

Psych 10: Intro to Statistics

Lecture 7 · z-scores and the normal distribution · Wed Oct
14, 2026

Prof. Judith Fan · Fall 2026

Reminders & Announcements

- CourseKata **Chapter 6** is due Friday.
- Friday: **Case #3, The Forged Average**, in your tutorial session. Check your discussion section's Slack channel for your time slot. Bring a pen.
- Milestone 3 is due Sunday, October 18 on Gradescope. Milestone 1 grades have now been posted on Canvas, and Milestone 2 grades will be posted by the end of the week.
- **OAE accommodations:** To take the exam (Wed Nov 18) at the Centralized Testing Center, you set it up yourself, ahead of time. Go to the OAE website, open *Student CTC Resources* (oae.stanford.edu/ctc/student-ctc-resources), and follow its steps in OAE|CONNECT.

Words with exact meanings

The words on the left name a general idea. Each term on the right has a precise definition and should be used only in that sense.

Everyday language	Statistical terms and their definitions
center, typical, "the average"	mean (the sum divided by n) · median (the middle value in order) · mode (the most common value)
spread, variation, variability	range · IQR · variance (SS divided by $n - 1$) · standard deviation (the square root of the variance)
error, how far off	deviation (a value minus the mean) · sum of squares (the squared deviations, added up)
unusual, extreme, an outlier	z-score (a deviation divided by the SD)
normal	the normal distribution (one specific bell-shaped curve)
random	random sample (every member of the population equally likely to be chosen)
related, goes with	correlation (r , the strength of a straight-line relationship)

Today

z-scores and the normal distribution



How many SDs from the mean?



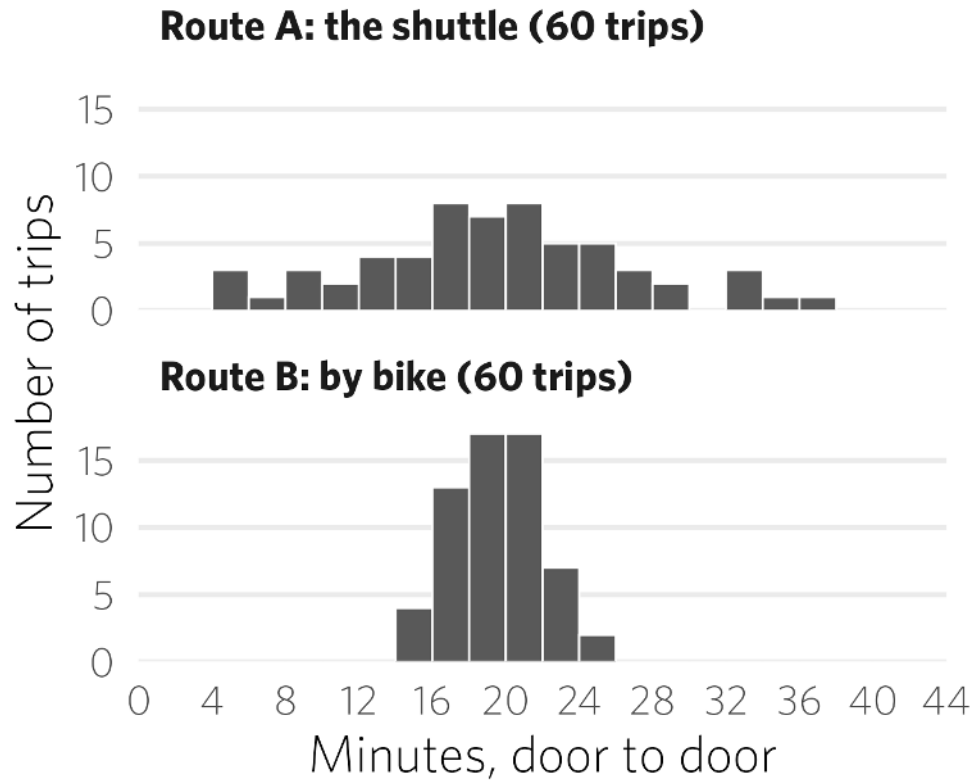
The normal distribution



Does the curve fit?

On Monday we defined the SD as a measure of spread. Today it is the unit for stating how far one value is from the mean, and, with the mean, it specifies a model for the whole distribution.

When to leave for the exam



Monday's two routes. Both have a mean of 20 minutes; Route A's SD is 8.5 minutes and Route B's is 2.2. The exam is at 1:30.

With your neighbor: come up with a suggestion for when to leave, in terms of the mean and the SD, and give the departure time for each route.

