Probability

Topic 5 • Tuesday, January 19, and Thursday, January 21

EDUC 15: Statistics for Education Research
UCI School of Education

Where are we?

Description

Summarize data

Visualize data

We are here, heading right

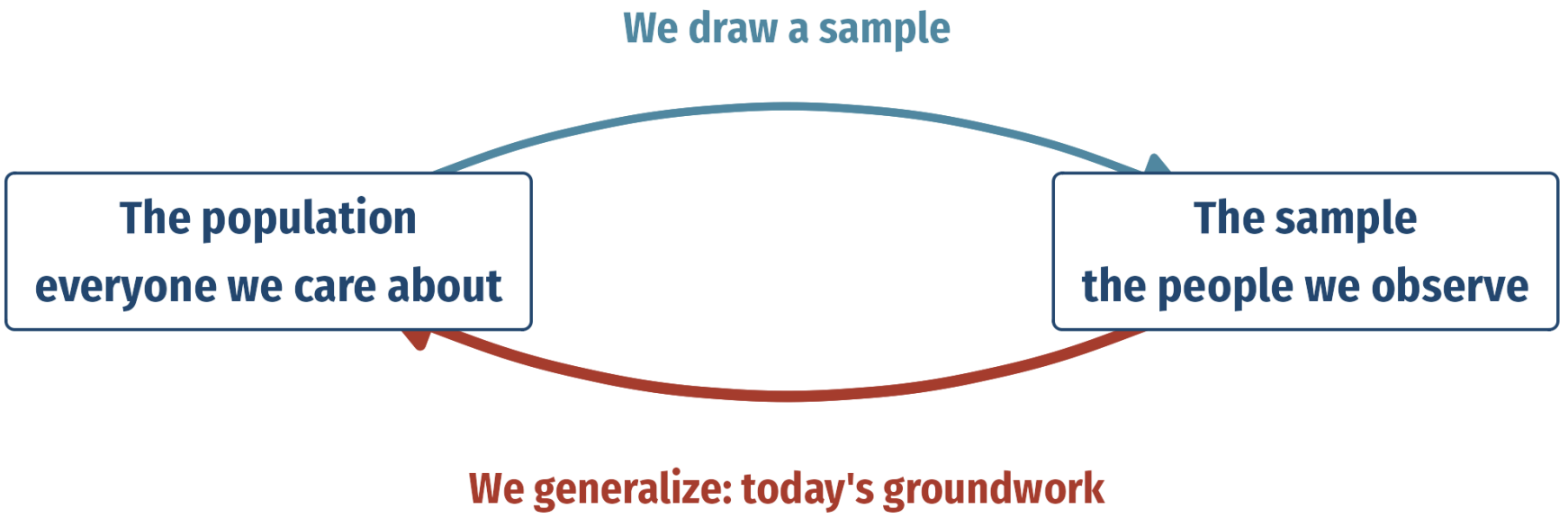


Prediction

Explanation

Peter the Anteater at a UC Irvine basketball game.

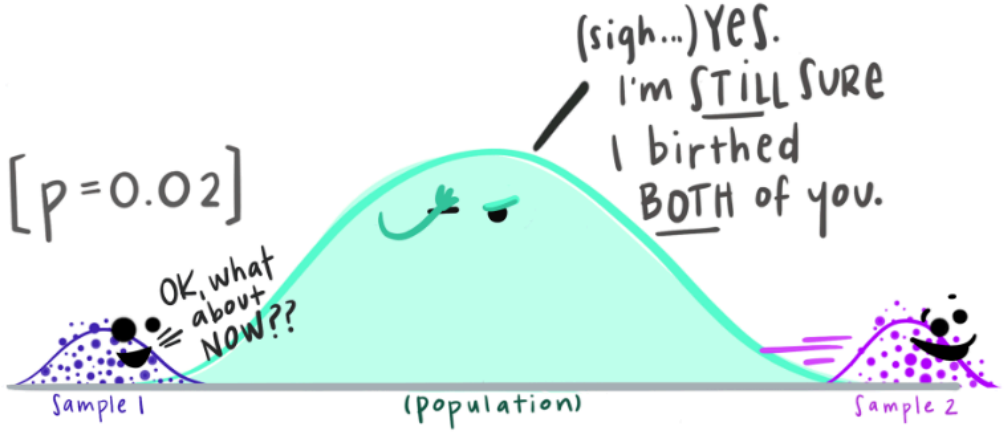
Where are we? From sample to everyone



A sample is never the whole story. Probability says how much it can fool us

After a figure in Gravetter et al., *Essentials of Statistics for the Behavioral Sciences*.

The question we'll keep asking



If what I think about the population is true, how weird would a sample like mine be?

And how weird must it be before I decide I was wrong?

Artwork by Allison Horst, allisonhorst.com.

Plan for today

What is probability?

From bars to curves

The normal distribution

What is probability?

Probability is a share of outcomes

$$p(A) = \frac{\text{outcomes that count as } A}{\text{all possible outcomes}}$$

Roll a fair die: p(6) = 1 / 6 = 0.167

A share always lies between 0 (never) and 1 (always)

0.25, 1 in 4 and 25% all say the same thing

This only works if we draw at random

Random

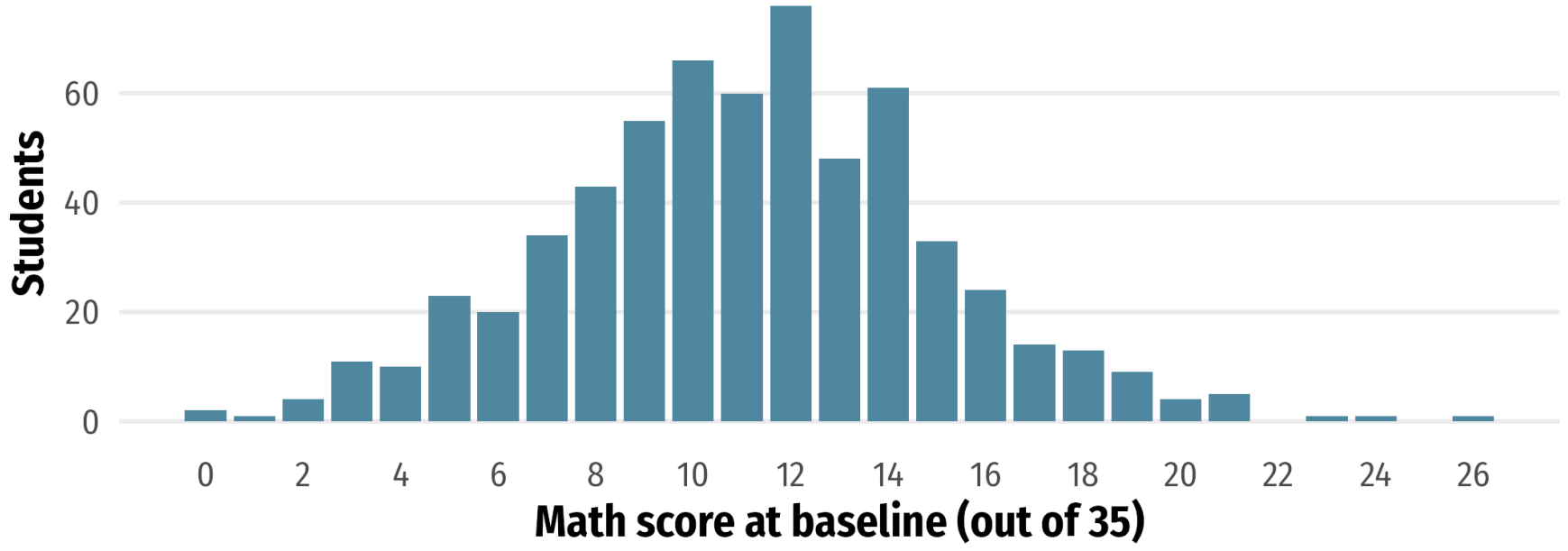
Everyone has the same chance of being picked

Independent

One draw doesn't change the chances for the next

Together these are called *independent and identically distributed*, or i.i.d. Every formula in this course leans on it

