Central tendency and variability

Topic 3 · Tuesday, January 12

EDUC 15: Statistics for Education Research
UCI School of Education

Where are we?

Description

Summarize data

Visualize data

We are here

Prediction

Explanation



Peter the Anteater at a UC Irvine basketball game.

Plan for today

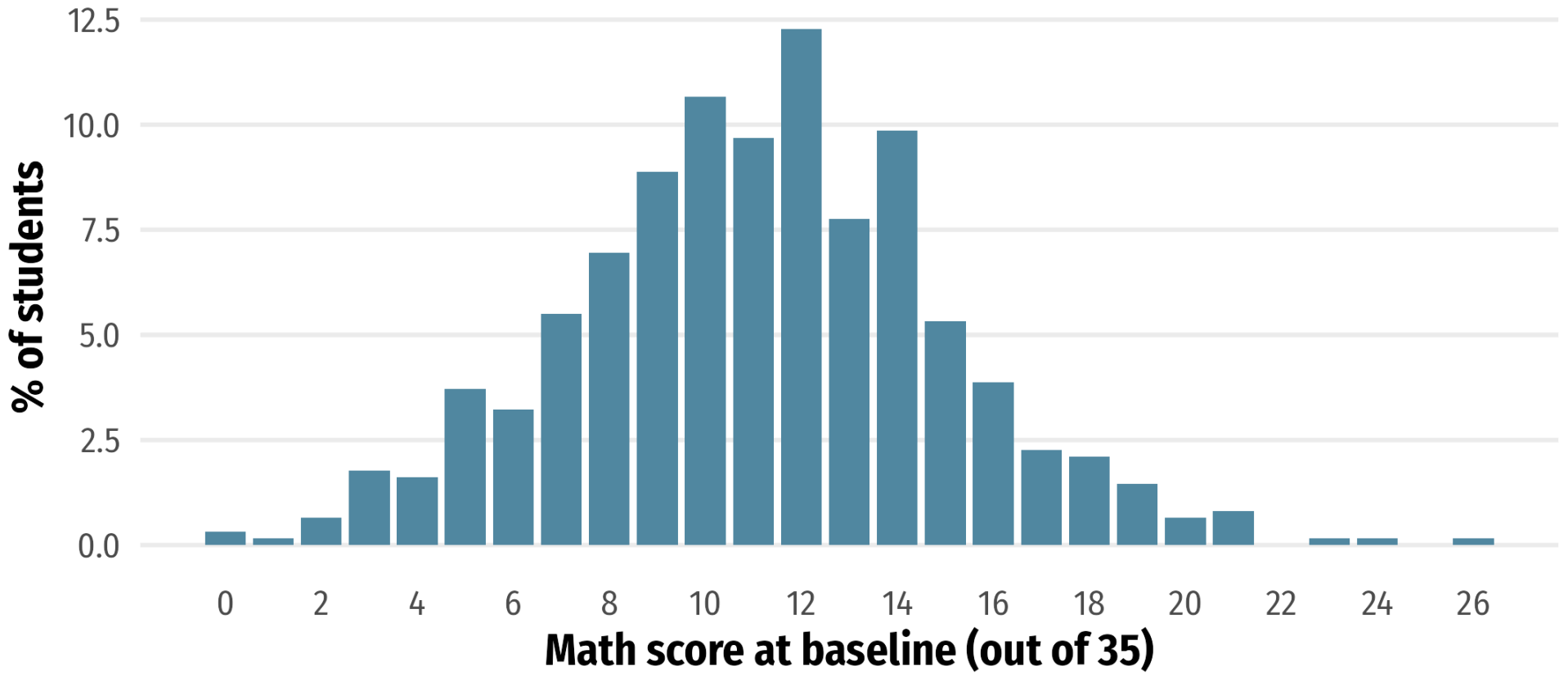
What's typical?

Shape and spread

The standard deviation

What's typical?

What value best describes this group?