Leaving for the exam

Route A: $20 + 2(8.5) = 37$ min

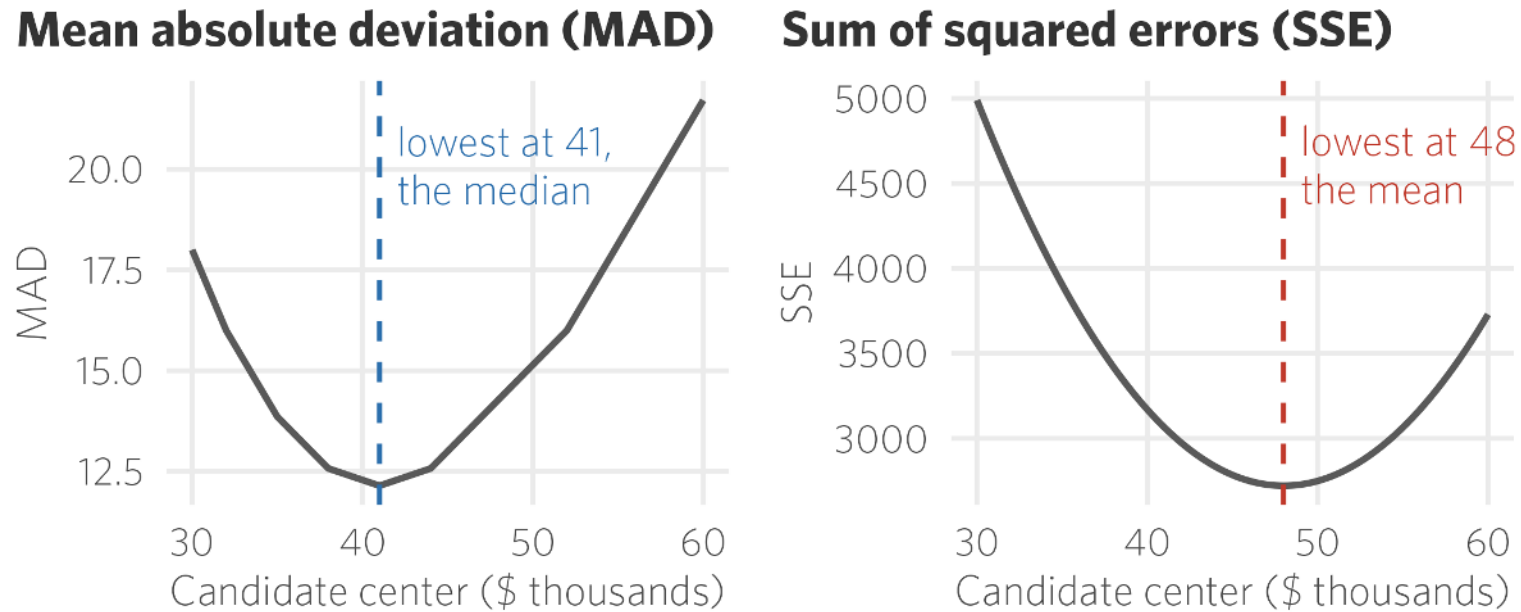
Route B: $20 + 2(2.2) = 24.4$ min

in seconds: $\bar{Y} \times 60, s \times 60$

$SS \times 60^2, s^2 \times 60^2$

- Allowing the mean plus two SDs gives 37 minutes on Route A and 24.4 on Route B: leave by 12:53 or by 1:05.
- The means are equal, so only the SD distinguishes the routes. How many SDs to allow depends on the cost of arriving late.
- The SD is in minutes, like the mean, so their sum is also in minutes. In seconds, the mean and the SD would each be 60 times larger, and the SS and the variance 3,600 times larger.

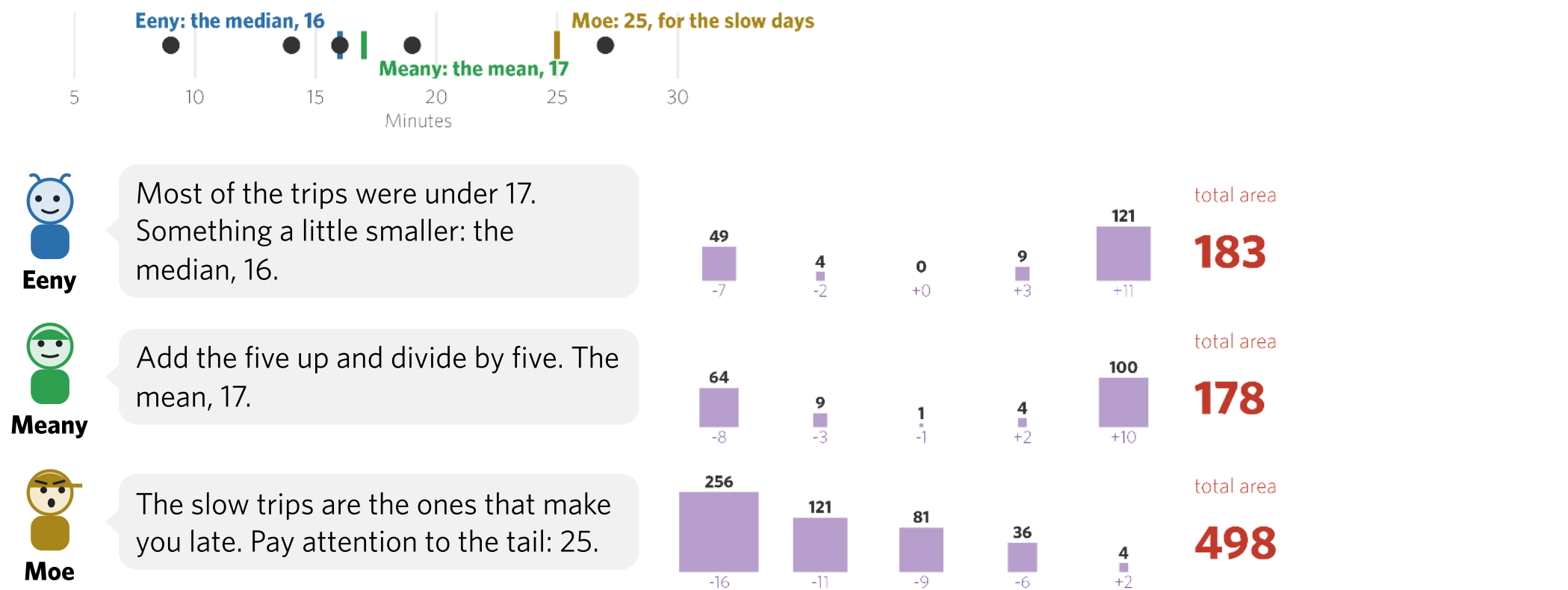
The mean minimizes the sum of squares



The median minimizes the mean absolute deviation (MAD), and the mean minimizes the sum of squared errors (SSE). The next slide returns to what the second claim means, with three proposed centers for five trips.