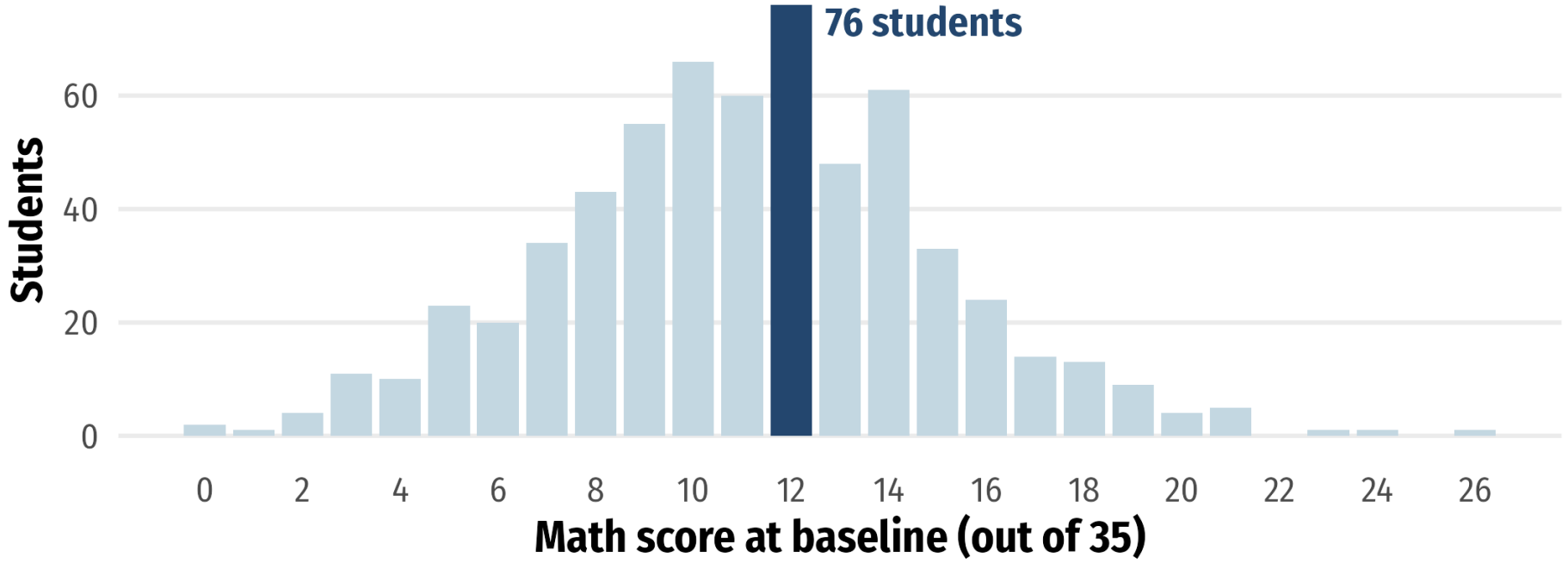
Pretend these 619 are everyone



They're really a sample from Delhi. For today they're everyone, so we write $\mu = 11.12$ and $\sigma = 3.90$

Data: Muralidharan, Singh, and Ganimian (2019). 619 students in Delhi.

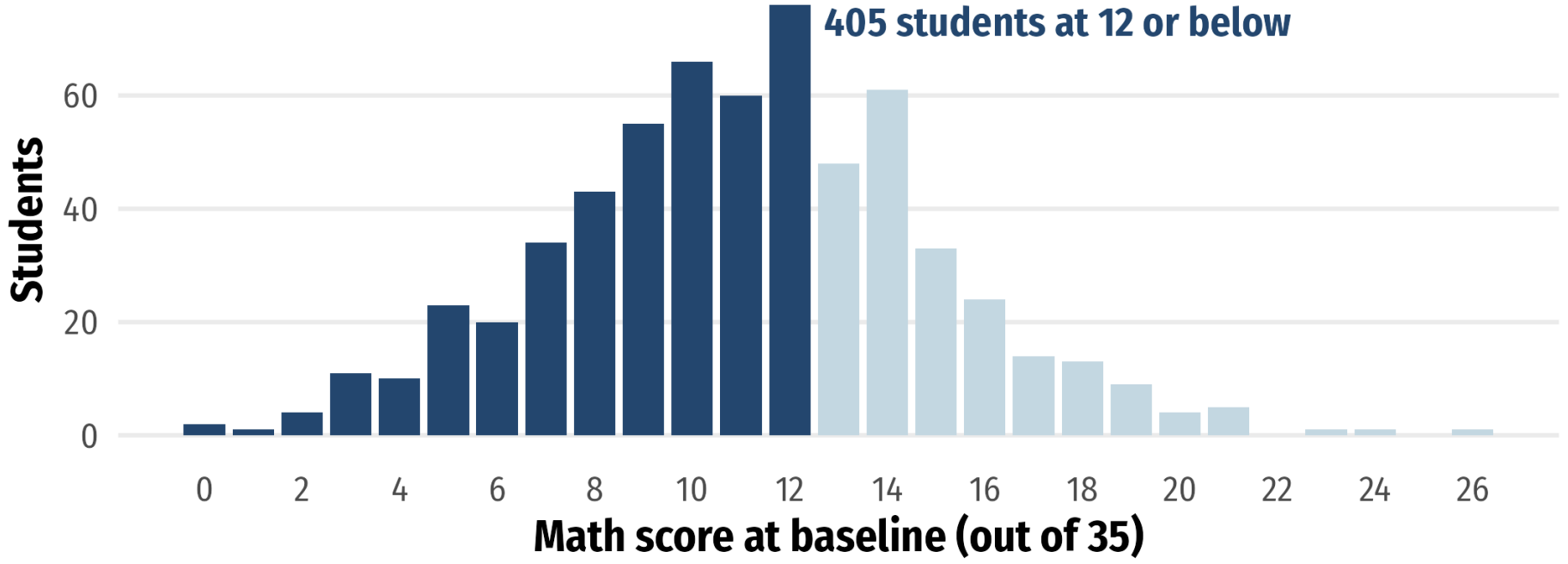
Pick one student: what's p(12)?



$$p(X = 12) = 76 \ / \ 619 = 0.123$$

Data: Muralidharan, Singh, and Ganimian (2019).

A range of scores just adds up bars



$$p(X \leq 12) = (2 + 1 + 4 + \ldots + 76) \,/\, 619 = 405 \,/\, 619 = 0.654$$

Data: Muralidharan, Singh, and Ganimian (2019).

Try it: Read the counts

Score	9	10	11	20	21	22	23	24	25	26
Students	55	66	60	4	5	0	1	1	0	1

Pick one of the 619 students at random. Find each probability

$p(X = 10)$

$p(9 \leq X \leq 11)$

$p(X \geq 20)$

Data: Muralidharan, Singh, and Ganimian (2019).

Try it: Answer

$p(X = 10)$

66 / 619 = **0.107**

$p(9 \leq X \leq 11)$

(55 + 66 + 60) / 619 = **0.292**

$p(X \geq 20)$

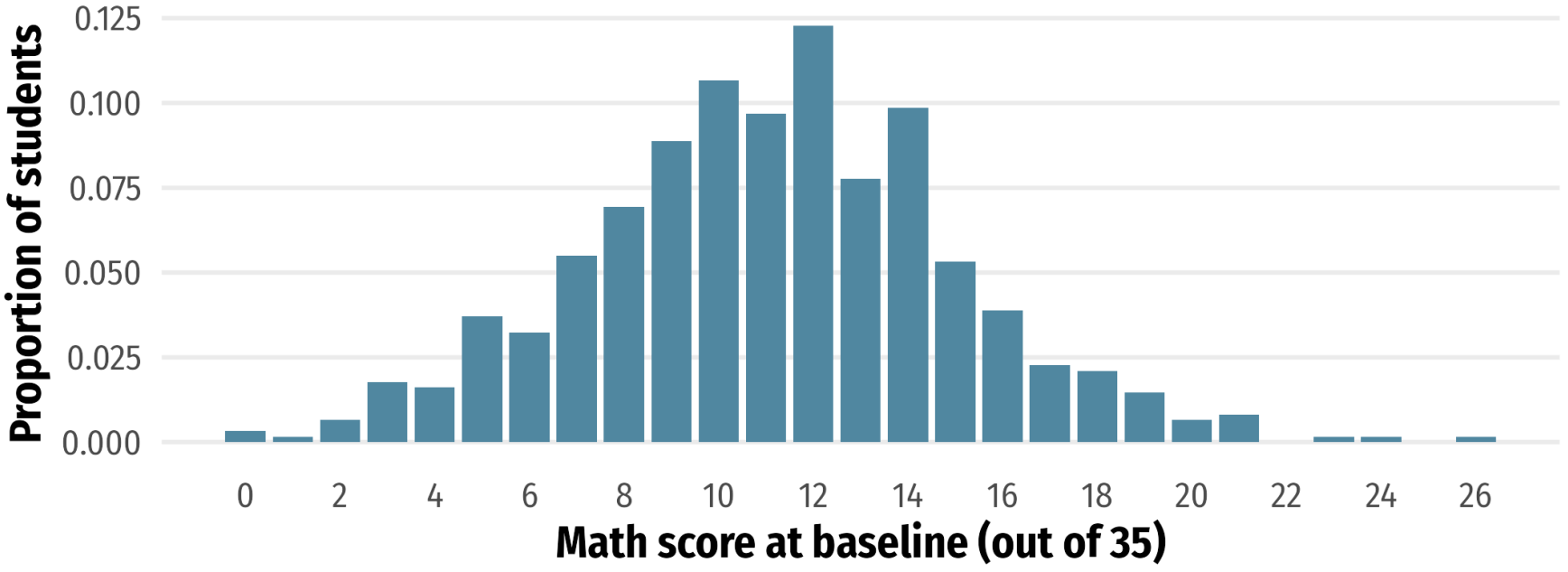
12 / 619 = **0.019**

A score of 20 or more is rare here. That's the sense of "weird" we'll use all quarter

Data: Muralidharan, Singh, and Ganimian (2019).

