has that chance every year. Across 50 blocks you would expect about one block like this in a typical year, with no cause at all.

What I would want to know. How many blocks and years were looked at before this one was noticed? Do the four cases share anything beyond the place, such as a type of cancer or a time? Has the block's population grown?

A sentence to quote: "Four cases where we expect one is worth checking. On its own it does not show a cause, because chance alone produces a block like this somewhere in most years."

The working is for you, not for the reporter. At $\lambda = 1$, the chance of 0, 1, 2 or 3 cases is $e^{(-1)} \times (1 + 1 + 1/2 + 1/6)$, which is 0.9810. So $P(X \geq 4)$ is about 0.019, and 50 times 0.019 is 0.95. The 50 blocks are an assumption for the example; use the district's real number.

1.

```
P(X = 0) = e^(-3) = 0.049787
P(X = 2) = e^(-3) × 3^2 ÷ 2!
0.049787 times 9 divided by 2 = 0.2240
```

$P(X = 0)$ is about 0.0498 and $P(X = 2)$ is about 0.2240.

2. Add the first three chances. $P(X = 1)$ is $e^{(-3)} \times 3$.

```
0.049787 times 3 = 0.1494
```

```
0.0498 plus 0.1494 plus 0.2240 = 0.4232
1 minus 0.4232 = 0.5768
```

$P(X \leq 2)$ is 0.4232 and $P(X \geq 3)$ is 0.5768. "3 or more" has no upper end, so work it out as 1 minus the chance of 0, 1 or 2.

3.

```
12 times 10 divided by 60 = 2
e^(-2) = 0.1353
```

λ is 2 for ten minutes, and the chance of none is 0.1353.

4. Each chance is the one before times λ ÷ k.

```
2.5 divided by 4 = 0.625
```

The ratio is 0.625.

For the second count, $P(X = 0) = e^{(-\lambda)} = 0.05$. Take natural logarithms of both sides, as in Book 2, section 4: $-\lambda = \ln 0.05$, so $\lambda = \ln 20$.

```
1 divided by 0.05 = 20
ln 20 = 2.996
```

λ is about 3.0.

5.

```
e^(-4.4) = 0.01228
365 times 0.01228 = 4.48
```

The chance of a day with none is about 0.0123. Expect about 4 or 5 such days in a year of 365. A day with no heart attacks in the city is as much a product of chance as a day with nine.

6.

```
200 times 0.0102 = 2.04
```

λ is 2.04. No days with low activity means 200 failures in the binomial.

```
(1 - 0.0102)^200 = 0.9898^200 = 0.1287
e^(-2.04) = 0.1300
```

Binomial 0.1287, Poisson 0.1300. They differ in the third decimal place, because p is small and n is large.

7. The third line broke. λ = 6 is the average for two hours. The question is about 15 minutes, one eighth of two hours.

```
6 divided by 8 = 0.75
e^(-0.75) = 0.4724
0.75 times 0.4724 = 0.3543
1 minus 0.4724 minus 0.3543 = 0.1733
```

The chance is about 0.173, as OpenStax gives it (0.1734 by calculator), not 0.983.

Why the wrong answer looks reasonable: 6 is the only rate in the question, and the formula accepts any λ. Nothing in the arithmetic warns you. Only the interval does.

8. The second line broke. It chose a Poisson model without checking the spread. The sample variance of the ten counts is 108 divided by 9, which is 12. A Poisson count with mean 2 has variance 2.

```
108 divided by 9 = 12
12 divided by 2 = 6
```

The counts spread six times as much as a Poisson count would. The rate was not steady. So the 0.0011 is a chance under a model the data reject, and it says nothing about days 5 and 8.

Why it looks reasonable: the mean is right, and the formula for P(X ≥ 8) is applied correctly. The flaw is the model, not the arithmetic. If those were camp days, they need no "explaining" beyond the camp.

9. Compute the chance of 3 or more in one block in one month, then the chance that at least one of 30 blocks shows it.

```
e^(-0.5) = 0.6065
0.6065 times 0.5 = 0.3033
0.3033 times 0.5 divided by 2 = 0.0758
1 minus 0.6065 minus 0.3033 minus 0.0758 = 0.0144
```

For one block named in advance, the chance is about 0.014. Now all 30 blocks, if they are independent.

```
1 minus 0.0144 = 0.9856
0.9856^30 = 0.6472
1 minus 0.6472 = 0.3528
```

About a 35 per cent chance that some block in the district shows three or more in a given month. Over a year of months it becomes very likely.

What this does not establish. It does not show the three cases are chance; it shows chance is a reasonable explanation. Whether to investigate turns on other facts: whether the children share an exposure, a type, or a week of onset. And the 0.5 a month is itself an estimate from past years, not a known λ.

10. Ask how surprising 8 would be if nothing had changed. Take the three-year average of 4 as λ.

```
e^(-4) = 0.0183
```

The chances of 0 to 7 are 0.0183, 0.0733, 0.1465, 0.1954, 0.1954, 0.1563, 0.1042 and 0.0595.

```
0.0183 plus 0.0733 plus 0.1465 plus 0.1954 plus 0.1954 plus 0.1563 plus 0.1042 plus
0.0595 = 0.9489
1 minus 0.9489 = 0.0511
```

With a steady λ of 4, a month of 8 or more comes about one month in 20. With a standard deviation of √4, which is 2, a count of 8 is two standard deviations above 4.

What this does not establish. One month of 8 does not show the rate doubled. It also says nothing about the area. Diagnoses depend on how many people were screened. A camp or a new blood-pressure machine raises them with no change in anyone's blood pressure. A "rise in hypertension" needs a prevalence, with a denominator.

15 · The normal distribution: what generates it, and areas under it

EXERCISE 1. A sum of many small independent contributions, none dominating, is near-normal: the central limit theorem.

$z = (x - \mu) \div \sigma$, the distance from the mean in standard deviations.

About 68, 95 and 99.7 per cent (0.6827, 0.9545 and 0.9973). 95 per cent lies within 1.96 standard deviations (to two decimals).

The sorted data values go up the side. Along the bottom go z-scores. Each is the point with the same share of a normal distribution's area to its left as the value has in the data. Near-normal data fall near a straight line.

EXERCISE 2. The plot shows the top end peeling away above the line while the bottom end stays on it. The largest values are larger than a normal distribution would give: a long right tail. With 160 points, a run of ten rising steadily away from the line is not wobble. It is a pattern.

What to do: look at the histogram and compare the mean with the median. If the median sits clearly below the mean, report the median with its quartiles, or work on the log scale. Do not use mean ± 2 SD as a range for these women.

A sentence: “Waist circumference was right-skewed (mean 84.1 cm, median 81.0 cm), so we report the median and interquartile range.” The two numbers here are placeholders for the authors’ own.

1.

```
130 minus 120 = 10
10 divided by 8 = 1.25
```

z is 1.25: 130 lies 1.25 standard deviations above the mean.

2. The whole area is 1. The curve is symmetric about 0, so the area left of −1.25 equals the area right of 1.25.

```
1 minus 0.8944 = 0.1056
1 minus 0.1056 minus 0.1056 = 0.7888
```

Right of 1.25: 0.1056. Left of −1.25: 0.1056. Between them: 0.7888.

3.

```
50 minus 6 = 44
50 plus 6 = 56
50 minus 12 = 38
50 plus 12 = 62
50 minus 18 = 32
50 plus 18 = 68
```

About 68 per cent lie between 44 and 56. About 95 per cent lie between 38 and 62. About 99.7 per cent lie between 32 and 68.

4. Turn $z = (x - \mu) \div \sigma$ around: $x = \mu + z \times \sigma$.

```
1.2816 times 15 = 19.224
100 plus 19.224 = 119.224
```

The 90th percentile is about 119.2.

5.

```
0.9772 minus 0.1587 = 0.8185
```

About 0.8185. The area between two points is the area left of the upper one minus the area left of the lower one.

6. Sort first: 12, 15, 17, 18, 23. The shares:

```
0.5 divided by 5 = 0.1
1.5 divided by 5 = 0.3
2.5 divided by 5 = 0.5
3.5 divided by 5 = 0.7
4.5 divided by 5 = 0.9
```

The points, as (z, value): (−1.28, 12), (−0.52, 15), (0.00, 17), (0.52, 18), (1.28, 23).

The step from 18 to 23 is larger than the others. With five points that is not evidence of anything; it is what a plot of five values looks like.

7.

```
168 minus 170 = -2
-2 divided by 6.28 = -0.32
1.27 times 6.28 = 7.98
170 plus 7.98 = 177.98
```

168 cm is 0.32 standard deviations below the mean. A z-score of 1.27 is 177.98 cm, about 178 cm. These match OpenStax's answers.

8.

```
74 minus 70.0 = 4.0
4.0 divided by 3.3 = 1.21
1 minus 0.8869 = 0.1131
```

About 0.113, or 11 per cent, as OpenIntro gives it. The model is fitted to 100 men; the answer is only as good as the bell's fit at the upper tail.