Eeny, Meany, and Moe pick a center

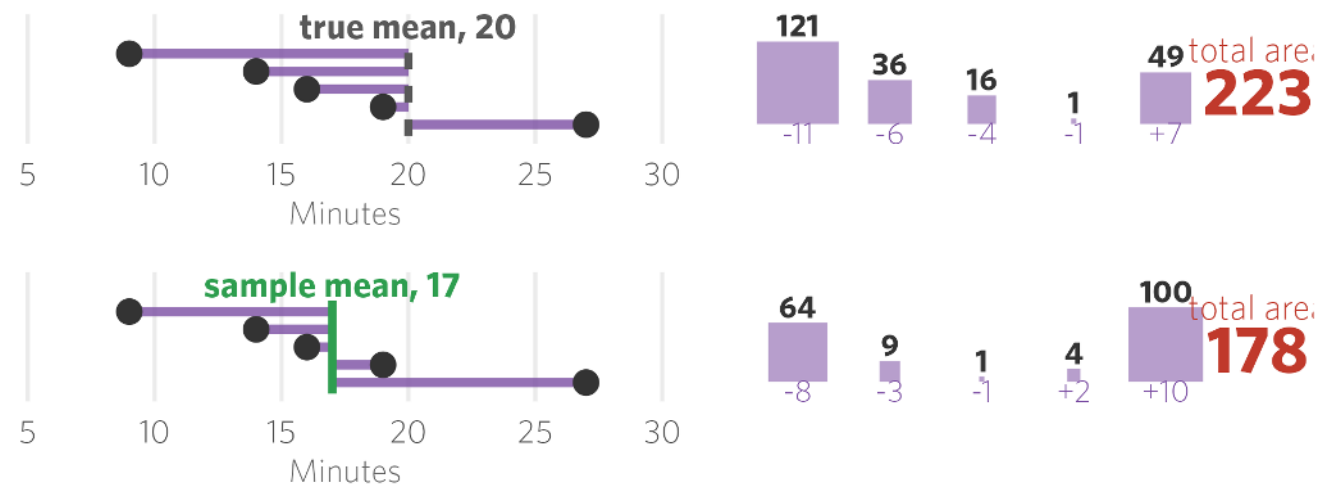
Five trips took 9, 14, 16, 19, and 27 minutes. Which one number stands for all five?



Meany's total is the smallest. Any other center gives a larger total: a little larger near the mean, much larger out in the tail. The mean minimizes the sum of squares.

Why the sample variance divides by $n - 1$

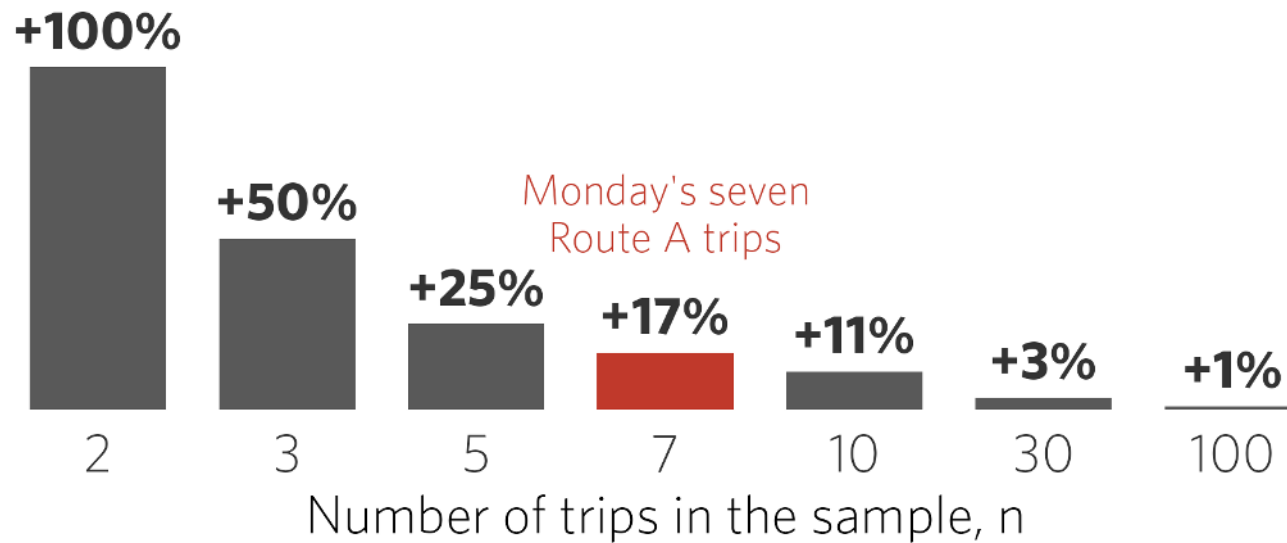
Suppose the five trips are a sample from a route whose true mean is 20 minutes. The true mean is never known; the sample mean, 17, takes its place.



From the true mean the squares total 223; from the sample mean, 178. This holds in general: the sample mean minimizes the sum of squares, so the true mean, like any other center, gives a larger total. The variance computed from the sample mean is therefore too small, and dividing by $n - 1$ rather than n corrects for the shortfall.