From bars to curves

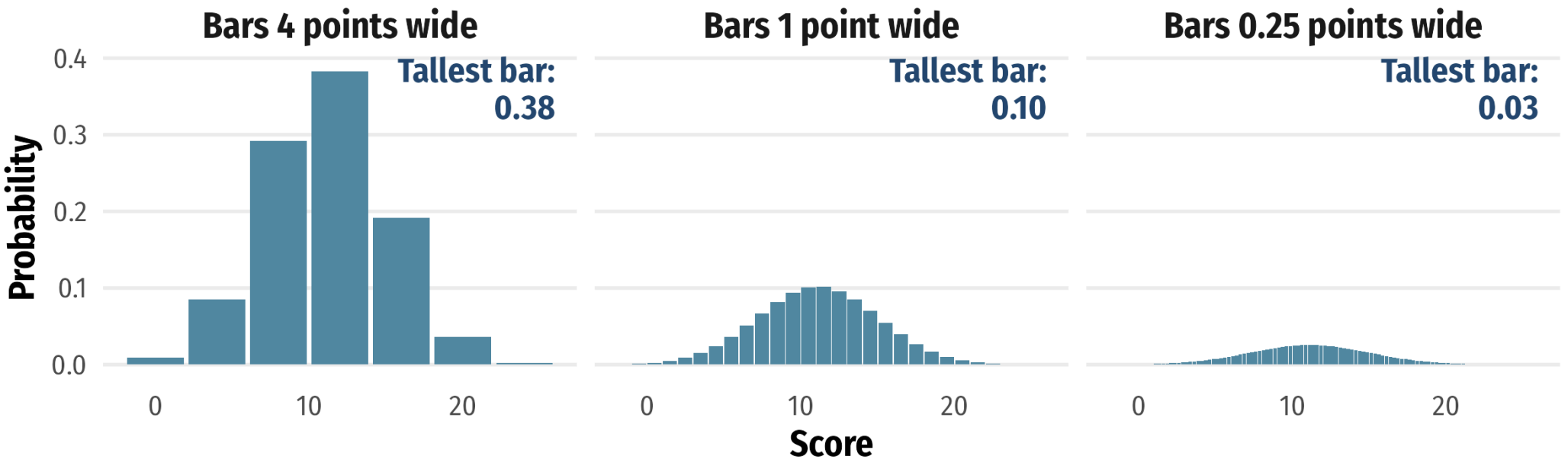
Shares make a probability distribution



Same shape, new axis. Each bar is now a probability, and all the bars add up to 1

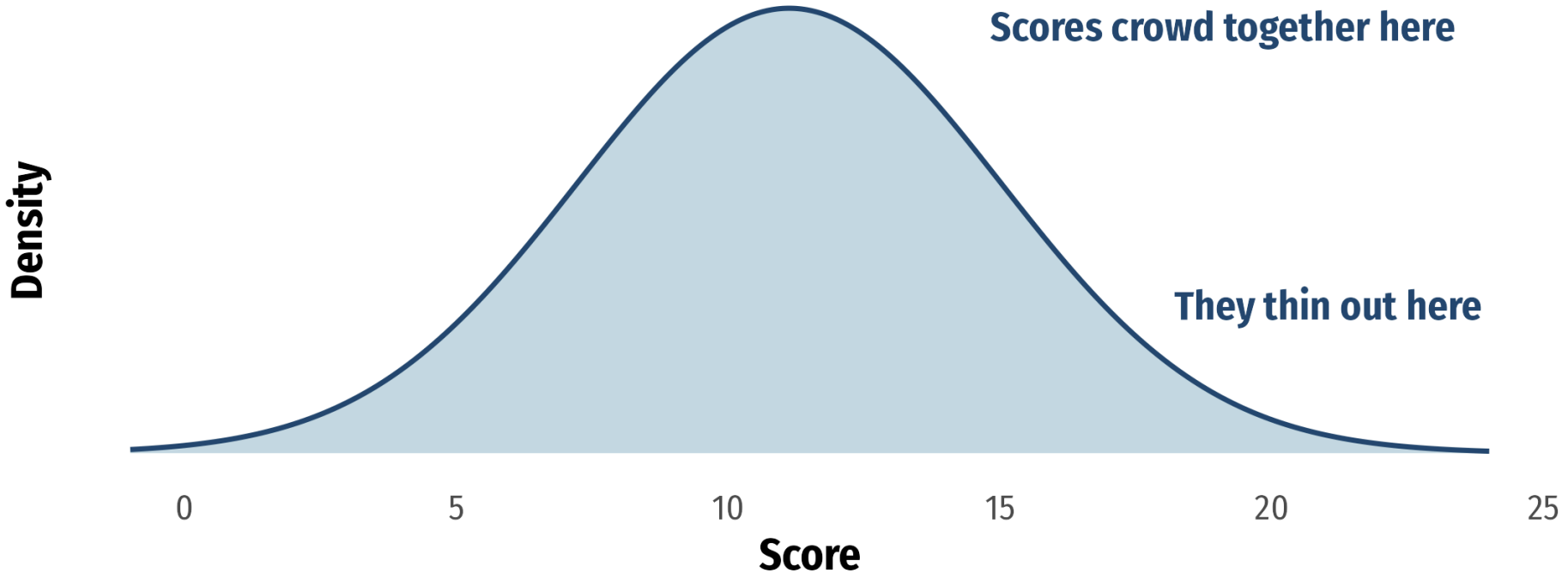
Data: Muralidharan, Singh, and Ganimian (2019).

Continuous scores can't be counted



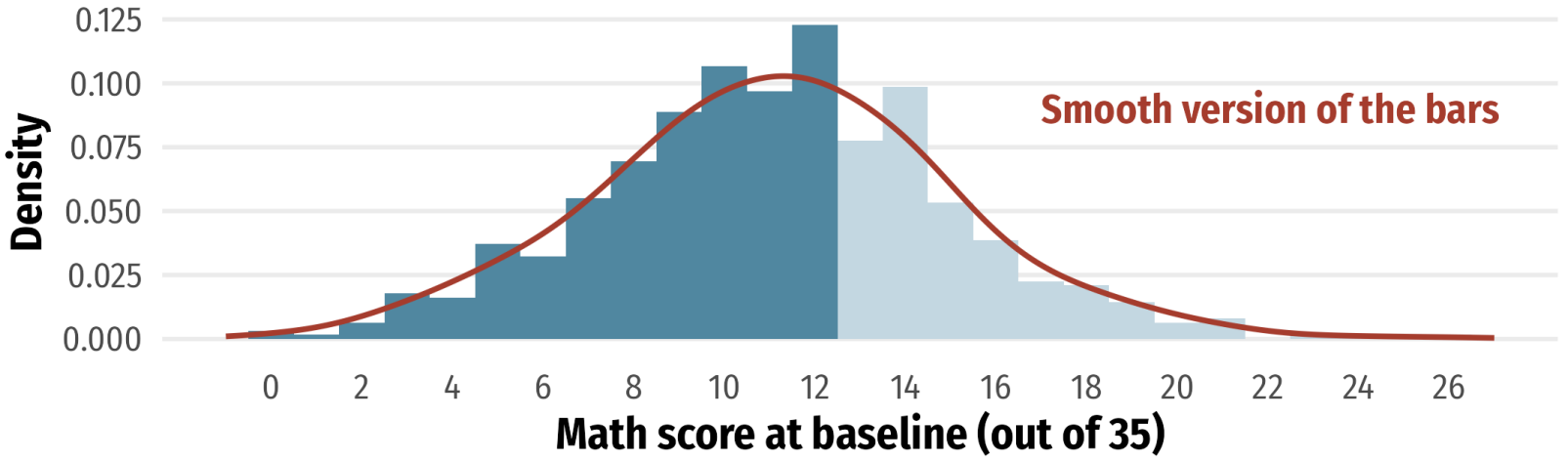
A continuous variable has endless possible values. The chance of any one exact value is zero

Density: where probability piles up



Height isn't a probability. It tells you how concentrated the probability is around that point