9. Check the condition, then find the mean and standard deviation.

```
200 times 0.307 = 61.4
200 times 0.693 = 138.6
200 times 0.307 times 0.693 = 42.5502
√42.5502 = 6.523
```

Both 61.4 and 138.6 are at least 10. "75 or more" starts at the whole number 75, so widen by a half: take the area above 74.5.

```
74.5 minus 61.4 = 13.1
13.1 divided by 6.523 = 2.01
1 minus 0.9778 = 0.0222
```

About 0.022. The exact binomial answer is about 0.024.

10.

```
16.7 minus 22.1 = -5.4
-5.4 divided by 7.3 = -0.74
28.8 minus 22.1 = 6.7
6.7 divided by 7.3 = 0.92
1 minus 0.8212 = 0.1788
```

The normal model puts about 23 per cent below 16.7 and about 18 per cent above 28.8. The quartiles say a quarter of the men lie below 16.7 and a quarter above 28.8.

The model is close at the bottom and misses at the top, where it puts 18 per cent where the data have 25. That is the right tail the model does not have.

11. The last line broke. Software and tables give the area to the left. The question asks for the area to the right.

```
1 minus 0.6554 = 0.3446
```

The chance is 0.3446, as OpenStax §6.2 gives it.

Why the wrong answer looks reasonable: every step before the last is right, and 0.6554 is a real area under the curve. A sketch would have caught it: 65 is just above the mean, so less than half the area lies beyond it.

12. The second line broke. np is 1, far below 10. The normal approximation does not apply. Work out the binomial chance directly.

```
1 minus 0.05 = 0.95
0.95^20 = 0.3585
```

The chance of none is about 0.36, not 0.31. With np = 1 the binomial is piled up against 0 and cannot go below it, and a symmetric bell cannot copy that shape.

Why it looks reasonable: the mean and standard deviation are right, and the half-unit widening is right. The answer is even in the right region. Only the condition was skipped.

13. The last line broke. Points above the end-to-end line are the signature of left skew, not right. Check two points against the line.

```
26 minus 6.29 times 0.97 = 19.90
26 plus 6.29 times 0.28 = 27.76
```

The values there are 24 and 33, both above the line. The low end stretches out: 16 is 8 below the next value, while the top values are 1 apart. The tail is on the left. The median, 32, is above the mean, which is about 29.8.

```
16 plus 24 plus 28 plus 30 plus 32 plus 33 plus 34 plus 35 plus 36 = 268
268 divided by 9 is about 29.8
```

Why it looks reasonable: “the plot bends, so it is skewed” is right. Which way it is skewed comes from which side of the line the points fall, and that is easy to get backwards.

14. Check the cut-off, and the share above it if BMI were normal.

```
2 times 3.9 = 7.8
24.2 plus 7.8 = 32.0
```

The cut-off is right. Above z = 2, a normal distribution has 1 − 0.9772 of its area, which is 0.0228: about 2.3 per cent, not 2.5. The 2.5 is the share above 1.96.

What this does not establish. “Above two SDs” defines rare in this group, not abnormal. The share flagged is about 2.3 per cent in any group, healthy or not. A group where many students have obesity would flag the same few. A referral threshold should come from outcomes, as BMI cut-offs do.

And the 2.3 per cent assumes BMI is normal. The paper gives a median of 23.8, below the mean of 24.2. That hints at a right tail, so more than 2.3 per cent may lie above 32.0.

15. Ask how often a random sample of 400 at p = 0.307 gives 104 or fewer.

```
400 times 0.307 = 122.8
400 times 0.693 = 277.2
400 times 0.307 times 0.693 = 85.1004
√85.1004 = 9.225
104.5 minus 122.8 = -18.3
-18.3 divided by 9.225 = -1.98
```

The area to the left is about 0.024. The exact binomial gives about 0.022. Suppose the district's share were 30.7 per cent and the survey picked women independently. Then a count this low would come in about 1 survey in 45. Unusual, not "almost never".

What this does not establish. District surveys usually sample villages or households, then women within them. Women in one cluster are alike, so the count spreads more than the binomial says, and 104 is less surprising than 0.024 suggests. The national 30.7 per cent is not the district's own past figure, so this is not a change over time. And a lower share of overweight is not by itself "better". It may reflect more undernutrition, an older or younger sample, or a different measuring team.

16 · Skewed quantities and the log scale: when effects multiply

EXERCISE 1. Many small independent effects that add give a near-normal shape. Many small independent effects that multiply give a near-log-normal shape, because the logarithm turns their product into a sum.

To get a geometric mean, take the log of every value. Average the logs. Then raise e to that average. It is the same as multiplying the n values and taking the n-th root. For a log-normal quantity it estimates the median. It lies below the ordinary mean.

EXERCISE 2. For the men, quote the median and quartiles. The mean sits 1.8 points above the median, and the upper quarter-band, 8.5 points, is more than twice the lower one, 3.6. A mean and standard deviation would describe both sides as equally spread, and they are not.

For the women, either summary will do. The mean, 33.8, and the median, 34.0, nearly agree, and the quarter-bands, 5.9 and 5.2, are close.

What the table does not show: the whole shape. Three numbers per group can flag a tail and its direction; they cannot show how long it is. Nor does it say whether these medical students resemble any other group.

EXERCISE 3. A strong answer draws the 16 paths of four doubling-or-halving steps from 1 million cells. It lists the five end counts: 0.0625, 0.25, 1, 4 and 16 million, with 1, 4, 6, 4 and 1 paths. On the board the counts bunch low with one far out at 16. That is the long right tail.

Then take ln of each. Each doubling adds 0.693 and each halving takes it away, so the logarithms sit at minus 2.77, minus 1.39, 0, 1.39 and 2.77: balanced around 0. Multiplying became adding, and adding gives balance.

It shows that the mean count, about 2.44 million, sits far above the median, 1 million.

It ends on the limit. Logging balances a quantity only if its effects multiply. Kiran's women's body fat was nearly balanced already and logging would unbalance it. And zero or a negative value cannot be logged at all.

1.

```
ln 8 = 2.079
ln 0.125 = -2.079
e^1.5 = 4.482
```

The two logarithms are equal and opposite, because 0.125 is 1 divided by 8. The number whose logarithm is 1.5 is about 4.482.

2. Take ln of each value, average them, and turn the average back.

```
ln 2 = 0.693
ln 8 = 2.079
ln 32 = 3.466
0.693 plus 2.079 plus 3.466 = 6.238
6.238 divided by 3 = 2.079
e^2.079 = 8.0
```

The geometric mean is 8. Check it as the cube root of the product: 2 times 8 times 32 is 512, and 8 times 8 times 8 is 512.

```
2 plus 8 plus 32 = 42
42 divided by 3 = 14
```

The ordinary mean is 14. The geometric mean is the lower of the two.

3. They are already in order, so the median is the fourth value, 8.

```
1 plus 2 plus 4 plus 8 plus 16 plus 32 plus 64 = 127
127 divided by 7 is about 18.14
```

The ordinary mean is about 18.14. For the geometric mean, use $\log_{10}$, since each value is 2 to a power and any base gives the same answer.

```
log10 of 1 plus log10 of 2 plus log10 of 4 plus log10 of 8 plus log10 of 16 plus log10
of 32 plus log10 of 64 = 6.32
6.32 divided by 7 = 0.903
10^0.903 = 8.0
```

The geometric mean is 8, the same as the median. Each value is the one before times 2, so the logarithms are evenly spaced and balanced around the middle one. The ordinary mean, pulled by 64, is more than twice the median.

4. Divide and multiply the geometric mean by 1.16 for the first range.

```
60 divided by 1.16 = 51.7
60 times 1.16 = 69.6
```

About 68 per cent of onsets between about 51.7 and 69.6 years. For the second range, use 1.16 squared.

```
1.16 times 1.16 = 1.3456
60 divided by 1.3456 = 44.6
60 times 1.3456 = 80.7
```

About 95.5 per cent between about 44.6 and 80.7 years. Each range is wider above 60 than below it, but only a little, because a factor as small as 1.16 makes a log-normal shape almost balanced.

5.

```
83.5 minus 77.0 = 6.5
92.0 minus 83.5 = 8.5
83.5 divided by 77.0 = 1.084
92.0 divided by 83.5 = 1.102
```

In centimetres the upper band, 8.5 cm, is wider than the lower, 6.5 cm. As ratios, 1.084 and 1.102, they are closer but still a little lopsided. The log scale would move the men's waist towards balance without reaching it, as it did for their body fat.

Three numbers cannot show the whole shape. This is a direction, not a verdict.

6. The last line broke. The logarithms were taken to base 10, so the average must be turned back with 10, not with e.

```
10^1 = 10
```

The geometric mean is 10. Check it with natural logarithms all the way through.

```
ln 1 plus ln 10 plus ln 100 = 6.908
6.908 divided by 3 = 2.303
e^2.303 = 10.0
```

Why the wrong answer looks reasonable: "back-transform with e" is the usual instruction, because most software uses natural logarithms. The base you take logs with is the base you must turn back with. A geometric mean of 2.72 for three counts of at least 1 is a warning in itself: it sits nearer the smallest value than the middle one.

7. The third line broke. The back-transformed mean of the logarithms is the geometric mean. It estimates the median of a log-normal quantity, not its mean. The ordinary mean really is 4.1, and it is not an error. Both numbers are right; they answer different questions.