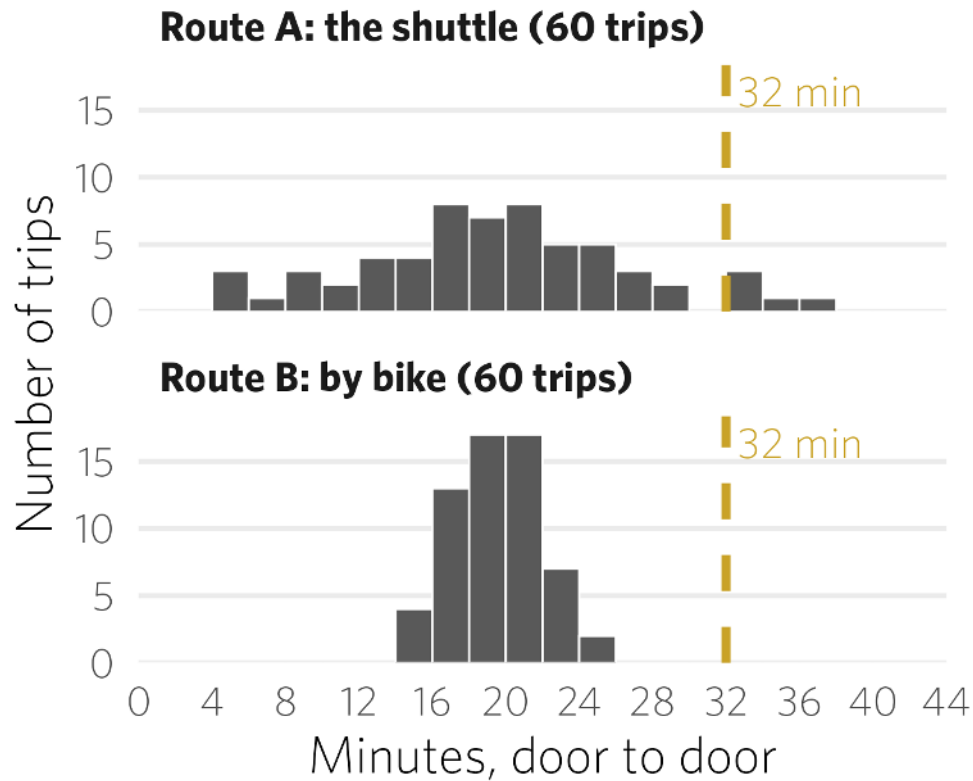
The size of the $n - 1$ correction

How much larger the estimate is with $n - 1$ as the divisor than with n



Dividing by $n - 1$ rather than n gives an estimate $n/(n - 1)$ times as large: double for two trips, a sixth larger for Monday's seven Route A trips, 1% larger for a hundred. With many trips the sample mean is close to the true mean, and the shortfall is a small part of a large total.

Is 32 minutes a long trip?



Monday's two routes, both with a mean of 20 minutes. Today a trip took 32 minutes.

On Route A, is that a bad day or an ordinary one? What about on Route B?

The deviation divided by the SD

On either route, 32 minutes is 12 minutes above the mean, so neither the value nor the deviation says which case is unusual. What differs between the routes is the SD.

$$\text{Route A: } \frac{32 - 20}{8.5} = \frac{12}{8.5} = 1.4$$

$$\text{Route B: } \frac{32 - 20}{2.2} = \frac{12}{2.2} = 5.5$$

- On Route A, 12 minutes is 1.4 SDs, a longer trip than most but not far outside the usual range for that route.
- On Route B, 12 minutes is 5.5 SDs, far outside the usual range for that route, which suggests that something unusual happened on that trip.
- The number that answers “how unusual is this value?” is the deviation measured in SDs.

The z-score

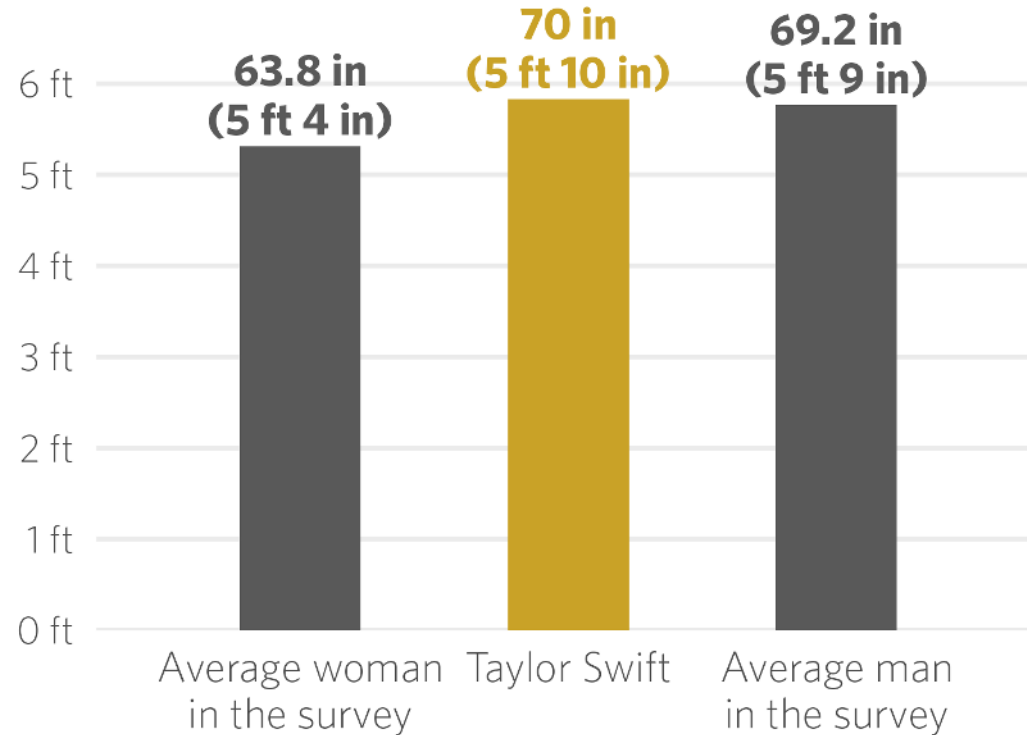
The **z-score** of a value is its distance from the mean, measured in SDs:

$$z = \frac{Y - \bar{Y}}{s}$$

Y is the value, **$\bar{Y}$** the mean, and **s** the SD, so **z** is the deviation divided by the SD: 1.4 for the trip on Route A, 5.5 on Route B.