Ask about ranges: area is the answer

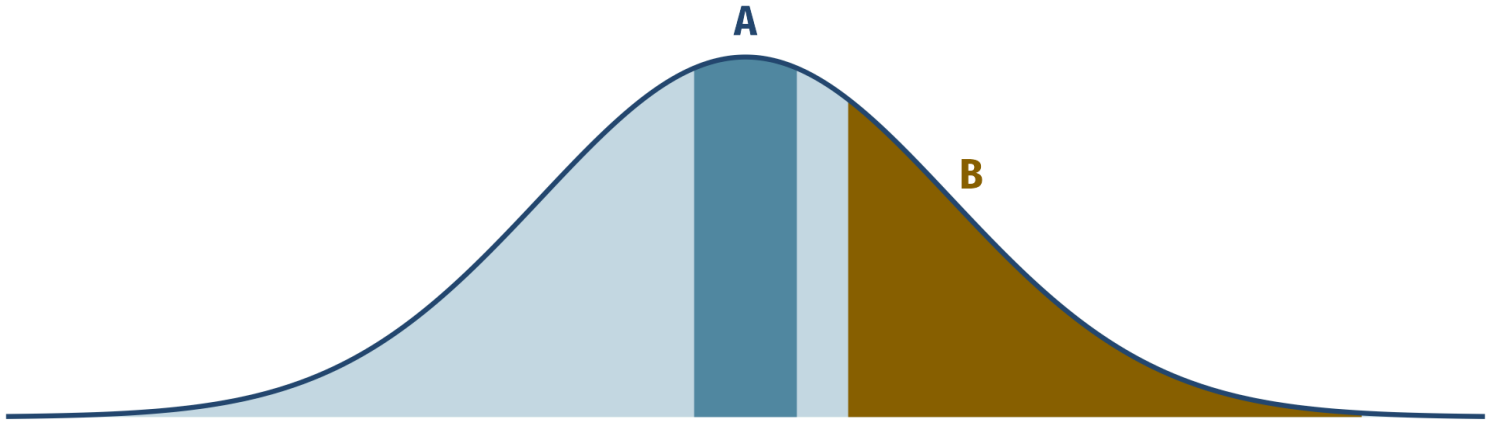


What's the chance of 12 or below?

The whole area is 1, and the dark area is 405 / 619 = 0.654: the same answer as counting

Data: Muralidharan, Singh, and Ganimian (2019).

Predict it: Tall or wide?

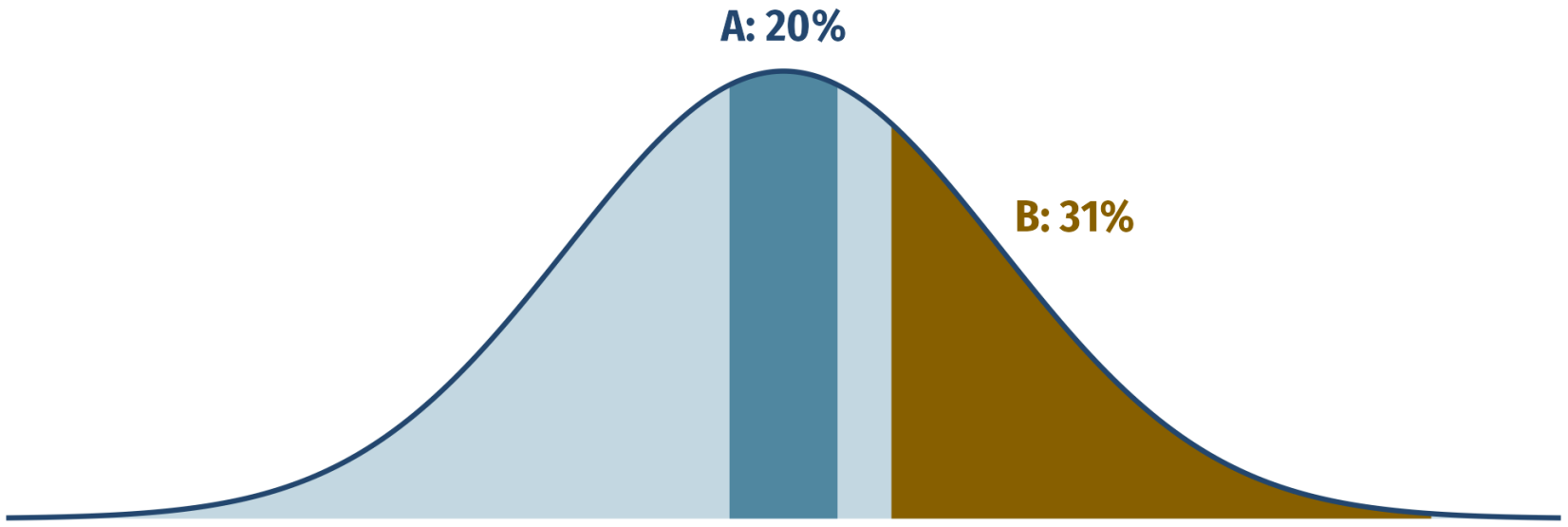


Draw one person at random. Which is more likely?

A The person lands in region A, where the curve is tallest

B The person lands in region B, where the curve is lower

Predict it: Area wins, not height

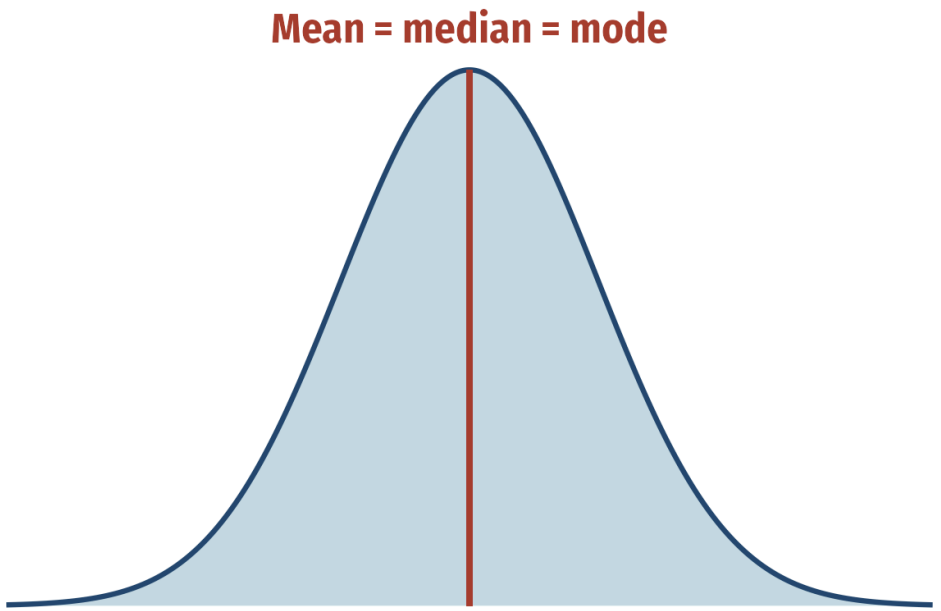


Region B is lower everywhere, but it's so much wider that it holds more of the area