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

Three ways to name a "typical" value

Mode

The most frequent value

Find the value that occurs most often

Median

The midpoint of a distribution

Rank the scores and take the middle one. With an even number, average the middle two

Mean

The average, or the balance point

Add up all the scores, then divide by how many there are

The mean: one idea, two symbols

$$\mu = \frac{\sum X}{N}$$

Population

μ is the mean of everyone we care about. There are **N** of them

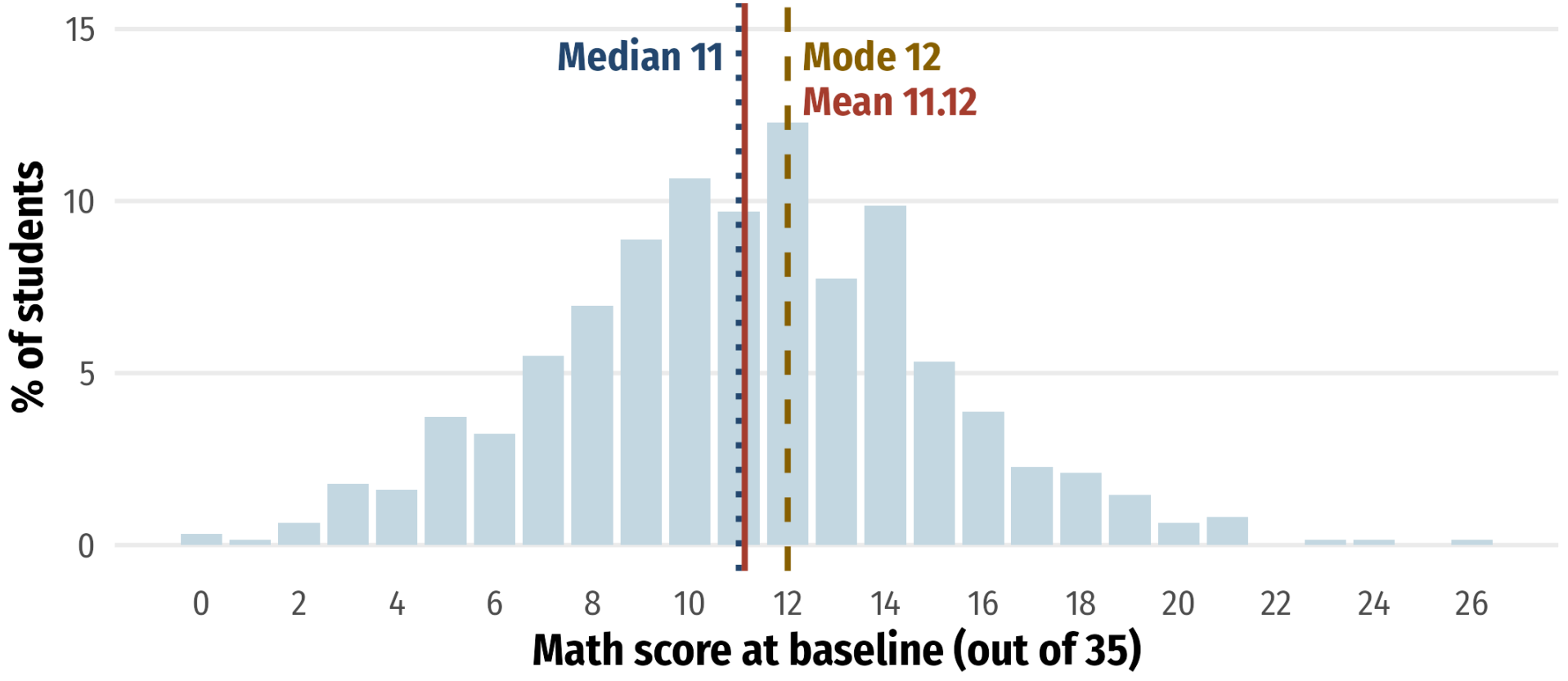
$$M = \frac{\sum X}{n}$$

Sample

M is the mean of the **n** people we observed. It estimates μ

Greek means population. A Latin letter or a hat means an estimate

Here, all three land close together



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

Predict it: One perfect score

Respond at **PollEv.com/adb**

Five students scored 8, 12, 12, 13, 13. Their mean is 11.6 and their median is 12. Suppose one 13 becomes a perfect 35. What happens?