Minutes divided by minutes leaves no units, so a z-score is a pure number, and z-scores from different variables, inches and dollars, are comparable: each states a distance from its own mean in units of its own SD.

Taylor Swift is 5 ft 10 in

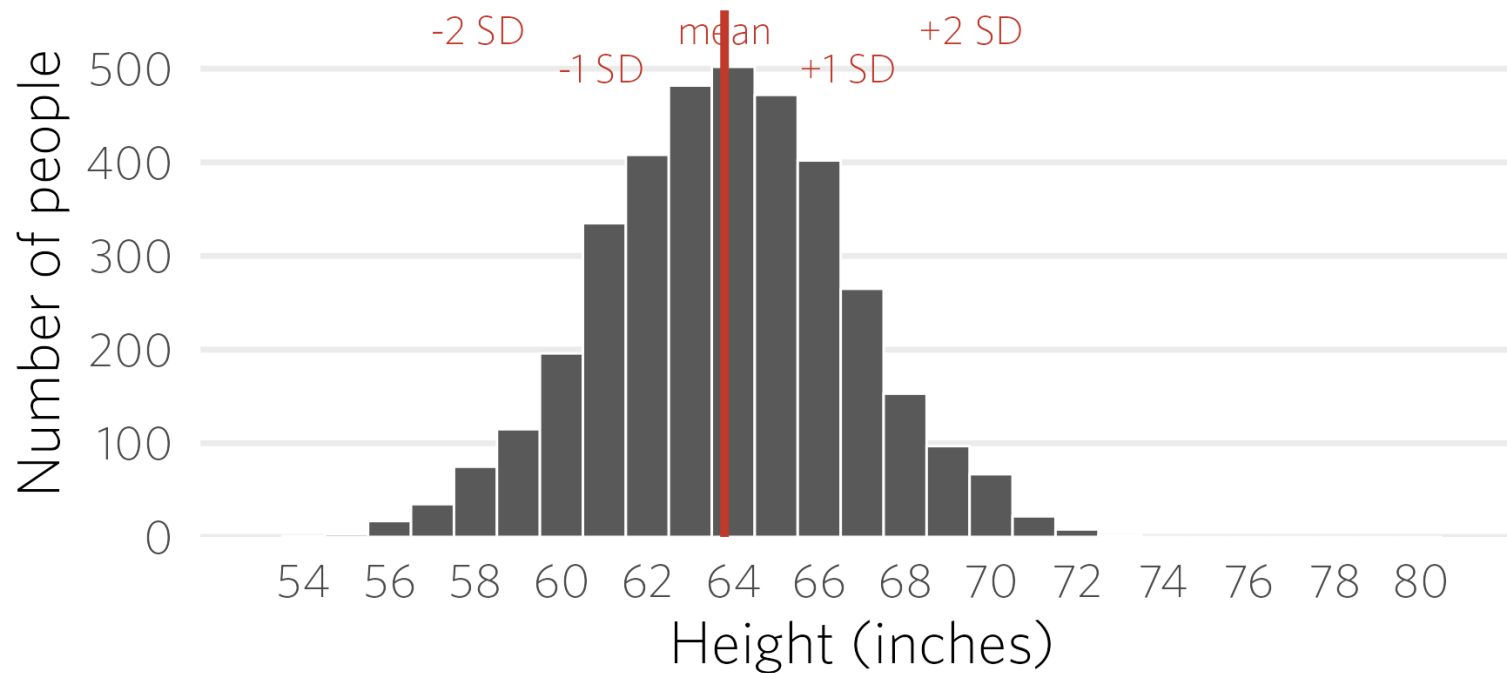


Seventy inches is about 6 inches above the mean height of adult women in a national survey and about an inch above the men's mean.

Photo: Paolo V, the Eras Tour, Inglewood, August 2023, CC BY 2.0, Wikimedia Commons.