Probability is area. Height alone can't tell you

The normal distribution

A bell curve, set by just two numbers



Symmetric, so one side mirrors the other

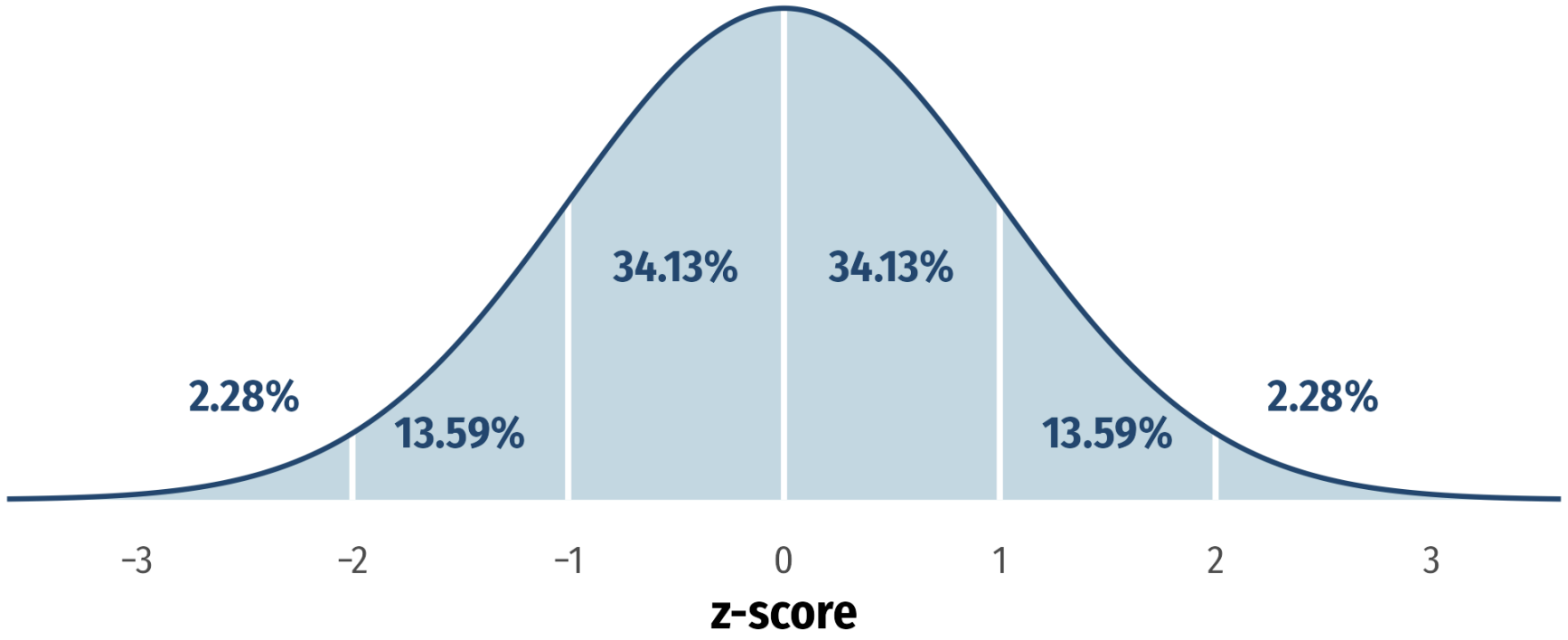
Most common in the middle

Tapers off toward both ends

$$Y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(X-\mu)^2/2\sigma^2}$$

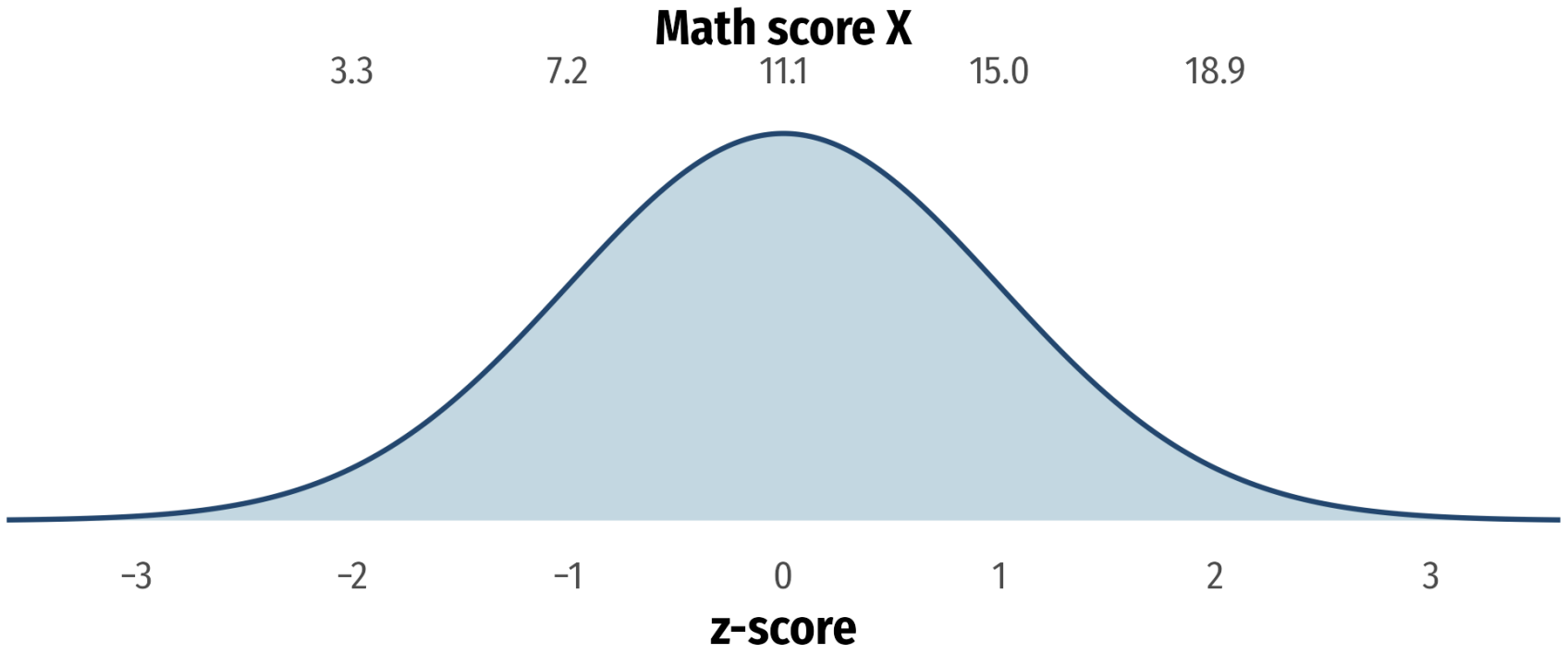
Only μ and σ change. You'll never compute this by hand

Every normal curve splits the same way



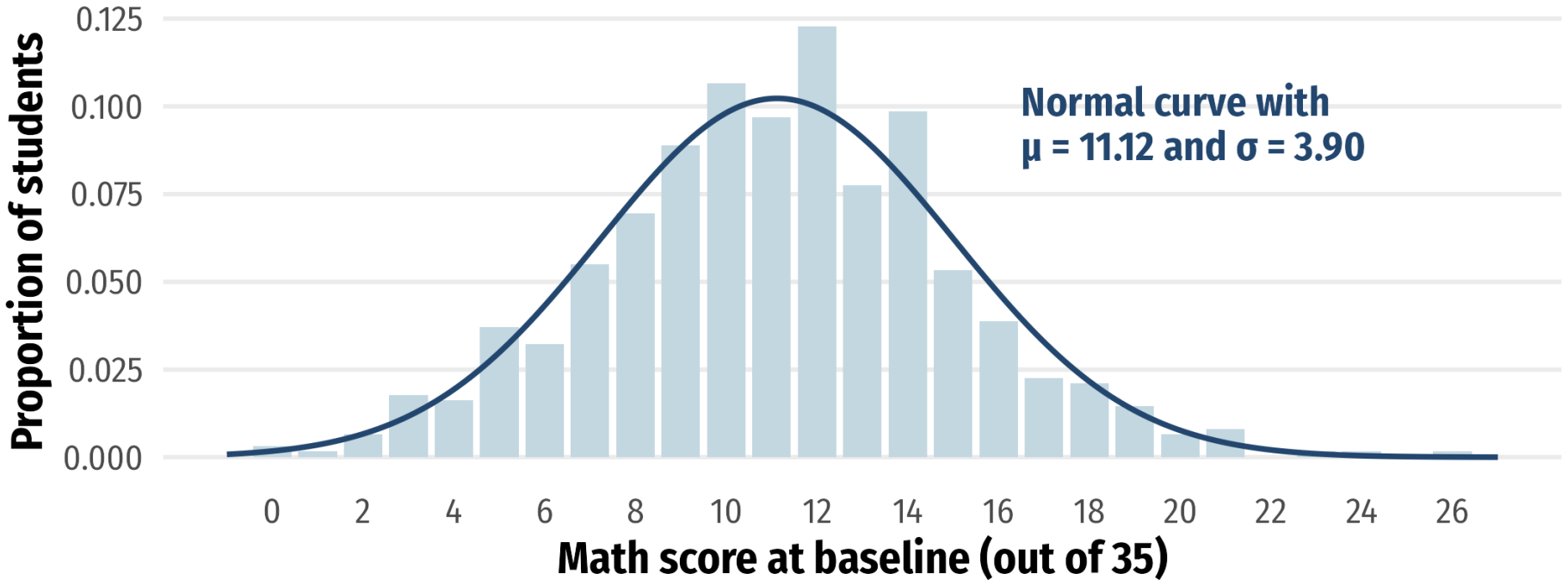
Each mark is one standard deviation. The areas are always the same

The standard normal relabels the axis



$z = (X - \mu) / \sigma$ moves the mean to 0 and the SD to 1. The shape doesn't change

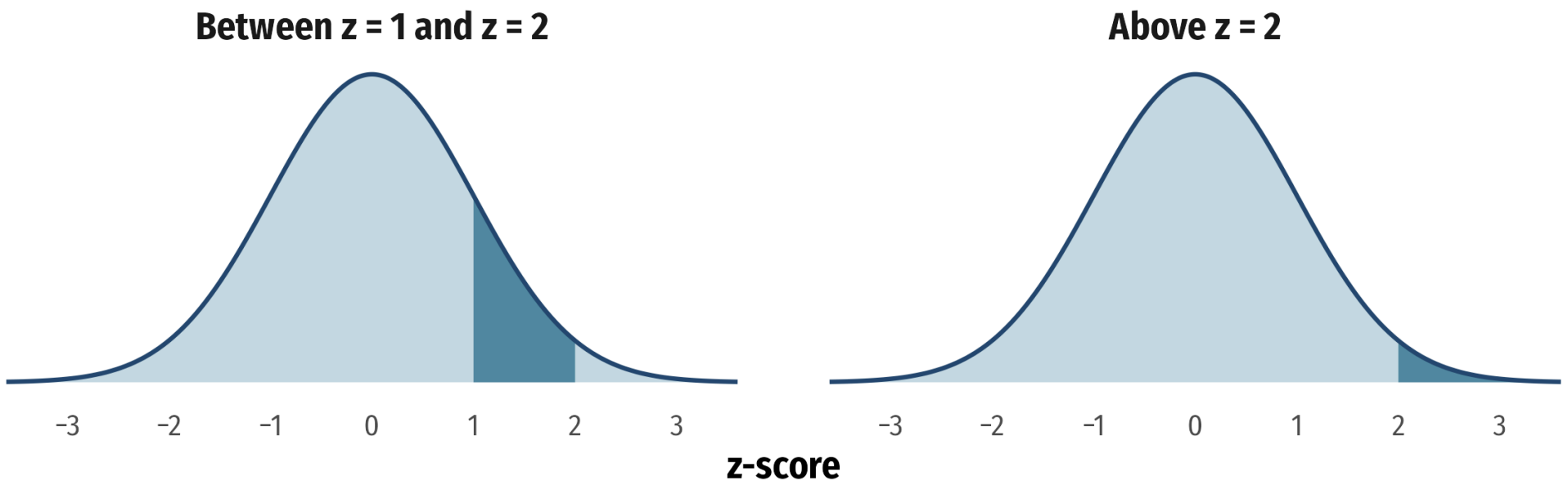
Our 619 scores stay close to the curve



Within 1 σ: 71% of students, 68% of the curve. Within 2 σ: 94% and 95%

Data: Muralidharan, Singh, and Ganimian (2019).