A Only the mean changes

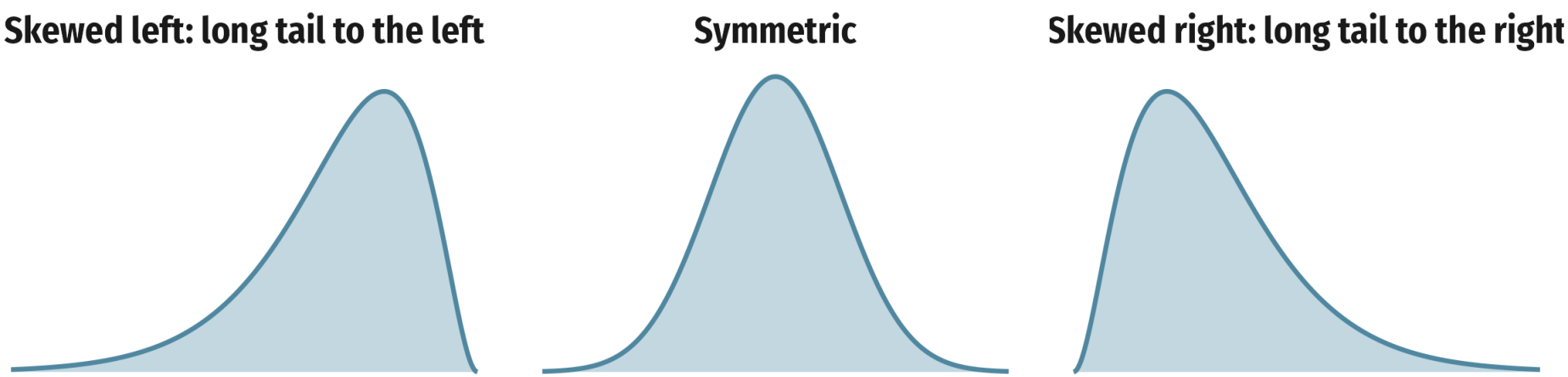
B Only the median changes

C Both change

D Neither changes

Shape and spread

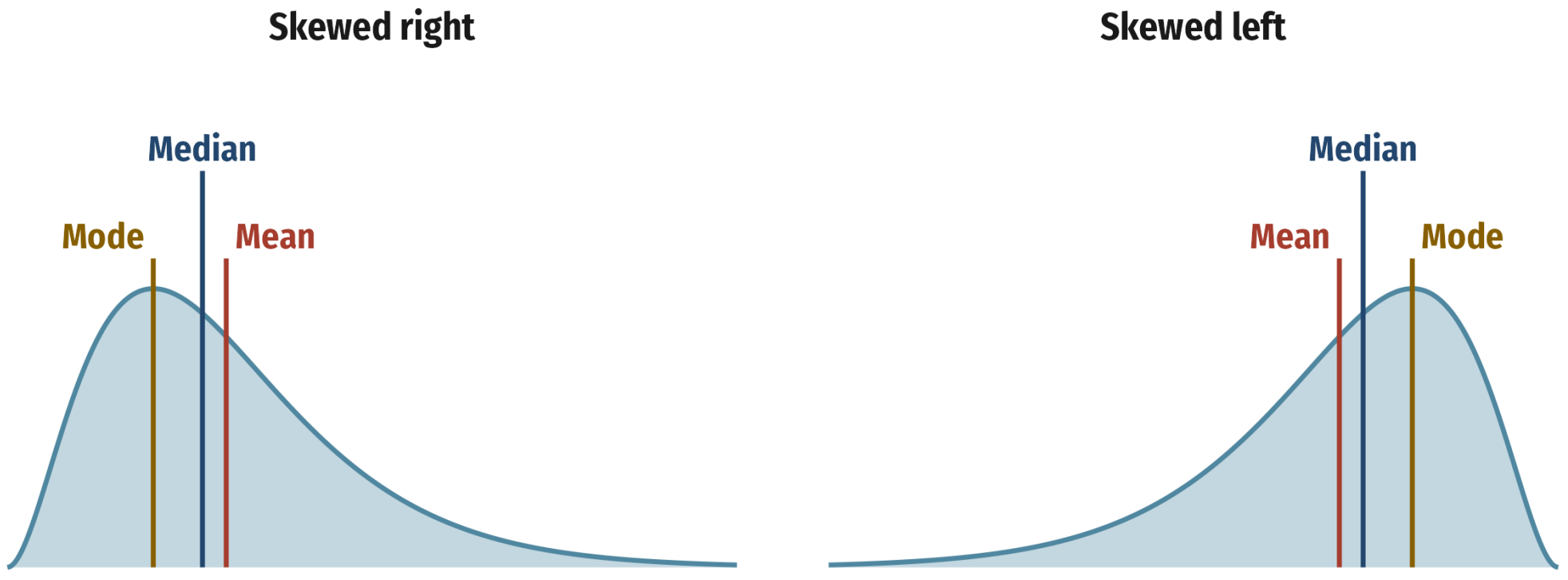
Skew is named after the tail



Our math scores are almost symmetric, so all three agreed

Dogs: Jawlik, *Statistics from A to Z: Confusing Concepts Clarified* (Wiley, 2016).