Heights in a national survey

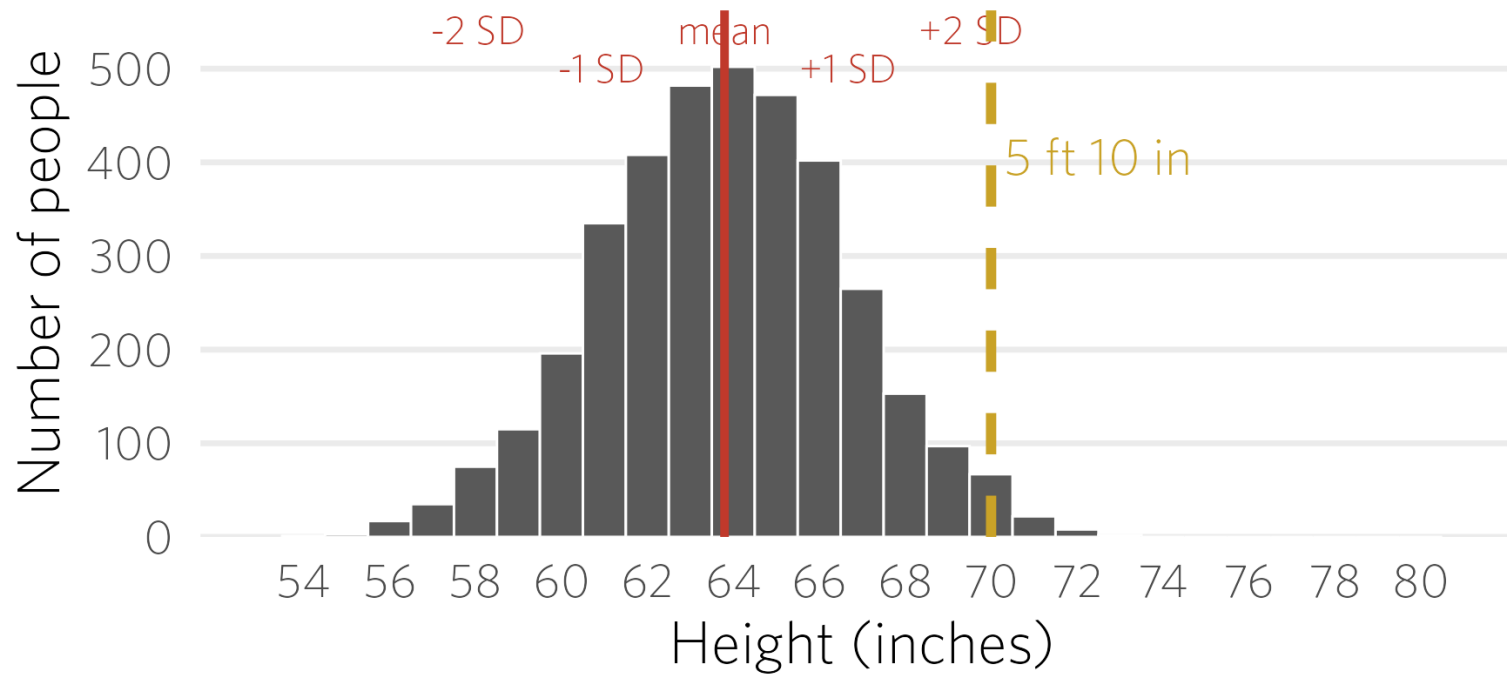
3,658 women in the NHANES survey: mean 63.8 inches, SD 2.9 inches



The CDC's National Health and Nutrition Examination Survey (NHANES) measures, rather than asks, each participant's height. The histogram shows the 3,658 women aged 20 and over in the 2009–2012 surveys.

How unusual is 5 ft 10 in?

3,658 women in the NHANES survey: mean 63.8 inches, SD 2.9 inches



Her 70 inches on the women's histogram. **Is that unusually tall for a woman? Would it be for a man?**

Worksheet · Part 1

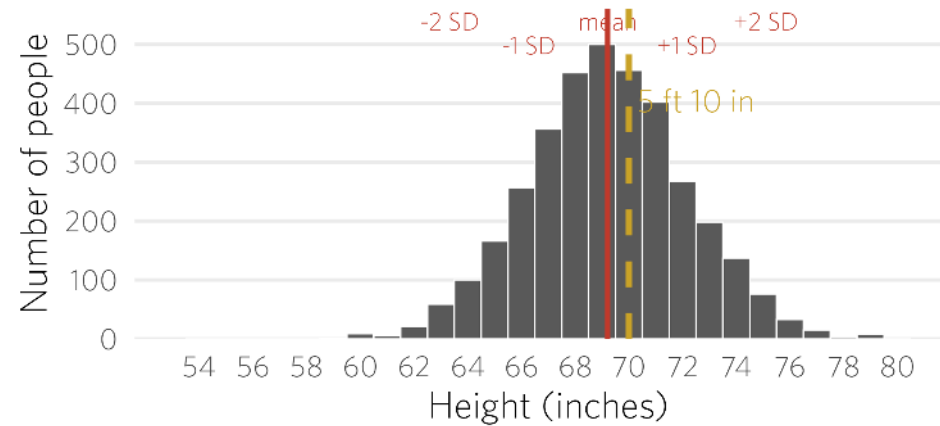
$$z = \frac{Y - \bar{Y}}{s}$$

In the NHANES survey, adult women's heights have a mean of 64 inches and an SD of 3 inches, rounded to whole inches. Men's heights have a mean of 69 inches and an SD of 3 inches.

- a.** Taylor Swift is 5 ft 10 in, or 70 inches. How many SDs above the women's mean is 70 inches? Show your arithmetic.
- b.** How many SDs from the men's mean is a man of 70 inches?
- c.** What is the z-score of a woman who is 61 inches tall (5 ft 1 in)?
- d.** What do the sign of a z-score (whether it is positive or negative) and its size each tell you about the value?

Part 1a-b · 70 inches among women and among men

3,524 men in the NHANES survey: mean 69.2 inches, SD 2.9 inches



$$Z_{\text{women}} = \frac{70 - 64}{3} = \frac{6}{3} = 2$$

$$Z_{\text{men}} = \frac{70 - 69}{3} = \frac{1}{3} = 0.3$$

- Among women, 70 inches is two SDs above the mean (2.2 unrounded), a tall height; among men it is 0.3 SDs above the mean, an ordinary one.
- The same height has two z-scores because a z-score is defined relative to one distribution, and the choice of distribution is part of the question.