Area is probability, so we can read it off

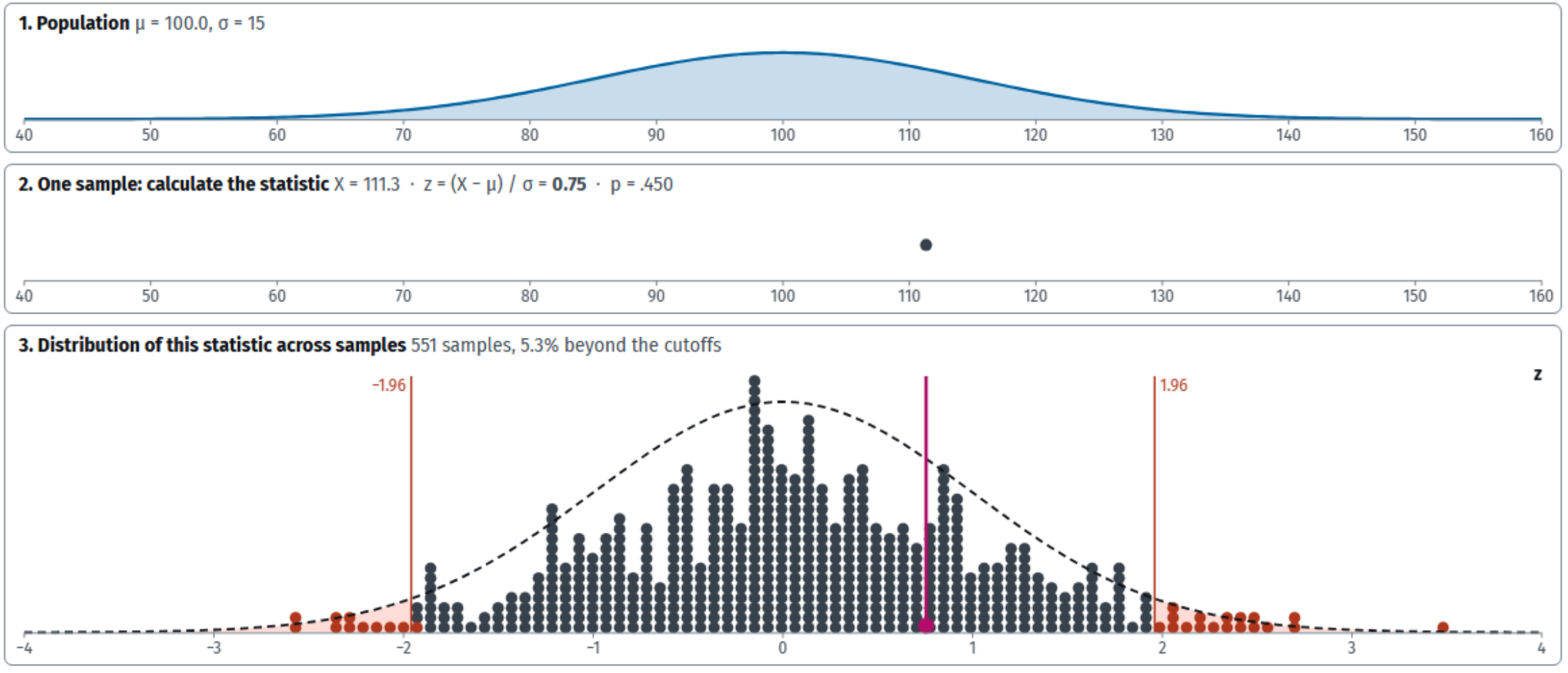


Draw one person from a normal population. What's the chance of each?

$p(1 < z < 2) = 0.1359$

$p(z > 2) = 0.0228$

Watch it happen, one person at a time



Simulator after Kristoffer Magnusson, rpsychologist.com.

Predict it: Halfway out

Respond at **PollEv.com/adb**

You know that p(z > 1) = 0.1587 and p(z > 2) = 0.0228. Now pick the spot halfway between. What's p(z > 1.5)?

A About 2%

B About 7%

C About 9%

D About 16%

Before Thursday

Finish Chapter 6

Read sections 7-1 and 7-2 for the second half of class

Bring a calculator: we'll use the table

Session 2: Thursday, January 21

Tuesday in one line: probability is area under a curve

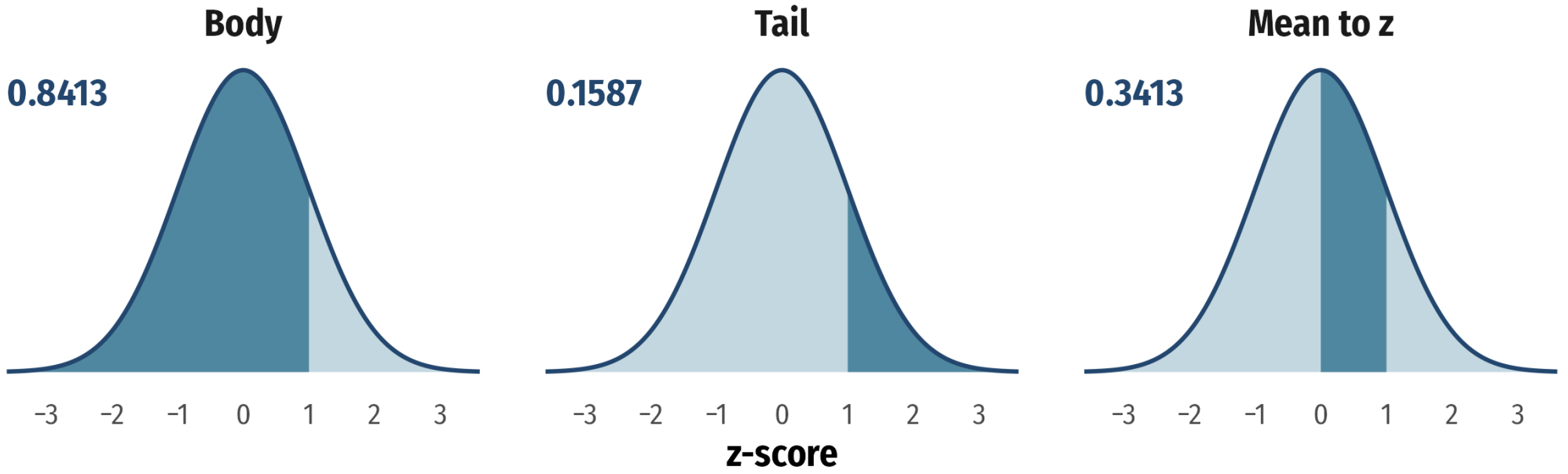
The unit normal table

From scores to percentiles

Then Topic 6 begins: sampling distributions

The unit normal table

The table answers one z in three ways



These are the three areas for z = 1. Body and tail always add up to 1

A piece of the table

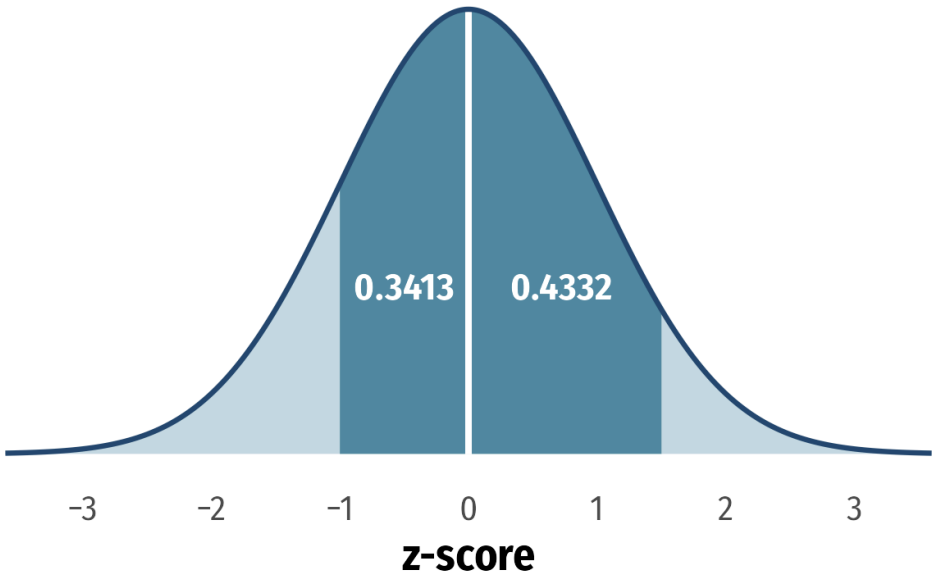
z	Body	Tail	Mean to z
0.25	0.5987	0.4013	0.0987
0.50	0.6915	0.3085	0.1915
1.00	0.8413	0.1587	0.3413
1.50	0.9332	0.0668	0.4332
1.96	0.9750	0.0250	0.4750
2.00	0.9772	0.0228	0.4772