In a skewed distribution, they pull apart



The mean gets pulled toward the tail. That's the perfect score you just saw

So which one should you report?

Measure	Works well for	Can mislead with
Mean	Interval or ratio scales, when you know every score	Extreme scores, skewed distributions, open-ended categories
Median	Extreme scores, skewed distributions, open-ended categories, ordinal scales	Nominal scales
Mode	Nominal scales, discrete variables, describing shape	Interval or ratio data, unless it comes with the mean or median

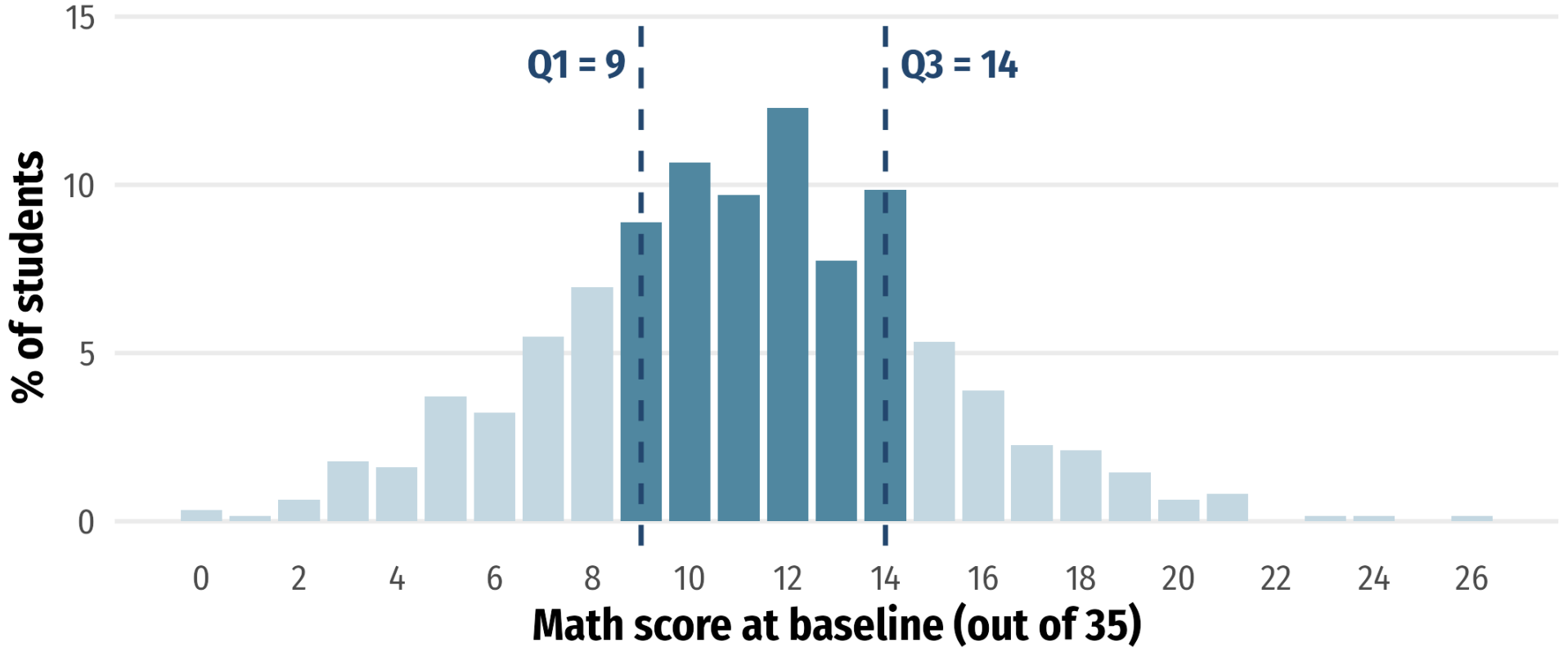
The range: how far the scores stretch

Range = highest score − lowest score

In our data: 26 − 0 = 26 points

It's easy to find, but only two scores set it

The interquartile range: the middle 50%



IQR = 14 − 9 = 5 points. One extreme student can't move it

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

Try it: Five real students

Scores: 8, 13, 12, 12, 16

Find the mean, the median and the range

Work alone for two minutes. Compare with a neighbor. Then we vote

Try it: Answer

Mean

(8 + 13 + 12 + 12 + 16) / 5 = **12.2**

Median

In order: 8, 12, 12, 13, 16. The middle one is **12**

Range

16 − 8 = **8**

The standard deviation

How far is a typical score from μ?