Why the wrong answer looks reasonable: the working did take a mean, of the logarithms, and it feels as if turning it back should give a mean. It gives the middle value instead, because the long right tail pulls the ordinary mean up and does not move the middle. Label 3.0 as the geometric mean every time it is printed.

8. Compute her range, then a range that respects the skew. If the quantity is log-normal, a mean and standard deviation can be converted to the log scale. First the coefficient of variation: the standard deviation divided by the mean.

```
3.7 divided by 4.1 = 0.902
1 plus 0.902 times 0.902 = 1.814
square root of 1.814 = 1.347
4.1 divided by 1.347 = 3.04
ln 1.814 = 0.596
square root of 0.596 = 0.772
e^0.772 = 2.16
```

The median is about 3.04 and the multiplicative standard deviation about 2.16. These match Limpert's 3.0 and 2.2, which the rest uses.

```
4.1 minus 7.4 = -3.3
4.1 plus 7.4 = 11.5
```

Her range runs from minus 3.3 to 11.5. Its lower end is impossible for a positive quantity. That is why she concluded that nothing can be too low; the method made it so.

```
2.2 times 2.2 = 4.84
3.0 divided by 4.84 = 0.62
3.0 times 4.84 = 14.5
```

On the log scale, about 95.5 per cent of values lie from 0.62 to 14.5. Values below 0.62 are as unusual as values above 14.5. Her upper limit, 11.5, is too low, and she has no lower limit at all.

What this does not establish. The 0.62 to 14.5 range assumes the quantity is log-normal; a paper that gives only a mean and a standard deviation does not show that. And a range that holds 95.5 per cent of a group describes the group. Whether a value outside it matters for a patient is a clinical question, and the range cannot answer it.

17 · Covariance and correlation of two random variables

EXERCISE 1. Var(X + Y) = Var(X) + Var(Y) + 2 Cov(X, Y). The last term is 0 when X and Y are independent, and also in some cases where they are not.

A correlation of 0 says there is no straight-line tie between the two. It does not say they are independent: a U-shaped tie, such as $Y = X^2$ with X balanced around 0, has a correlation of exactly 0 while X fixes Y.

EXERCISE 2. The 400 times 0.3 times 0.7 is the binomial variance. It needs the 400 adults to be independent. Two adults in one household share food, salt, a home and often family. So their results are likely to be tied, and in the same direction. Each household's count is then X + Y, with variance Var(X) + Var(Y) + 2 Cov(X, Y). A positive covariance adds to every household's share. So the true variance is larger than 84. The protocol understates the spread. Any precision worked out from it will look better than it is.

The fix is to find out how strongly two adults in one house are tied, from a pilot or an earlier survey. Then carry the covariance into the sum. Doing that in full belongs to the later book on study design and sample size.

1.

```
E[X] = 0.2 plus 0.3 = 0.5
E[Y] = 0.1 plus 0.3 = 0.4
E[XY] = 1 times 1 times 0.3 = 0.3
Cov(X, Y) = 0.3 minus 0.5 times 0.4 = 0.1
```

Cov(X, Y) is 0.1. Only the cell with X = 1 and Y = 1 has a product XY that is not 0.

2.

```
4 plus 9 plus 2 times 3 = 19
4 plus 9 minus 2 times 3 = 7
```

Var(X + Y) is 19 and Var(X − Y) is 7. With a positive covariance, the sum spreads more than the 13 the adding rule gives, and the difference spreads less.

3.

```
3 times 4 = 12
-6 divided by 12 = -0.5
```

The correlation of X and Y is minus 0.5. For the second, multiply both sides of the definition by the two standard deviations: the covariance is the correlation times SD(U) times SD(V).

```
0.8 times 2 times 5 = 8
```

Cov(U, V) is 8.

4. The dal's error is the second reading's error minus the first's. Use the rule for a difference.

```
50 plus 50 minus 2 times 30 = 40
square root of 40 = 6.32
50 plus 50 = 100
square root of 100 = 10
```

With the tie, the variance is 40 square grams and the standard deviation about 6.32 grams. Without it, 100 square grams and 10 grams. The shared part of the error cancels when you subtract, so a positive tie helps a difference.

5.

```
10 times 0.1 = 1
0.5 divided by 1 = 0.5
```

The correlation is 0.5. In centimetres every height is multiplied by 100. Its SD becomes 0.1 times 100, which is 10 cm. Every height deviation is 100 times larger, so the covariance is too.

```
0.5 times 100 = 50
10 times 10 = 100
50 divided by 100 = 0.5
```

The covariance becomes 50 kg·cm, a hundred times larger, and the correlation is still 0.5. The units changed the covariance and left the correlation alone.

6. First the covariance, from the correlation.

```
-0.3 times 150 times 200 = -9000
```

The covariance is minus 9,000, in kcal squared. Now the variance of the sum.

```
150 times 150 = 22500
200 times 200 = 40000
22500 plus 40000 plus 2 times -9000 = 44500
square root of 44500 = 210.95
```

The SD of the two meals together is about 211 kcal. If they were independent:

```
22500 plus 40000 = 62500
square root of 62500 = 250
```

250 kcal. The negative tie makes the total steadier than the parts suggest.

7. The last line broke. It adds the variances as if X and Y were independent. The first line shows they are not. Independence would need P(X = 1 and Y = 1) to be 0.5 times 0.2, which is 0.1, and it is 0.15.

```
0.15 minus 0.5 times 0.2 = 0.05
0.25 plus 0.16 plus 2 times 0.05 = 0.51
```

Var(X + Y) is 0.51, not 0.41.

Why the wrong answer looks reasonable: Book 2's adding rule is the first one learned, and the two variances were right. The joint probability that settles independence was given, and the working never compared it with the product.

8. The fourth line broke. The covariance is 0, which the first three lines get right. But a covariance of 0 does not make two variables independent. Here Y is fixed by X. When X is 1, Y is 1, so P(X = 1 and Y = 1) is P(X = 1), which is 0.25. P(Y = 1) is 0.25 plus 0.25, which is 0.5.

```
0.25 times 0.5 = 0.125
```

Independence would need the joint probability to be 0.125. It is 0.25, so X and Y are dependent.

Why the wrong answer looks reasonable: "independent gives zero covariance" is true and easy to remember, and the reverse sounds like the same sentence. It is not.

9. Compute the covariance of the three pairs. Read the three people as a random variable that takes each person's pair with probability one third. A correlation worked from data does the same sums. A covariance of 0 gives a correlation of 0 either way.

```
(5 plus 7 plus 9) divided by 3 = 7
(27 plus 23 plus 27) divided by 3 is about 25.67
-2 times 1.33 = -2.67
0 times -2.67 = 0
2 times 1.33 = 2.67
-2.67 plus 0 plus 2.67 = 0
```

The covariance, and so the correlation, is 0. The note's number is right.

What it does not establish. The pattern is a U: both the short and the long sleeper have the higher BMI. A correlation of 0 rules out only a straight-line tie. With three people it rules out very little at all. And a tie, if one were found, would not show that sleep changes BMI.

10. The protocol has added standard deviations. Compute the total's SD for the possible ties between the subscores: none, a moderate one, and a perfect one.

```
10 times 10 plus 15 times 15 = 325
square root of 325 = 18.03
0.4 times 10 times 15 = 60
325 plus 2 times 60 = 445
square root of 445 = 21.10
1 times 10 times 15 = 150
325 plus 2 times 150 = 625
square root of 625 = 25
```

Independent subscores give an SD of about 18.0. A correlation of 0.4 gives about 21.1. Only a correlation of exactly 1 gives the protocol's 25. Adding SDs assumes the two subscores are locked together.

What it does not establish. The correct SD depends on the correlation, which the protocol does not give; it has to be measured, in a pilot or from the instrument's own validation. And nothing here says whether adding the two subscores makes a sensible score at all.

18 · Sums and averages of many independent draws, simulated by hand

EXERCISE 1. The sum has mean nμ and SD σ times √n. The average has mean μ and SD σ divided by √n.

The central limit theorem says that as n grows, the distribution of the sum and of the average comes closer to normal, whatever the shape of one draw. It needs independent draws and a finite SD, and the n needed depends on how skewed one draw is.

EXERCISE 2. A good plan has these parts.

Each resident rolls four dice, writes the four faces and their average, and repeats this 10 times. That gives 10 averages each. Twelve residents give 120 averages.

Pool the averages on the board as a tally, in steps of 0.25 from 1 to 6. Beside it, pool the first die of every set as a tally from 1 to 6.

What the group should see: the single dice spread flat across 1 to 6; the averages crowd near 3.5 in a rounded hump. Work out the SD of each pool. The averages' should be near $\sigma \div \sqrt{4}$, about 0.854, against about 1.708 for one die.

End by naming what was built: the distribution of an average over many samples, the sampling distribution of the mean. Make clear that n is 4, the dice in one average, not 120, the number of averages.

1.

```
25 times 5 = 125
25 times 4 = 100
square root of 100 = 10
4 divided by 25 = 0.16
square root of 0.16 = 0.4
```

The sum has mean 125, variance 100 and SD 10. The average has mean 5, variance 0.16 and SD 0.4. Check: the SD of one draw is 2, and 2 divided by the square root of 25 is 0.4.