Tuesday's poll: the row for 1.50, column Tail, gives $p(z > 1.5) = 0.0668$

The full table is at the back of the book, and exams come with a copy

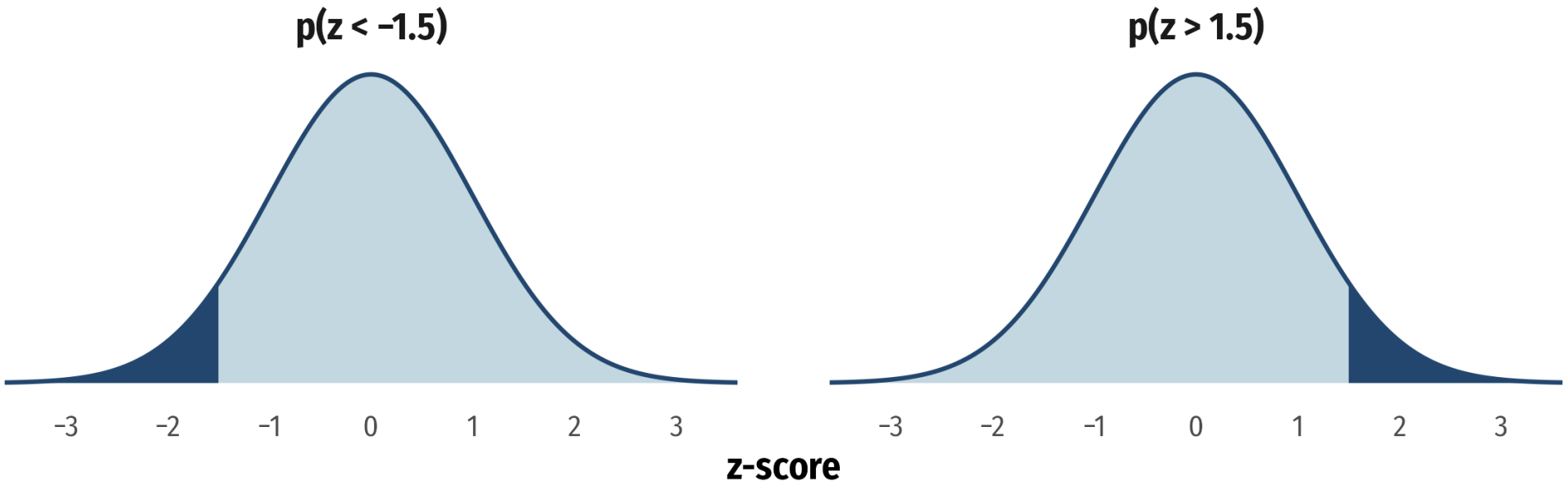
Sketch first, then pick the column

1. Draw the curve and shade the area you want
2. Mark your z-scores
3. Is that a body, a tail, or a piece from the mean?
4. Read the column, then add or subtract



$p(-1 < z < 1.5) = \mathbf{0.7745}$

Negative z? The curve is a mirror



The table lists only positive z. Both tails here equal 0.0668, so drop the sign and read the tail

The table also runs backwards



Find 0.0500 in the Tail column. It falls between z = 1.64 and 1.65, and we use 1.65

Split 5% into 0.0250 per tail. The Tail column gives z = 1.96

A spreadsheet does the same lookup

You want	Formula, in Excel or Google Sheets	Result
Body for z = 1.5	=NORM.DIST(1.5, 0, 1, TRUE)	0.9332
Tail for z = 1.5	=1 - NORM.DIST(1.5, 0, 1, TRUE)	0.0668
z for a body of 0.95	=NORM.INV(0.95, 0, 1)	1.645

The 0 and the 1 are μ and σ. The body is everything below z, so a tail is 1 minus the formula

Try it: Use the table

z	Body	Tail	Mean to z
0.25	0.5987	0.4013	0.0987
0.84	0.7995	0.2005	0.2995
1.28	0.8997	0.1003	0.3997

1 Find $p(z < 0.25)$

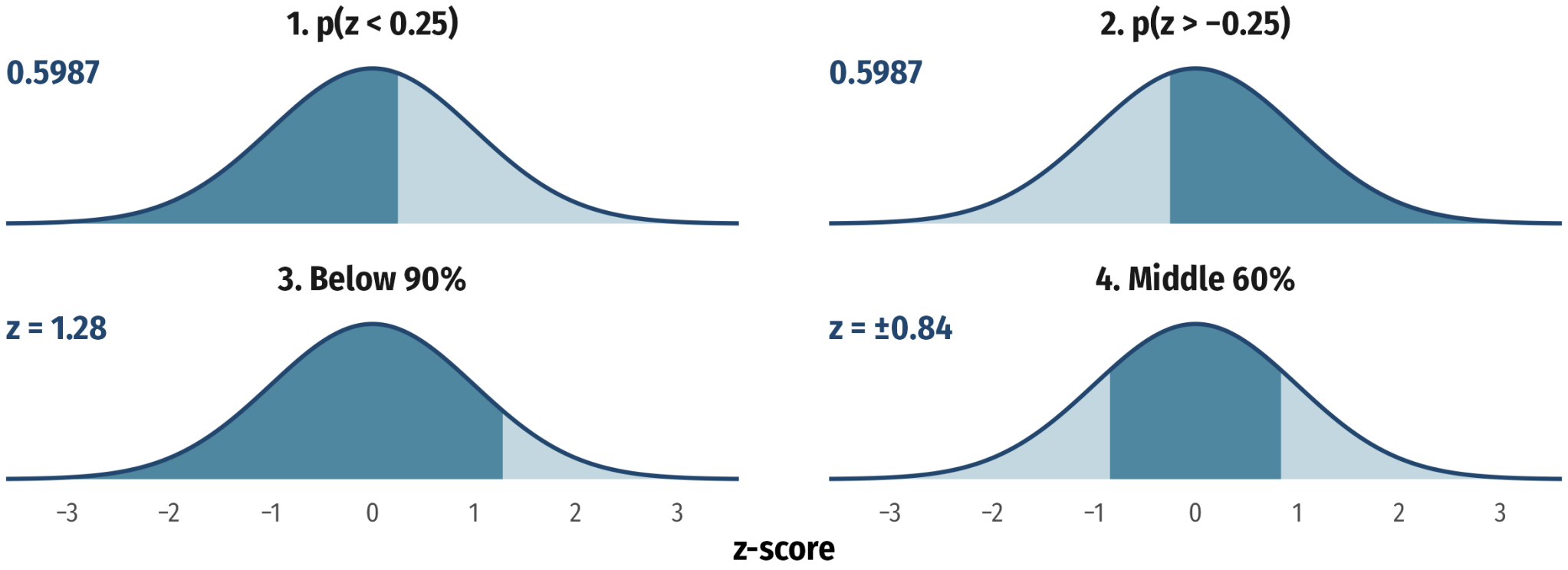
2 Find $p(z > -0.25)$

3 90% of scores fall below which z?