Say these five students are everyone we care about, so μ = 12.2

Score X	Deviation X − μ	Squared deviation
8	−4.2	17.64
13	0.8	0.64
12	−0.2	0.04
12	−0.2	0.04
16	3.8	14.44

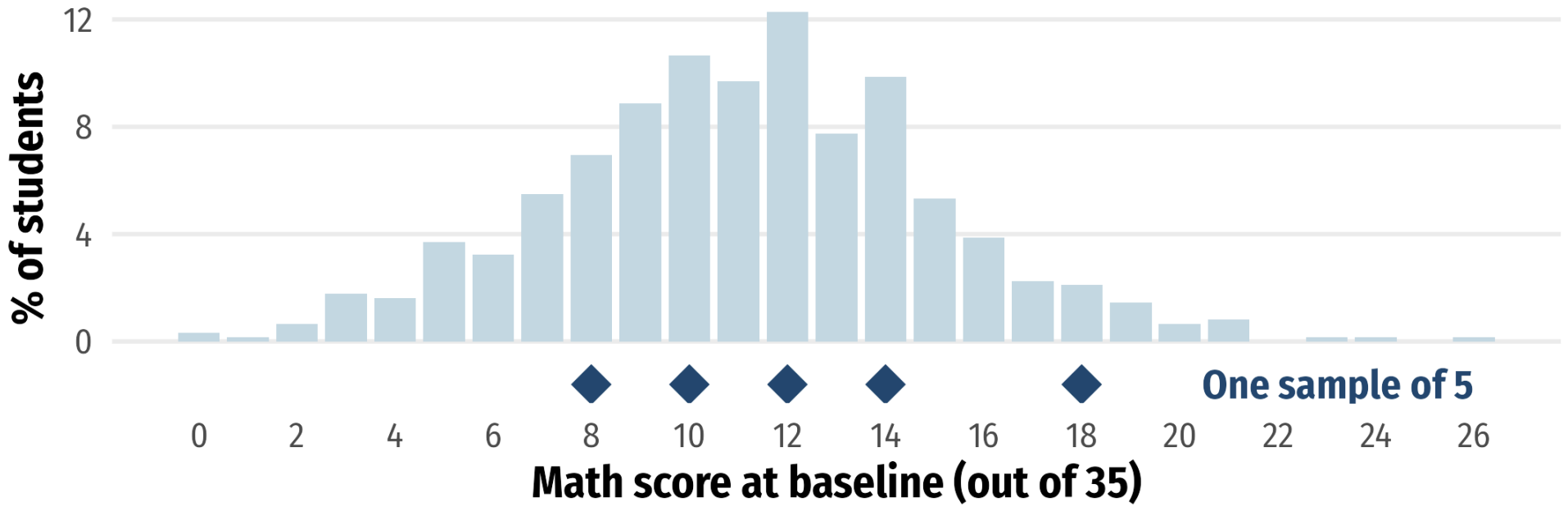
Add up the last column to get the sum of squares: SS = 32.8

Four steps to the standard deviation

Step	In symbols	Our five students
1. Find each deviation	$X - \mu$	The table we just built
2. Square them and add them up	$SS = \sum(X - \mu)^2$	SS = 32.8
3. Average them: the variance	$\sigma^2 = SS/N$	32.8 / 5 = 6.56
4. Take the root: the SD	$\sigma = \sqrt{\sigma^2}$	$\sqrt{}$6.56 = 2.56

Variance is the average squared distance. The SD puts it back in points

We rarely see the whole population



A small sample rarely catches the extreme scores, so it tends to look less spread out than its population

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

Predict it: Treat the 619 students as a population

Their variance is $\sigma^2 = 15.22$. Draw 5,000 samples of 5. In each one, compute SS / n