2. The SD of the average is 12 divided by $\sqrt{n}$. Set it to 3: $\sqrt{n}$ must be 12 divided by 3.

```
12 divided by 3 = 4
4 times 4 = 16
12 divided by 1 = 12
12 times 12 = 144
```

16 draws for an SD of 3, and 144 for an SD of 1. Cutting the SD to a quarter took 16 times the draws; to a twelfth, 144 times.

3.

```
4 times 4 = 16
5 times 16 = 80
square root of 80 = 8.94
4 divided by square root of 5 = 1.79
```

The weekly total has an SD of about 8.94 minutes, as OpenIntro gives. The average day of the week has an SD of about 1.79 minutes.

Both rest on the five days being independent. OpenIntro's own follow-up asks whether that holds: if Fridays are always worse, the days are tied, and the weekly SD is larger.

4.

```
0.307 times 0.693 = 0.2128
0.2128 divided by 400 = 0.000532
square root of 0.000532 = 0.02306
1.96 times 0.02306 = 0.0452
0.307 minus 0.0452 = 0.262
0.307 plus 0.0452 = 0.352
```

The SD of the share is about 0.023, or 2.3 percentage points. About 95 per cent of such samples would give shares between about 26.2 and 35.2 per cent. The normal shape is fair here because np is 400 times 0.307, which is about 123, and n(1 − p) is about 277, both well above 10 (section 15).

This is a statement about samples drawn from a population whose share is known. Turning one sample's share into a statement about an unknown share is the next rung's work.

5.

```
square root of 131 = 11.45
7.3 divided by 11.45 = 0.638
```

About 0.64 percentage points.

It assumes the 131 men are independent random draws from one population. They are one college's students who agreed to take part, so this is the spread of their mean only if they behave like a random sample. It also borrows the sample's own 7.3 as if it were σ. And it says nothing about whether the college's men resemble any wider group.

6. The third line broke. It uses the SD of one man as the spread of the mean of 131. The mean of 131 independent draws spreads with SD 7.3 divided by √131.

```
7.3 divided by square root of 131 = 0.638
```

About 0.64 points, not 7.3. The fourth line inherits the error and is about eleven times too wide.

Why the wrong answer looks reasonable: 7.3 is printed right beside 22.1 in the paper's table, and "mean (SD)" looks like "a value and its uncertainty". The SD in that table describes how the men differ from each other, not how far their mean might be from the population's.

7. The third line broke. The n in σ ÷ √n is the number of values in one average, which is 4, not the number of averages, which is 50.

```
1.708 divided by square root of 4 = 0.854
```

The averages should spread with an SD of about 0.854. Fifty repeats make the class's tally of averages look smoother. They do not make each average steadier.

Why the wrong answer looks reasonable: 50 is the biggest number in the problem and "more makes it steadier" is true of something here. It is true of how well the 50 averages show their own distribution, not of the spread of that distribution.

8. First pin down "precise". If it means the SD of the estimate, which sets the width of any interval built on it, compute how much that SD shrinks when n doubles. It goes as 1 divided by √n.

```
square root of 2 = 1.414
1 divided by 1.414 = 0.707
1 minus 0.707 = 0.293
```

Doubling n cuts the SD to about 0.707 of what it was: about 29 per cent smaller, not half. To halve it, the study needs 4 times 200, which is 800 participants.

If "precise" means 1 divided by the variance, doubling n does double it. The interval still narrows by only 29 per cent. So the proposal should say which it means, and quote the interval width.

What this does not establish. The rule holds only for independent draws; a clustered design shrinks less. And more people make the estimate steadier around whatever the sampling method actually measures. If the way people were recruited is biased, 400 or 800 do not remove the bias (Book 0, D6).

9. Compute the mean and SD of one patient and the SD of the average of 30. Then compute the normal range, and the chance that the average is exactly 0.

```
0.01 times 1000 = 10
1000 times 1000 times 0.01 times 0.99 = 9900
square root of 9900 = 99.5
99.5 divided by square root of 30 = 18.17
10 minus 1.96 times 18.17 = -25.6
0.99 to the power 30 = 0.740
```

One patient has mean 10 and SD about 99.5 units. The average of 30 has SD about 18.2. The normal range would run from about minus 25.6, which is impossible, upwards. And three times in four, none of the 30 has an admission, so the average is exactly 0. A distribution with three quarters of its probability on one value is not close to normal.

What this does not establish. The cost model is made up; real costs have their own shape. The point is that "30 is enough" is a rule of thumb for mild skew. For rare, large values the n needed is far larger, and no single n works for every quantity.

19 · Parameter, statistic, estimator, estimate

EXERCISE 1. An estimator is a rule that turns a sample into a number meant to stand for a parameter. An estimate is the number the rule gives on the one sample you drew.

The estimator is a random variable before the sample is drawn, because its result depends on who is picked. The estimate is a fixed number once the sample is in hand.

A "95 per cent" is about the rule: how often it does something over all the samples it could meet. It is never about the one estimate.

EXERCISE 2. 600 is the sample size, the n the rule was fed.

31.5% is an estimate, $\hat{p}$, of the district's parameter p. That parameter is the true share of the district's women aged 15 to 49 with an index of 25 or more.

30.7% is also an estimate. It is NFHS-6's estimate of the national share for the same age group, excluding pregnant women and women with a birth in the two months before.

25 is a definition: the cut-off in kg/m^2 that sorts women into the two groups. It is not estimated from anything.

The last sentence compares two estimates as if they were two parameters. Before accepting it you need each estimate's spread, so you can tell whether the gap is larger than samples of these sizes usually produce. The next rung teaches how. You also need to know that the two used the same definition and comparable women. And "at higher risk" is a further claim. A higher share of women with a high index is not, by itself, a measure of anyone's risk of anything.

1.

```
4 plus 6 plus 7 plus 9 plus 9 = 35
35 divided by 5 = 7
```

$\bar{x} = 7$. Three of the five values are 7 or more.

```
3 divided by 5 = 0.6
```

$\hat{p} = 0.6$.

2. The mean first.

```
5 plus 8 plus 11 = 24
24 divided by 3 = 8
```

The squared deviations, added, divided by n − 1, which is 2.

```
(5 minus 8) squared = 9
(8 minus 8) squared = 0
(11 minus 8) squared = 9
9 plus 0 plus 9 = 18
18 divided by 2 = 9
```

s is the square root of 9, so s = 3.

3. The rule: count the members of the sample who have the feature, and divide by the number of members in the sample.

"The number in the sample with the feature, divided by the sample size" is the same rule, so it is the estimator. "$\hat{p}$ = 0.18" is an estimate: the value the rule gave on one sample.

4.

```
41 plus 5 = 46
46 divided by 151 is about 0.305
```

$\hat{p}$ = 0.305, or 30.5 per cent.

The parameter is the true share with an index of 25 or more in the population these women stand for.

Before calling it an estimate for young Indian women, you would need the women to have been picked at random from that population. They were not. They were medical students at one college who came forward when the study was announced. So 0.305 estimates the share in some population that resembles these volunteers, and the paper cannot say which.

5. 22.1 per cent estimates μ, the population mean body-fat percentage. Its rule: add the 131 values and divide by 131. Written $\bar{x}$ = 22.1 per cent.

7.3 per cent estimates σ, the population SD. Its rule, if the paper divided by n − 1 as Book 0, D5 does (the paper does not say): the sample standard deviation, with n − 1 = 130. Written s = 7.3 per cent.

20.3 per cent estimates the population median. Its rule: sort the 131 values and take the middle one, the 66th. This book gives that estimate no symbol; write "sample median, 20.3 per cent".

All three are estimates. None is the population's value, and the population is whatever these 131 volunteers resemble.

6. The arithmetic is right. Two steps broke.

The second line writes p, the parameter, for a number computed from the sample. It should be $\hat{p}$ = 0.397, an estimate.

The third line then treats that estimate as the parameter, and for a population the sample was never drawn from. The 131 men are volunteers at one college. The sentence should read: "Of 131 male medical students at one college who volunteered, 52, or 39.7 per cent, had an index of 25 or more."

Why it looks reasonable: the symbol p is the one used in every binomial formula, so it is the first letter to hand. Writing it erases the difference between the number you know and the number you want.

7. Compute the true mean, and then check the colleague's range against it.

```
21 plus 23 plus 27 plus 29 = 100
100 divided by 4 = 25
```

μ is 25. The colleague's range, 23 to 27, holds it. So the statement "μ is between 23 and 27" is simply true. Its probability is 1, not 0.625.

The 62.5 per cent is right about the rule: 10 of the 16 equally likely pairs give an average between 23 and 27.

```
10 divided by 16 = 0.625
```

What the last sentence gets wrong: it moves the 62.5 per cent from the rule to the one range. Another colleague might draw 29 and 29. They would say "a 62.5 per cent chance that μ is between 27 and 31". For that range the truth is 0. Both used the same rule and the same reasoning. Only a statement about the rule over all samples can carry the 0.625.

In real work μ is unknown, so you cannot check the range as you can here. That does not make the range's probability 0.625. It means you do not know whether this range is one of the 62.5 per cent.

20 · How good is an estimator: bias and spread, found by listing every sample

EXERCISE 1. The bias is the estimator's expected value minus the parameter. The standard error is the estimator's standard deviation over all the samples it could meet.

The standard error shrinks as the sample grows: for a mean it is $\sigma \div \sqrt{n}$. The bias does not.