Part 1c-d · Reading a z-score

$$\text{c. } z = \frac{61 - 64}{3} = \frac{-3}{3} = -1$$

$$\text{d. } z = 0 \iff Y = \bar{Y}$$

- **c.** $z = -1$: this woman is one SD below the mean.
- **d.** The sign gives the side of the mean; the size gives the distance from it in SDs. The mean itself has a z-score of 0.

Today

z-scores and the normal distribution

1

How many SDs from the mean?

2

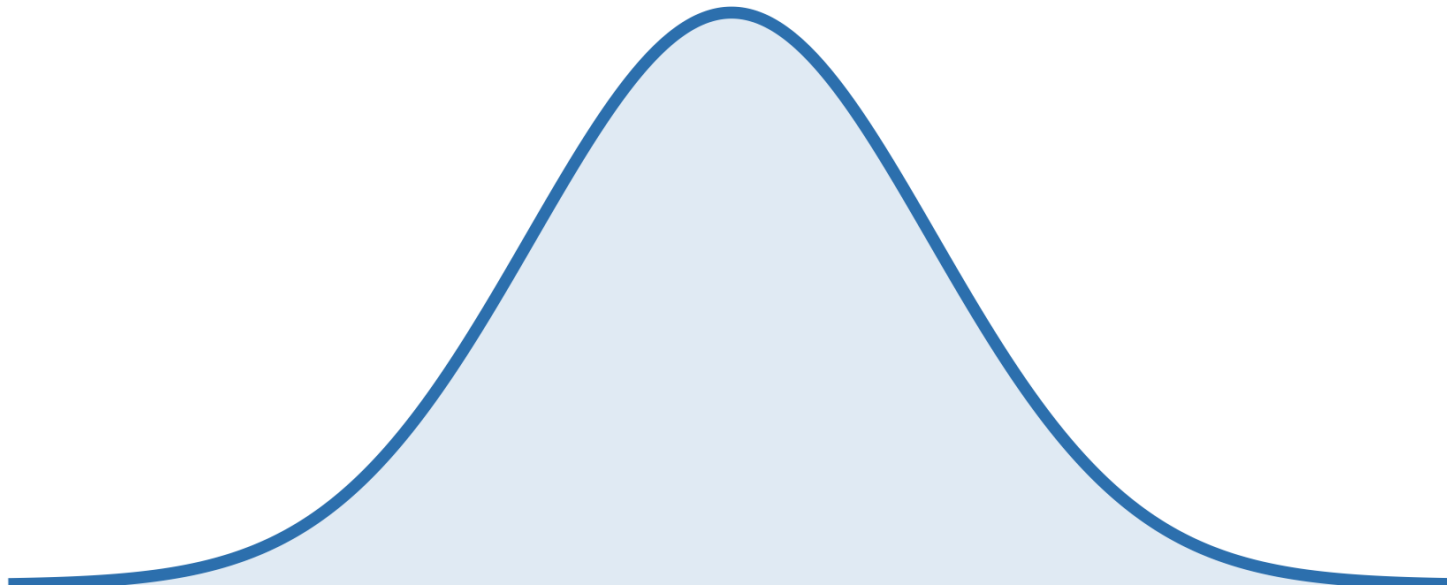
The normal distribution

3

Does the curve fit?

A z-score states how far one value is from the mean, in SDs. The next step is a model for the whole distribution, specified by the same two quantities, the mean and the SD.

Where does the normal distribution come from?

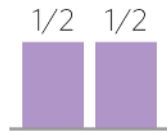
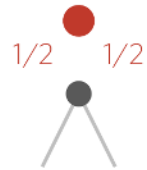


Histograms of many measured variables, heights among them, have roughly this shape. The next few slides build the shape from a simple random process.

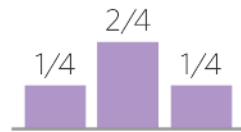
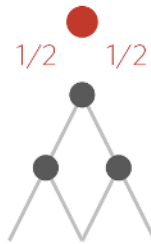
Where will the balls land?

At each peg, a ball bounces left or right with equal probability. With one row of pegs, which bin does it end up in? With two rows? With three?

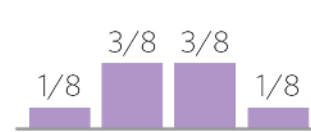
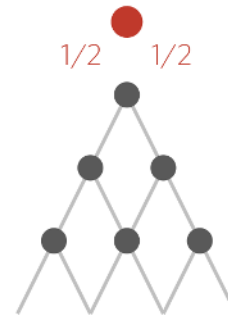
one row of pegs