4 The middle 60% lies between which two z-scores?

Sketch first. Work alone, compare with a neighbor, then we vote

Try it: Answer

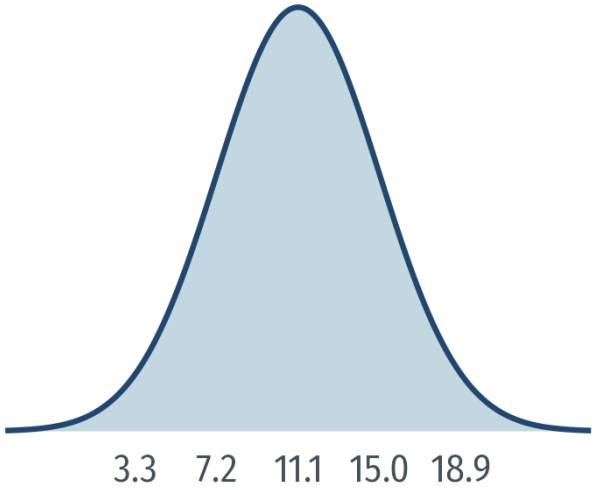


The first two match because the curve is symmetric. For the last one, each tail holds 20%

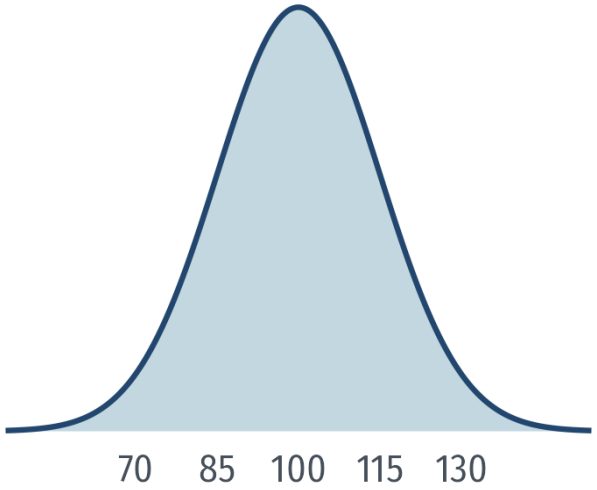
From scores to percentiles

Normal curves differ only in their labels

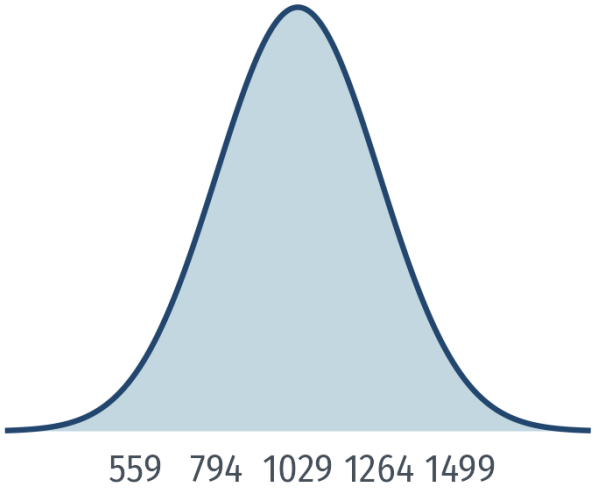
Mindspark: μ = 11.12, σ = 3.90



IQ test: μ = 100, σ = 15



SAT: μ = 1029, σ = 235



Convert to z and one table serves them all. But the population must be normal, or the table's areas are wrong

Mindspark: Muralidharan, Singh, and Ganimian (2019). SAT: College Board, 2025 SAT Suite Annual Report.

Three stops: score, z, probability

Score X — $z = (X - \mu) / \sigma$, $X = \mu + z\sigma$ — **z-score** — Unit normal table — **Probability**

Have a score?

Go right: find z, then read the area

Have a share?

Go left: read z from the table, then find X

You can't jump straight from X to a probability. Every trip goes through z

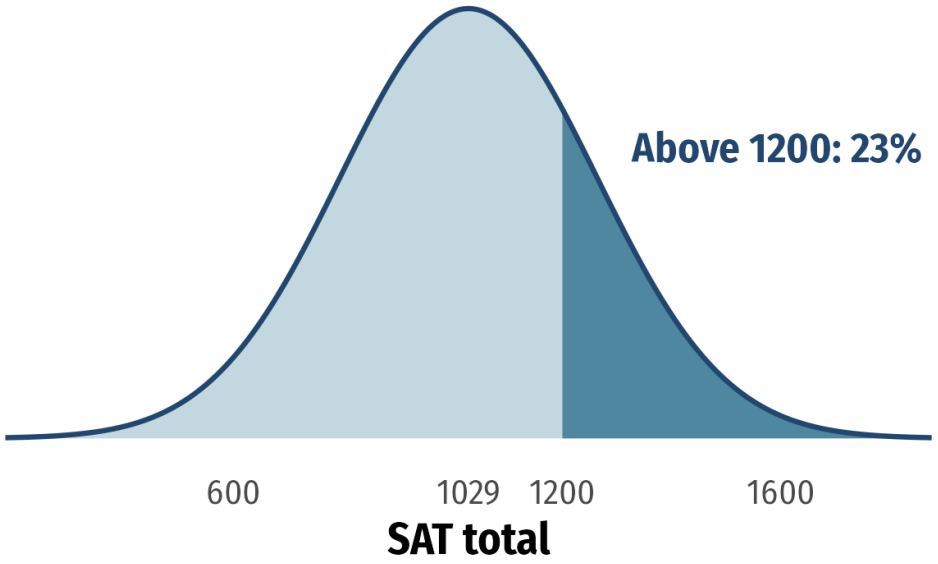
After a figure in Gravetter et al., *Essentials of Statistics for the Behavioral Sciences*.

How many students scored above 1200?

Class of 2025: μ = 1029, σ = 235. Treat it as normal

z = (1200 − 1029) / 235 = 0.73

Tail for z = 0.73: 0.2327, so about 23%



Real count at 1200 or above: 507,341 of 2,004,965, or 25%. Close, not exact

Data: College Board, 2025 SAT Suite Annual Report, Total Group.

Percentile rank: the share at or below

1200 has a percentile rank of 77

23% scored above it, so 100 − 23 = 77% scored at or below it

It's the body of the curve at that score

Percentile rank describes a person's place in the group. It isn't the percent they got right

Data: College Board, 2025 SAT Suite Annual Report, Total Group.

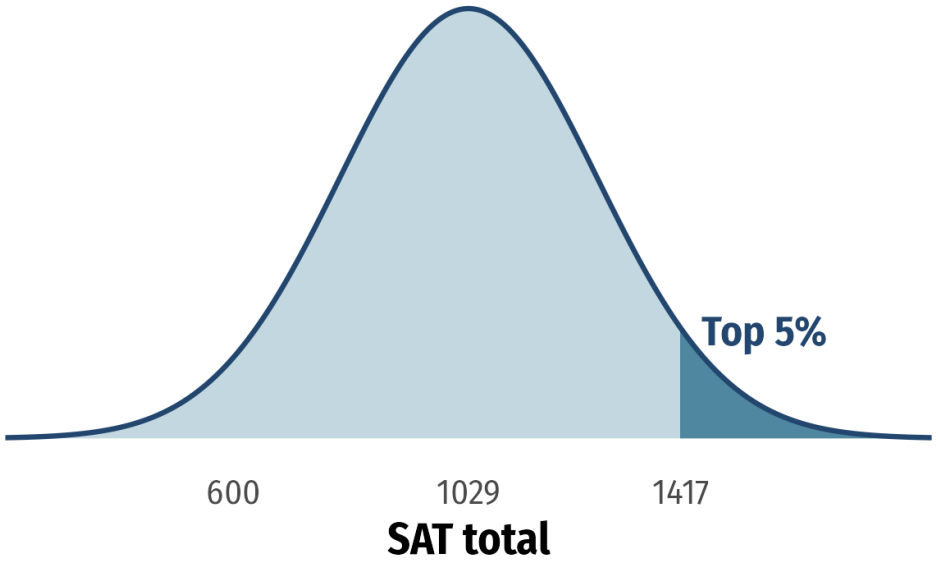
A percentile: the score at a given share

To be in the top 5%, what score did you need?

The 95th percentile has a tail of 0.05, so z = 1.65

$$X = \mu + z\sigma = 1029 + 1.65 \times 235$$

X = 1417, or about 1420 on the SAT's scale



Data: College Board, 2025 SAT Suite Annual Report, Total Group.

Play it: Warm-up for Friday

Curves and percentiles. Answer as a question. Nothing is scored

For 100

The percentile rank of z = 0

What is the 50th?

For 200

The share of a normal population above z = −1

What is about 84%?

Play it: For 300

Respond at **PollEv.com/adb**

Mindspark's curve has μ = 11.12 and σ = 3.90. Treating it as normal, this score is the 90th percentile. In the table, the tail 0.10 is closest at z = 1.28

A What is 6.1?

B What is 12.0?

C What is 14.6?

D What is 16.1?

Data: Muralidharan, Singh, and Ganimian (2019).

From one score to a whole sample

So far: how weird is one score?

Next: how weird is one sample's mean?

Our 619 students stay a population. We draw many samples, and their means pile up into a curve of their own

Areas under that curve drive the four steps of a hypothesis test

What you can do now

- Find a probability as a share of outcomes, or as an area
- Say why height on a density curve isn't a probability
- Use the unit normal table, forward and backward
- Go from a score to a z-score to a probability, and back
- Find a percentile rank and a percentile

Next steps

Right now

Topic 6 starts: sampling distributions, from sections 7-1 and 7-2

This Friday: Lab 3

Quiz, covering everything through z-scores and probability. Bring a calculator

Mastery and Problem Set for Chapter 6 are due Sunday, January 24 at 6 PM