Squared deviations measured from the sample's own mean are smaller, on average, than squared deviations from μ. That is why the sample variance divides by $n - 1$. Their sum has expected value $(n - 1)\sigma^2$, so dividing by $n - 1$ gives σ^2 on average.

EXERCISE 2. The sentence runs together random error, the standard error, and systematic error, the bias.

The 2,000 controls the standard error. With that many people, the estimated prevalence will move little from one random sample of attenders to the next.

It does nothing for bias. Outpatient attenders are not the district's adults. They are people who are unwell, or live near the hospital, or can get there. Whatever makes them attend pulls every sample of attenders the same way, and 2,000 of them carry that pull as surely as 200.

To claim accuracy for the district, the protocol needs one of two things. One is a sample drawn at random from the district's adults. The other is evidence that attenders do not differ from the district's adults in waist size. It should then say "precise" for what the sample size buys, and keep "accurate" for what the sampling method buys.

1. There are 3 times 3 = 9 ordered samples.

sample	average
2, 2	2
2, 4	3
2, 6	4
4, 2	3
4, 4	4
4, 6	5

sample	average
6, 2	4
6, 4	5
6, 6	6

```
2 plus 3 plus 4 plus 3 plus 4 plus 5 plus 4 plus 5 plus 6 = 36
36 divided by 9 = 4
2 plus 4 plus 6 = 12
12 divided by 3 = 4
```

The expected value of the average is 4, the same as μ. The average is unbiased.

2.

```
3 plus 7 = 10
10 divided by 2 = 5
(3 minus 5) squared = 4
(7 minus 5) squared = 4
4 plus 4 = 8
8 divided by 2 = 4
8 divided by 1 = 8
```

Dividing by n gives 4. Dividing by n − 1 gives 8.

3.

```
the square root of 36 = 6
12 divided by 6 = 2
the square root of 144 = 12
12 divided by 12 = 1
```

2 for n = 36, and 1 for n = 144. Four times the sample halves the standard error.

4. Use s = 3.4 cm in place of σ.

```
the square root of 131 = 11.45
3.4 divided by 11.45 = 0.297
```

About 0.30 cm. It describes how far the mean of 131 would typically move between random samples of that size from the population the men came from. It does not describe how far these volunteers' mean sits from that of all medical students.

5.

```
the square root of 131 = 11.45
7.3 divided by 11.45 = 0.638
```

About 0.64 percentage points. To halve a standard error, multiply n by 4.

```
4 times 131 = 524
the square root of 524 = 22.89
7.3 divided by 22.89 = 0.319
```

524 men give about 0.32 points, half of 0.638.

6. The larger of the two is 21 only for the pair 21, 21. It is 23 for 21-23, 23-21 and 23-23. It is 27 for the five pairs that hold a 27 and no 29. It is 29 for the seven pairs that hold a 29.

```
1 plus 3 plus 5 plus 7 = 16
21 times 1 plus 23 times 3 plus 27 times 5 plus 29 times 7 = 428
428 divided by 16 = 26.75
26.75 minus 29 = -2.25
```

The rule's expected value is 26.75 kg/m^2 against a largest index of 29. Its bias is minus 2.25. It can never overshoot, so it can only miss low.

7. The second line broke. The standard error of a mean is σ ÷ √n, not σ ÷ n.

```
the square root of 131 = 11.45
3.4 divided by 11.45 = 0.297
```

The estimated standard error is about 0.30 cm, more than eleven times the 0.026.

Why it looks reasonable: "the average of n things is n times more stable" sounds right, and dividing by n is what makes an average. But variances divide by n, so standard deviations divide by √n. The wrong answer makes a study look eleven times more precise than it is.

8. The second line broke. A square root does not pass through an expected value, as Book 2, section 14 warned for squares and ratios. Check it by listing. For a pair a and b, s is the gap between them divided by the square root of 2. The gaps are 0 for four pairs, 2 for four, 4 for two, 6 for four and 8 for two.

```
0 divided by 1.414 = 0
2 divided by 1.414 = 1.414
4 divided by 1.414 = 2.828
6 divided by 1.414 = 4.243
8 divided by 1.414 = 5.657
0 times 4 plus 1.414 times 4 plus 2.828 times 2 plus 4.243 times 4 plus 5.657 times
2 = 39.598
39.598 divided by 16 = 2.475
```

The average of s is about 2.475, against σ of about 3.162. s is biased low.

Why it looks reasonable: the first line is true, and "the square root of an unbiased thing is unbiased" sounds like a rule of arithmetic. It is not one. The small values of s, including the four zeros, pull its average down more than the square root lets the large ones pull it up.

9. Compute each share's standard error. For a share, section 18 gives the SD as the square root of p(1 − p) ÷ n. Use each survey's own p̂ in place of p.

```
0.35 times 0.65 = 0.2275
0.2275 divided by 10000 = 0.00002275
the square root of 0.00002275 = 0.0048
0.28 times 0.72 = 0.2016
0.2016 divided by 400 = 0.000504
the square root of 0.000504 = 0.0224
```

The mall survey's standard error is about 0.5 percentage points. The household survey's is about 2.2 points. So the mall figure is far more precise.

What this does not establish: that it is more accurate. The two estimates differ by 7 points, more than three of the household survey's standard errors. That gap is more than random sampling usually produces. The mall survey is the one with a reason to be off, because women at malls are not the city's women. Its standard error says nothing about that bias. The household figure is less precise and has no such built-in pull. Neither number shows how large the mall's bias is; only a comparison like this one hints that it exists.

21 · The likelihood: the same formula read as a function of the parameter

EXERCISE 1. The likelihood of a value of p is the probability of the observed data if the parameter were that value. The data are held fixed and p is varied.

Each is a probability of the same data under a different p. None is the probability of a p, and nothing in the model makes p a random variable. So they need not add to 1.

The ratio L(p1) ÷ L(p2) says how many times more probable the data are under p1 than under p2. It is the likelihood ratio of the odds form of Bayes' theorem. The two values of p are its two hypotheses.

EXERCISE 2. The ratio of the maximum to the likelihood at 0.273 is $(0.397 \div 0.273)^{52} \times (0.603 \div 0.727)^{79}$.

```
0.397 divided by 0.273 = 1.45421
0.603 divided by 0.727 = 0.82944
log10 1.45421 = 0.16263
log10 0.82944 = -0.08122
52 times 0.16263 = 8.4568
79 times -0.08122 = -6.4164
8.4568 plus -6.4164 = 2.0404
```

$10^{2.0404}$ is about 110. So the likelihood at 0.273 is about 1 in 110 of the maximum, and the sentence's "less than 1 in 100" is right.

What it shows: these 131 men's counts are far more probable under a share near 0.4 than under 0.273.

What "excluded" overstates. A likelihood ratio does not exclude a value; turning it into a decision needs a rule the next rung teaches. And the comparison is of different populations. The 0.273 is the 27.3 per cent of NFHS-6, 2023-24, for men aged 15 to 49 across India. The 0.397 is for male medical students who volunteered at one college. Nothing here says the national figure is wrong for the nation.

1. L(p) = (3 choose 2) × p^2 × (1 − p). 3 choose 2 is 3.

```
3 times 0.25 times 0.5 = 0.375
3 times 0.36 times 0.4 = 0.432
```

L(0.5) = 0.375 and L(0.6) = 0.432. The data support 0.6 better.

2. The factor 10 choose 6 is in both likelihoods, so it cancels.

```
0.6 divided by 0.4 = 1.5
0.4 divided by 0.6 = 0.6667
1.5^6 = 11.390625
0.6667^4 = 0.19757
11.390625 times 0.19757 = 2.25
```

The data are 2.25 times as probable if p is 0.6 as if p is 0.4.

3. $L(p) = 4 \times p \times (1 - p)^3$.

```
4 times 0.2 times 0.512 = 0.4096
4 times 0.4 times 0.216 = 0.3456
4 times 0.6 times 0.064 = 0.1536
4 times 0.8 times 0.008 = 0.0256
```

Of the four, 0.2 is best supported. The sample share is $1 \div 4 = 0.25$.

```
4 times 0.25 times 0.421875 = 0.421875
```

L(0.25) is about 0.4219, higher than any of the four.

4.

```
46 divided by 151 is about 0.305
151 minus 46 = 105
1 minus 0.305 = 0.695
1 minus 0.307 = 0.693
0.305 divided by 0.307 = 0.993485
0.695 divided by 0.693 = 1.002886
46 times -0.0028387 = -0.13058
105 times 0.0012516 = 0.13142
-0.13058 plus 0.13142 = 0.00084
```

Here −0.0028387 is $\log_{10}$ of 0.993485, and 0.0012516 is $\log_{10}$ of 1.002886. The ratio is $10^{0.00084}$, about 1.002.

The best-supported share is 0.305. The counts are almost exactly as probable at 0.307.

What it shows: these 151 women's counts fit 0.305 and 0.307 about equally. What it does not show: that these women stand for young Indian women. They are volunteer medical students at one college. A ratio near 1 says the data do not separate two values; it does not say the two groups are the same.

5. The first line is right. The second line broke. L(0.3) is the probability of 3 in 10 if p were 0.3. It is not the probability that p is 0.3.