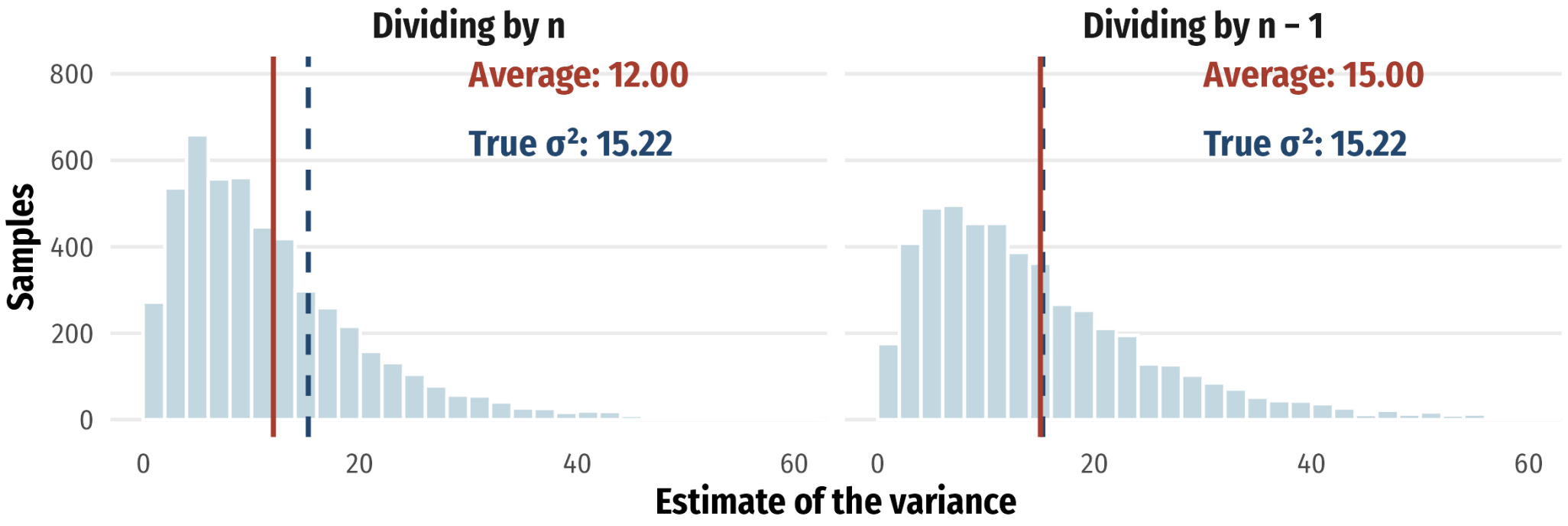
Across all 5,000 samples, what does SS / n average out to?

A Less than 15.22

B About 15.22

C More than 15.22

Predict it: SS / n falls short on average



SS / n tends to be too small, though 28% of samples still overshoot. Dividing by n − 1 lands on target on average

Data: Muralidharan, Singh, and Ganimian (2019). Samples drawn from those scores.

Our estimate of the population variance

$$\hat{\sigma}^2 = \frac{SS}{n-1}$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

The hat says this is an estimate. n – 1 is called the *degrees of freedom*

SS / n is still the sample's own variance. It just isn't our best guess about the population

In the book: $\hat{\sigma}^2$ is written s^2, and $\hat{\sigma}$ is written s

Our five students, now as a sample

	Population	Sample
How many	N	$n = 5$
Mean	μ	$M = 12.2$
Variance	$\sigma^2 = SS / N$	$\hat{\sigma}^2 = 32.8 / 4 = 8.20$
Standard deviation	σ	$\hat{\sigma} = 2.86$

With all 619 students, n – 1 hardly matters: $\hat{\sigma} = 3.90$, against 3.90 when dividing by n

What you can do now

Find and choose among the mean, median and mode

Say how skew pulls the mean toward the tail

Compute SS, the variance and the SD by hand

Label a symbol as a parameter or an estimate, and say why we divide by n − 1

Next steps

Before Thursday

Read Chapter 5

This Friday: Lab 2

Bring a calculator and a laptop. It includes a 15-minute individual check

Mastery and Problem Set for Chapter 5 are due Sunday at 6 PM

Bonus: A faster way to compute SS

Start from the definition, multiply out the square, and use μ = ΣX / N:

$$SS = \sum (X - \mu)^2 = \sum X^2 - 2\mu \sum X + N\mu^2$$

$$SS = \sum X^2 - \frac{(\sum X)^2}{N}$$

This needs no deviations, so it's quicker with a calculator