The third line inherits the error. Add the likelihoods at 0.1, 0.2, up to 0.9 and they come to about 0.91. Add more values of p between them and the total grows past 1. They are not a distribution for p, so 1 minus one of them means nothing.

Why it looks reasonable: the number came from a probability formula, and it is a probability. It is a probability of the data. Reading it as "the probability of p" swaps the two sides of the bar in $P(X = 3 \mid p)$. That is the transposed conditional of section 9.

6. Compute each value's likelihood against the best-supported share, $52 \div 131 = 0.397$.

```
0.3 divided by 0.397 = 0.75567
0.7 divided by 0.603 = 1.16086
52 times -0.12167 = -6.3268
79 times 0.06478 = 5.1176
-6.3268 plus 5.1176 = -1.2092
0.5 divided by 0.397 = 1.25945
0.5 divided by 0.603 = 0.82919
52 times 0.10018 = 5.2094
79 times -0.08134 = -6.4259
5.2094 plus -6.4259 = -1.2165
```

$10^{-1.2092}$ is about 0.062, and $10^{-1.2165}$ is about 0.061. Each value has about one sixteenth of the best-supported value's likelihood.

What the colleague gets right: the ratio of 0.30 to 0.50 is about 1.02. So the data do not favour either of those two over the other.

What the claim gets wrong: the data are not uninformative. Both values sit on opposite sides of the peak at 0.397, about equally far down. The data favour a share near 0.4 over either by about 16 to 1. A ratio of two values tells you only about those two.

What none of this establishes: the share among young Indian men. The likelihood is built on volunteers at one college, and treats their results as independent.

22 · Parameter, estimate or prediction: what a reported number is a number of

EXERCISE 1.

1. A count of everyone: conditional on the group and period counted. Ask who was missed and how each person was measured.
2. A survey estimate: conditional on the sample and the survey's design. Ask which population, in which years, from what sample.
3. A model-based estimate: conditional on the data that went in and on the model. Ask which data went in and what the model filled in.
4. A prediction: a forecast about a future population, or a risk for one person. It is conditional on a scenario, or on the group the score was built on. Ask under what scenario, and for people like whom.

EXERCISE 2.

1. A count of everyone, but only of the children weighed at the camp this year. It is the parameter of that group. It is not the district's prevalence: children absent on camp day, and children not in school, are missing. It carries the scale's and the tape's errors.
2. A survey estimate of a parameter: the share of all men aged 15 to 49 in India with a body mass index of 25 or more, in 2023-24. Conditional on NFHS-6's sample and design.

3. A prediction, a forecast. GBD 2021 gives it under a reference scenario in which past trends go on. “Will have” drops the scenario. It counts adults aged 25 and over, and overweight and obesity together.
4. A prediction for one person, built from a group share. It is the positive predictive value of IDRS 60 or more in the Chennai study, at its prior of about 10 per cent. It holds for this woman only as far as your clinic’s patients resemble that study’s.
5. A model-based estimate of a 2021 parameter, for adults aged 25 and over. The interval is the range of the model’s draws, conditional on the model.

EXERCISE 3. Four things.

- “Obese”: the 450 million counts overweight and obesity together, a body mass index of 25 or more. The paper defines obese as 30 or more.
- “Adults”: the paper counts adults aged 25 years and older, not all adults.
- “Will have”: the figure is a forecast under a reference scenario, in which past trends and patterns continue. The paper says it has not built in wide use of new obesity medicines or other rapid advances in prevention or treatment. It is what would happen if nothing changes the trend.
- No range: the paper gives 359 to 535 million, and even that is the model’s spread under one scenario. A single number with no range claims more certainty than the authors do.

A fair headline: “If past trends continue, about 450 million Indians aged 25 and over could have overweight or obesity by 2050, a Lancet model forecasts.”

EXERCISE 4. Tell her both can be right, because they count different people in different ways. The national survey weighed and measured women aged 15 to 49 in 2023-24. The Lancet number is a model’s estimate for all adults aged 25 and over in 2021, built from many surveys, including some where people reported their own weight.

The sentence to quote: “The survey found about three women in ten aged 15 to 49 with overweight or obesity. The Lancet model estimates 180 million Indian adults aged 25 and over. They are not measuring the same thing.”

How sure: the survey’s figure is a direct measurement of a large sample. The model’s figure is a careful estimate whose range, 167 to 194 million, shows the model’s own spread. It cannot show what the model got wrong. Neither says anything about a single state or district.

23 · One clinic test, from the question to a handout a patient can read

EXERCISE 1.

1. The question: one condition, people like this patient, this clinic.
2. The test’s figures, read with their condition, cut-off and sample. They are estimates (section 19, section 22).
3. The prior, from your own register, through the four checks of section 10.
4. The calculation three ways: frequency tree (section 7), formula (section 5), odds and likelihood ratios (section 8). They must agree.
5. The prior moved: the other priors on offer, and why each does or does not pass the checks.
6. The colleague’s transposed sentence, corrected in four steps (section 9).

7. The checks: cells add up, routes agree, digits no finer than the inputs.
8. The handout, and what it does not establish.

EXERCISE 2. Two faults, one per sentence.

- "72.5% chance of having diabetes" is the transposed conditional. 72.5 per cent is the sensitivity, P(score 60 or more | undiagnosed diabetes). The patient's question is P(undiagnosed diabetes | score 60 or more), the PPV. In Mohan and colleagues' study, at its prior of about 10 per cent, that was 17.0 per cent.
- "The national prevalence" fails the checks of section 10. The score was tested against diabetes not yet diagnosed, by a 2-hour glucose test, in people not known to have diabetes. A national figure that includes known cases counts a different condition in different people. The prior must come from a register of people screened at your clinic who were not already known to have diabetes.

Count your own prior first. Then the line should read: "Of 100 people like you with a score of 60 or more, about [your PPV] have diabetes that has not been found yet. A blood test will tell."

EXERCISE 3. Draw the 1,000-man frequency tree. Start with 1,000 young men at the clinic, 200 in the band and 800 not. Of the 200, the tape catches 172: that is the 86 per cent, and it is out of men in the band. Of the 800, it wrongly flags 208, because its specificity is 74 per cent.

Now circle the men with a high reading: 172 plus 208, 380 men. Only 172 of them are in the band, about 45 in 100. The 86 per cent and the 45 per cent share the same top count, 172. They divide it by different totals, the 200 in the band and the 380 flagged.

Then rub out the 200 and write 100, for a clinic where one man in ten is in the band. The tape catches 86 of the 100 and wrongly flags 234 of the other 900.

```
0.86 times 100 = 86
0.26 times 900 = 234
86 plus 234 = 320
86 divided by 320 ≈ 0.269
```

About 27 in 100. The PPV fell with the prior while the 86 per cent stayed. Leave them with the question to ask of any percentage: out of whom?

EXERCISE 4. Check your handout against this list.

- The condition is the same in the question, the paper's figures and the register. If the paper's label and its numbers disagree, as Kiran's do, the handout names the one the numbers fit.
- The register's prior names its count, its denominator, its period and who was left out. Every person in the denominator was tested.
- The three routes agree to the rounding. The patient's page carries whole numbers out of 1,000 or 100, never decimals.
- The national and study priors are shown on the staff copy, with the check each fails.
- The correction of the transposed sentence restates both conditionals in words, asks "out of whom?", counts, and names the number that answers the question.

- The “does not establish” box names the sample the test was measured in, and how the cut-off was chosen. It says that all inputs are estimates. It says that nothing is claimed about whether the test or the condition matters for the patient’s health.
- Two patients read it, and the page was changed for anything either one misread.

If your clinic has no register in which everyone was tested by the reference test, the handout cannot be finished honestly. Say so on the page and start the count.

Glossary

Every term of art this book teaches, in the plain words it gets where it is first used, and the section that teaches it.

Term	Plain words	Section
axioms of probability	the three rules every probability obeys: never negative, the whole sample space has probability 1, and mutually exclusive events add	1
base rate	(1) the prevalence of a condition in the group being tested, before any test: the prior probability (Book 12); (2) how often claims of one kind have turned out true when they were tested; the evidence particular to a new claim of that kind then moves you from there (Book 10)	6
Bayes' theorem	P(A given B) = P(B given A) × P(A) ÷ P(B): a conditional probability turned round, from the two ways of writing P(A and B)	5
Bernoulli random variable	1 for a success with probability p, 0 for a failure	12
bias (of an estimator)	the estimator's expected value minus the parameter; the systematic error of a rule, which a bigger sample does not shrink	20
binomial distribution	the count of successes in a fixed number of independent trials with the same p	12
central limit theorem	a sum of many small independent pieces, none dominating, has a distribution close to normal; so, as n grows, does the sum or average of n independent draws, whatever the shape of one draw	15, 18
chance process	anything whose result is not known until it runs	1
conditionally independent (given C)	independent among the outcomes in C: P(A given B and C) = P(A given C)	3
correlation	1. a number worked out from paired measurements (Book 9); 2. the covariance divided by both standard deviations: it has no unit, lies between −1 and 1, and measures only a straight-line tie (Book 12)	17
count of everyone	every member of a stated group counted, so the number is that group's parameter, with no sampling error but with the errors of counting and measuring	22
counting rule	the number of ways to do a task in stages is the product of the numbers of ways for each stage, when the choices at a stage do not depend on earlier ones	11
covariance, Cov(X, Y)	the expected value of the product of the two variables' deviations from their own expected values; how they move together	17

Term	Plain words	Section
degree of belief	(1) a person's own probability for an event, shown by the bets they would accept (Book 12); (2) a number from 0 to 1 that states how strongly one person holds one claim to be true, given what that person knows; it obeys the same arithmetic as a probability (Book 10)	1
dependent	not independent: knowing one changes the chance of the other	3
estimator	a rule that turns a sample into a number meant to stand for a parameter; before the sample is drawn its result is a random variable	19
factorial, n!	the product of the whole numbers from n down to 1; 0! is 1	11
false-positive rate	the share of people without the condition whom the test still flags, P(test positive given no condition): 1 minus the specificity	4
forecast	a prediction about a population in the future, conditional on a scenario for how the future goes	22
frequency tree	a diagram that follows one group of people through each split in turn (condition, then test) and writes the count at each end	7
general addition rule	P(A or B) = P(A) + P(B) − P(A and B), for any two events	1
general multiplication rule	P(A and B) = P(B) × P(A given B), for any two events, independent or not	2
geometric distribution	the number of independent trials up to and including the first success	13
geometric mean	the mean of the logarithms of the values, turned back with the same base; the n-th root of their product	16
hypergeometric distribution	the number with a feature among n drawn without replacement from a group of known make-up	13
joint distribution	every pair of values two random variables can take together, with the probability of each pair	17
joint probability	a probability about two events together: one cell of a two-way table	2
law of total probability	the probability of an event is the sum, over groups that do not overlap and take in everyone, of its probability in each group times the group's share	4
likelihood ratio (LR)	how much more probable the evidence is if the hypothesis is true than if it is false: P(evidence given true) divided by P(evidence given false); it multiplies the odds	8
likelihood, L(p)	the probability of the observed data if the parameter were p, with the data held fixed and p varied; not the probability that p is true	21
log-normal	a positive quantity whose natural logarithm has a normal distribution; made by many small independent effects that multiply	16
marginal probability	a probability about one event alone: a row or column total	2

Term	Plain words	Section
maximum likelihood estimate	the value of the parameter with the largest likelihood; for a binomial share, k ÷ n	21
model-based estimate	a number for the present or past that a model fills in from many sources of data and its own assumptions; nobody measured it	22
multiplicative standard deviation	e raised to the standard deviation of ln X; a factor, always above 1, that the median is multiplied or divided by	16
n choose k	the number of ways to choose k things from n when order does not matter, n! ÷ [k! (n − k)!]	11
natural frequencies	(1) counts of people got by following one group through each split in turn, not divided by any group's size, so they still carry the base rate (Book 12); (2) whole counts of imagined cases, out of 1,000, or out of 100,000 when the base rate is rare, so every count stays whole (Book 10)	7
negative likelihood ratio (LR−)	(1 minus sensitivity) divided by specificity: what a negative result multiplies the odds by	8
normal distribution, N(μ, σ)	the symmetric bell-shaped continuous distribution fixed by its mean μ and standard deviation σ	15
normal probability plot	the sorted data plotted against the z-scores a normal distribution would give them; near-normal data fall near a straight line	15
overdispersion (overdispersed)	counts whose variance is well above their mean: the events cluster, or the rate changes	14
Poisson distribution	the distribution of a count of events at a steady rate, independent of each other; mean and variance both λ	14
positive likelihood ratio (LR+)	sensitivity divided by (1 minus specificity): what a positive result multiplies the odds by	8
posterior probability	the probability once you know the evidence; for a test, the PPV after a positive result	5
pre-test probability	another name for the prior probability: the chance a person like this patient, in this setting, has the condition before the test	10
prediction	(1) a number about something not yet observed: a future value, or one person (Book 12); (2) a statement of the form "if the hypothesis is true, then this observable result will follow", fixed before the result is seen (Book 10)	22
prior odds / posterior odds	the odds before and after the evidence; posterior odds = prior odds × likelihood ratio	8
prior probability	the probability before the evidence; for a test, the prevalence in the group tested	5
prosecutor's fallacy	reading the chance of the evidence if the accused is innocent as the chance the accused is innocent given the evidence	9
q-q plot	strictly, one data set's quantiles plotted against another's; against a normal distribution it is the normal probability plot	15
random match probability (RMP)	how often a DNA profile like this one turns up among people who are not its source; P(match given not the source)	9

Term	Plain words	Section
reference scenario (GBD)	GBD's forecast setting in which past trends and patterns continue	22
sampling distribution of the mean	the distribution of the average over all the samples of size n that could be drawn	18
simulation by hand	draw, record, repeat, tabulate, with dice or a random-number table	18
standard error (of an estimator, SE)	an estimator's standard deviation over every sample it could meet; the random error of a rule, which shrinks as n grows	20
standard normal distribution	the normal distribution with mean 0 and standard deviation 1, N(0, 1)	15
success (in a trial)	the kind of result being counted; a label, not praise	11, 12
transposed conditional	reading P(A given B) as P(B given A): the two conditions swapped, so the bottom of the fraction changes	9
tree diagram	the multiplication rule drawn in stages: marginal probabilities on the first split, conditional ones on the second, joint probabilities at the ends	2
trial (in probability)	one run of a chance process whose result is sorted into two kinds	11, 12
unbiased	an estimator whose bias is 0: averaged over every sample it could meet, it lands on the parameter	20
uncertainty interval (UI)	the range holding the middle 95 per cent of a model's draws; it shows the model's own spread, conditional on the model	22
Venn diagram	a box for the sample space, with a circle for each event; the overlap is "A and B"	1
weighted average	each value times its weight, added, where the weights are not negative and add to 1	4
with replacement / without replacement	each member drawn is put back and can be drawn again / can be drawn only once	3
Youden index	sensitivity plus specificity minus 1	23
z-score	a value's distance from the mean in standard deviations, $(x - \mu) \div \sigma$	15

The series: all 196 books

In build order: Book 0 first, then every Introductory rung, then every Intermediate rung, and so on; within a level, a subject comes after the subjects it depends on. "done" means that book's version 1 is finished. This is Book 12.

GROUND FLOOR		
0	Ground floor	done
INTRODUCTORY		
1	Energy balance and dynamic weight-change modelling · Rung 1	done
2	Mathematics for modelling and inference · Rung 1	done
3	Qualitative and mixed methods · Rung 1	done
4	Food systems, value chains and Indian food policy · Rung 1	done
5	Policy process, framing and advocacy craft · Rung 1	done
6	R, reproducibility engineering and data management · Rung 1	done
7	Question selection and research strategy · Rung 1	done
8	Teaching, mentorship and coalition building · Rung 1	done
9	Scientific writing, visualisation and public communication · Rung 1	done
10	Epistemics: philosophy of science and calibration · Rung 1	done
11	Adjacent specialist domains: recognise and route · Rung 1	
12	**Probability and statistical inference · Rung 1 ◀ you are here**	done
13	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 1	
14	Commercial determinants and corporate political activity · Rung 1	
15	System dynamics and complexity concepts · Rung 1	
16	Indian regulatory architecture and the live policy agenda · Rung 1	
17	Comparative and international policy · Rung 1	
18	Python for simulation, scraping and tooling · Rung 1	
19	Regression modelling for health outcomes · Rung 1	
20	Potential outcomes, identification and target trial emulation · Rung 1	
21	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 1	
22	Appetite, reward and the determinants of intake · Rung 1	
23	Nutrient science for obesity and cardiometabolic risk · Rung 1	
24	Physical activity and sedentary behaviour science · Rung 1	
25	Sleep science and obstructive sleep apnoea · Rung 1	
26	Participatory systems mapping and group model building · Rung 1	
27	Agent-based and network modelling · Rung 1	
28	Clustered, longitudinal, hierarchical and small-area models · Rung 1	
29	Evidence synthesis and meta-analysis · Rung 1	
30	Prediction modelling and machine learning · Rung 1	
31	Measurement error theory and correction · Rung 1	
32	Causal diagrams, structural bias and sensitivity analysis · Rung 1	
33	Study and trial design, power and missing data · Rung 1	
34	Clinical assessment, staging and secondary causes · Rung 1	
35	Indian dietary patterns, guidelines, food composition and diet cost · Rung 1	
36	Circadian biology and chrononutrition · Rung 1	

37	Behavioural theory, choice architecture and intervention design · Rung 1
38	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 1
39	G-methods, time-varying treatment and mediation · Rung 1
40	Instrumental variables and Mendelian randomisation · Rung 1
41	Quasi-experimental policy evaluation · Rung 1
42	Genetics, developmental origins and life-course epidemiology · Rung 1
43	Social epidemiology, equity measurement and surveillance in India · Rung 1
44	Nutritional epidemiology and its pathologies · Rung 1
45	Lifestyle, dietary and behavioural treatment · Rung 1
46	Pharmacotherapy of obesity · Rung 1
47	Metabolic and bariatric surgery: selection and co-management · Rung 1
48	Eating behaviour, mental health and weight stigma · Rung 1
49	Food environment measurement, retail and digital delivery · Rung 1
50	Food demand, elasticity and the economics of externality · Rung 1
51	Implementation science and programme craft · Rung 1
52	Protocol, preregistration, reporting standards and research ethics · Rung 1
53	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 1
54	Ultra-processed food and the effects of processing · Rung 1
55	Activity measurement and compositional 24-hour analysis · Rung 1
56	Fiscal instruments: taxation and subsidy design in India · Rung 1
57	Burden of disease, economic evaluation and decision-analytic modelling · Rung 1
58	The Indian health system and its financing mechanics · Rung 1
59	India's data infrastructure, access and complex survey analysis · Rung 1
60	Ethics of obesity intervention and industry engagement · Rung 1
61	Health financing, HTA and pharmaceutical access in India · Rung 1
INTERMEDIATE	
62	Energy balance and dynamic weight-change modelling · Rung 2
63	Mathematics for modelling and inference · Rung 2
64	Qualitative and mixed methods · Rung 2
65	Food systems, value chains and Indian food policy · Rung 2
66	Policy process, framing and advocacy craft · Rung 2
67	R, reproducibility engineering and data management · Rung 2
68	Question selection and research strategy · Rung 2
69	Teaching, mentorship and coalition building · Rung 2
70	Scientific writing, visualisation and public communication · Rung 2
71	Epistemics: philosophy of science and calibration · Rung 2
72	Probability and statistical inference · Rung 2
73	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 2
74	Commercial determinants and corporate political activity · Rung 2
75	System dynamics and complexity concepts · Rung 2
76	Indian regulatory architecture and the live policy agenda · Rung 2
77	Comparative and international policy · Rung 2
78	Python for simulation, scraping and tooling · Rung 2
79	Regression modelling for health outcomes · Rung 2
80	Potential outcomes, identification and target trial emulation · Rung 2
81	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 2

82	Appetite, reward and the determinants of intake · Rung 2
83	Nutrient science for obesity and cardiometabolic risk · Rung 2
84	Physical activity and sedentary behaviour science · Rung 2
85	Sleep science and obstructive sleep apnoea · Rung 2
86	Participatory systems mapping and group model building · Rung 2
87	Agent-based and network modelling · Rung 2
88	Clustered, longitudinal, hierarchical and small-area models · Rung 2
89	Evidence synthesis and meta-analysis · Rung 2
90	Prediction modelling and machine learning · Rung 2
91	Measurement error theory and correction · Rung 2
92	Causal diagrams, structural bias and sensitivity analysis · Rung 2
93	Study and trial design, power and missing data · Rung 2
94	Clinical assessment, staging and secondary causes · Rung 2
95	Indian dietary patterns, guidelines, food composition and diet cost · Rung 2
96	Circadian biology and chrononutrition · Rung 2
97	Behavioural theory, choice architecture and intervention design · Rung 2
98	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 2
99	G-methods, time-varying treatment and mediation · Rung 2
100	Instrumental variables and Mendelian randomisation · Rung 2
101	Quasi-experimental policy evaluation · Rung 2
102	Genetics, developmental origins and life-course epidemiology · Rung 2
103	Social epidemiology, equity measurement and surveillance in India · Rung 2
104	Nutritional epidemiology and its pathologies · Rung 2
105	Lifestyle, dietary and behavioural treatment · Rung 2
106	Pharmacotherapy of obesity · Rung 2
107	Metabolic and bariatric surgery: selection and co-management · Rung 2
108	Eating behaviour, mental health and weight stigma · Rung 2
109	Food environment measurement, retail and digital delivery · Rung 2
110	Food demand, elasticity and the economics of externality · Rung 2
111	Implementation science and programme craft · Rung 2
112	Protocol, preregistration, reporting standards and research ethics · Rung 2
113	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 2
114	Ultra-processed food and the effects of processing · Rung 2
115	Activity measurement and compositional 24-hour analysis · Rung 2
116	Fiscal instruments: taxation and subsidy design in India · Rung 2
117	Burden of disease, economic evaluation and decision-analytic modelling · Rung 2
118	The Indian health system and its financing mechanics · Rung 2
119	India's data infrastructure, access and complex survey analysis · Rung 2
120	Ethics of obesity intervention and industry engagement · Rung 2
121	Health financing, HTA and pharmaceutical access in India · Rung 2
ADVANCED	
122	Energy balance and dynamic weight-change modelling · Rung 3
123	Mathematics for modelling and inference · Rung 3
124	Qualitative and mixed methods · Rung 3
125	Food systems, value chains and Indian food policy · Rung 3
126	Policy process, framing and advocacy craft · Rung 3

127	R, reproducibility engineering and data management · Rung 3
128	Question selection and research strategy · Rung 3
129	Teaching, mentorship and coalition building · Rung 3
130	Scientific writing, visualisation and public communication · Rung 3
131	Epistemics: philosophy of science and calibration · Rung 3
132	Probability and statistical inference · Rung 3
133	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 3
134	Commercial determinants and corporate political activity · Rung 3
135	System dynamics and complexity concepts · Rung 3
136	Indian regulatory architecture and the live policy agenda · Rung 3
137	Regression modelling for health outcomes · Rung 3
138	Potential outcomes, identification and target trial emulation · Rung 3
139	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 3
140	Appetite, reward and the determinants of intake · Rung 3
141	Nutrient science for obesity and cardiometabolic risk · Rung 3
142	Physical activity and sedentary behaviour science · Rung 3
143	Sleep science and obstructive sleep apnoea · Rung 3
144	Participatory systems mapping and group model building · Rung 3
145	Clustered, longitudinal, hierarchical and small-area models · Rung 3
146	Evidence synthesis and meta-analysis · Rung 3
147	Measurement error theory and correction · Rung 3
148	Causal diagrams, structural bias and sensitivity analysis · Rung 3
149	Study and trial design, power and missing data · Rung 3
150	Clinical assessment, staging and secondary causes · Rung 3
151	Indian dietary patterns, guidelines, food composition and diet cost · Rung 3
152	Circadian biology and chrononutrition · Rung 3
153	Behavioural theory, choice architecture and intervention design · Rung 3
154	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 3
155	G-methods, time-varying treatment and mediation · Rung 3
156	Instrumental variables and Mendelian randomisation · Rung 3
157	Quasi-experimental policy evaluation · Rung 3
158	Genetics, developmental origins and life-course epidemiology · Rung 3
159	Social epidemiology, equity measurement and surveillance in India · Rung 3
160	Nutritional epidemiology and its pathologies · Rung 3
161	Lifestyle, dietary and behavioural treatment · Rung 3
162	Pharmacotherapy of obesity · Rung 3
163	Metabolic and bariatric surgery: selection and co-management · Rung 3
164	Eating behaviour, mental health and weight stigma · Rung 3
165	Food environment measurement, retail and digital delivery · Rung 3
166	Food demand, elasticity and the economics of externality · Rung 3
167	Implementation science and programme craft · Rung 3
168	Protocol, preregistration, reporting standards and research ethics · Rung 3
169	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 3
170	Ultra-processed food and the effects of processing · Rung 3
171	Activity measurement and compositional 24-hour analysis · Rung 3
172	Fiscal instruments: taxation and subsidy design in India · Rung 3
173	Burden of disease, economic evaluation and decision-analytic modelling · Rung 3

174	The Indian health system and its financing mechanics · Rung 3
175	India's data infrastructure, access and complex survey analysis · Rung 3
176	Ethics of obesity intervention and industry engagement · Rung 3

EXPERT

177	Energy balance and dynamic weight-change modelling · Rung 4
178	R, reproducibility engineering and data management · Rung 4
179	Question selection and research strategy · Rung 4
180	Probability and statistical inference · Rung 4
181	Indian regulatory architecture and the live policy agenda · Rung 4
182	Regression modelling for health outcomes · Rung 4
183	Potential outcomes, identification and target trial emulation · Rung 4
184	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 4
185	Appetite, reward and the determinants of intake · Rung 4
186	Measurement error theory and correction · Rung 4
187	Causal diagrams, structural bias and sensitivity analysis · Rung 4
188	Circadian biology and chrononutrition · Rung 4
189	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 4
190	Genetics, developmental origins and life-course epidemiology · Rung 4
191	Nutritional epidemiology and its pathologies · Rung 4
192	Eating behaviour, mental health and weight stigma · Rung 4
193	Activity measurement and compositional 24-hour analysis · Rung 4
194	India's data infrastructure, access and complex survey analysis · Rung 4
195	Ethics of obesity intervention and industry engagement · Rung 4

OBESITY EXPERTISE · BOOK 12

Probability and statistical inference · Rung 1

A short book on probability for people who order tests and read research. It covers the rules every probability obeys, conditional probability and independence, Bayes' theorem worked first by counting people and then by formula, and why most clinicians misjudge a positive result. It teaches the binomial, Poisson and normal patterns and what produces each, covariance, and what happens when you average many draws. It ends by separating a parameter from an estimate and a prediction, and with a handout on a real test, for your own clinic.

Book 12 of the series' 196 books, free to read at drharshmaheshwari.com.



Dr. Harsh Maheshwari
MD Community Medicine Resident, AIIMS Raipur
contact@drharshmaheshwari.com
ORCID 0000-0001-8848-1918
drharshmaheshwari.com

Version 1.0 · 13 October 2026

CC BY-NC-SA 4.0 · free to share, not to sell