two rows

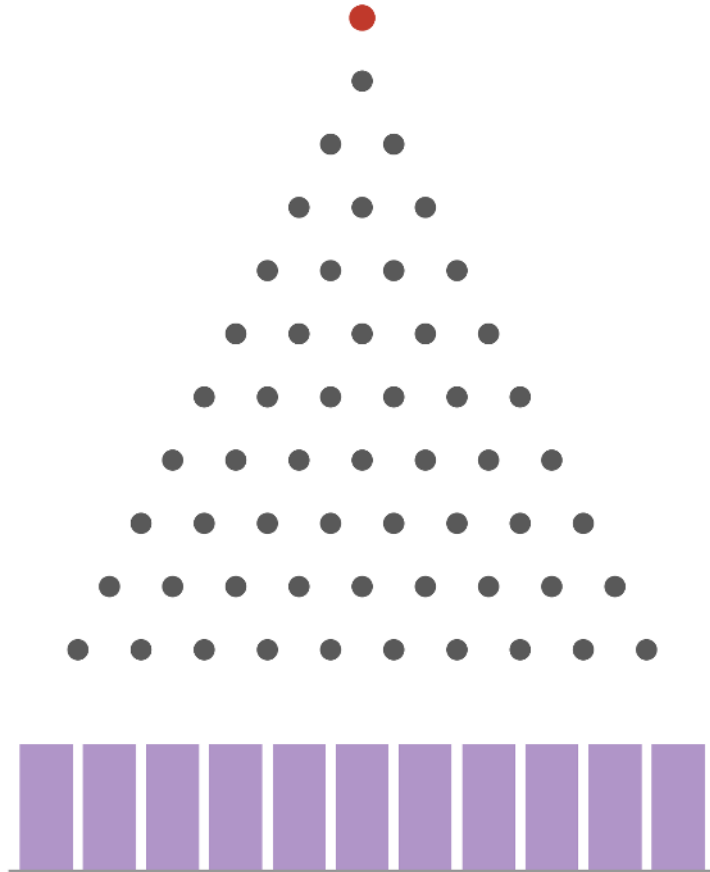


three rows



With your neighbor: a board has ten rows of pegs and eleven bins, and a thousand balls drop through it. Sketch the pile at the bottom.

Is this what the pile will look like?



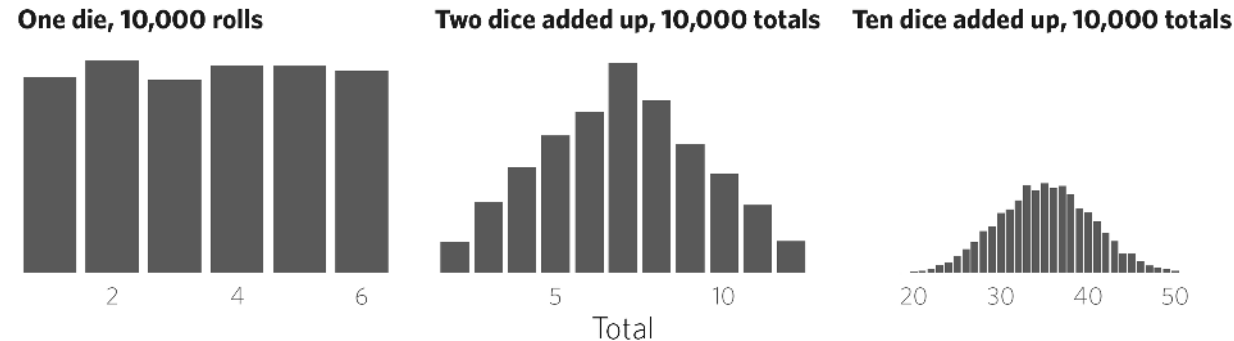
One guess: a ball can end in any of the eleven bins, so after a thousand balls the bins hold about equal numbers.

Is that what you sketched? Hands: flat, highest in the middle, or something else.

Video: *Matemateca* (IME/USP), exhibit by Estes Objethos Atelier, CC BY-SA 4.0, Wikimedia Commons.



Why so many distributions are bell-shaped

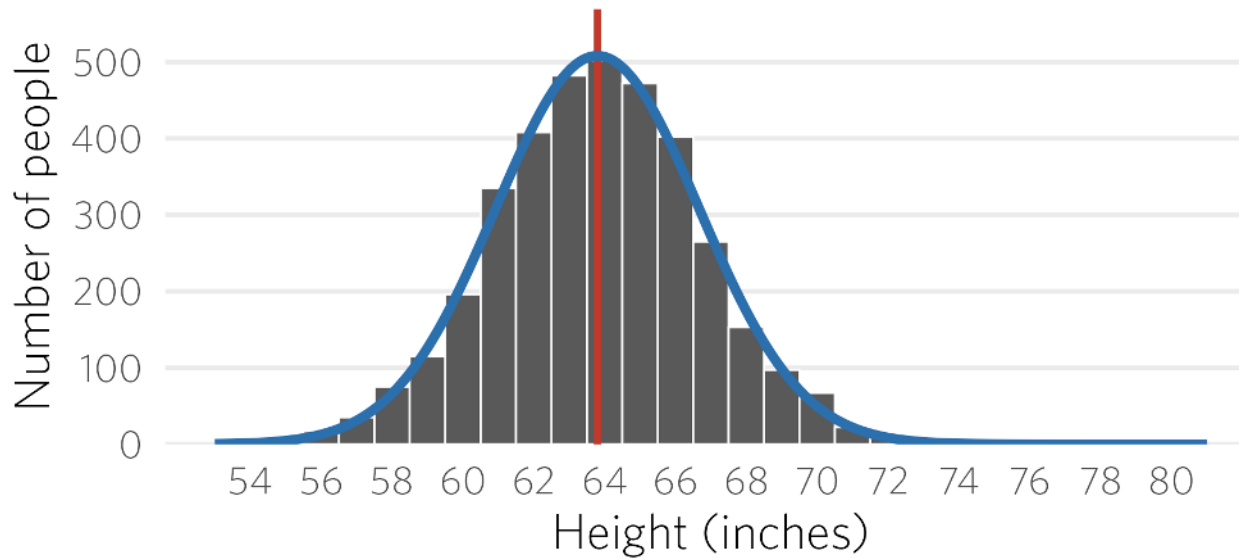


A single die lands on each face with equal frequency. The sum of two dice is concentrated at 7, because more combinations produce 7 than produce 2 or 12. The sum of ten dice is bell-shaped, because high and low rolls largely offset one another.

Height arises in the same way, from many small genetic and nutritional influences that largely offset one another. Many measured variables are sums of this kind, which is why so many are approximately bell-shaped and why the normal curve is the usual first model.

A model for the whole distribution

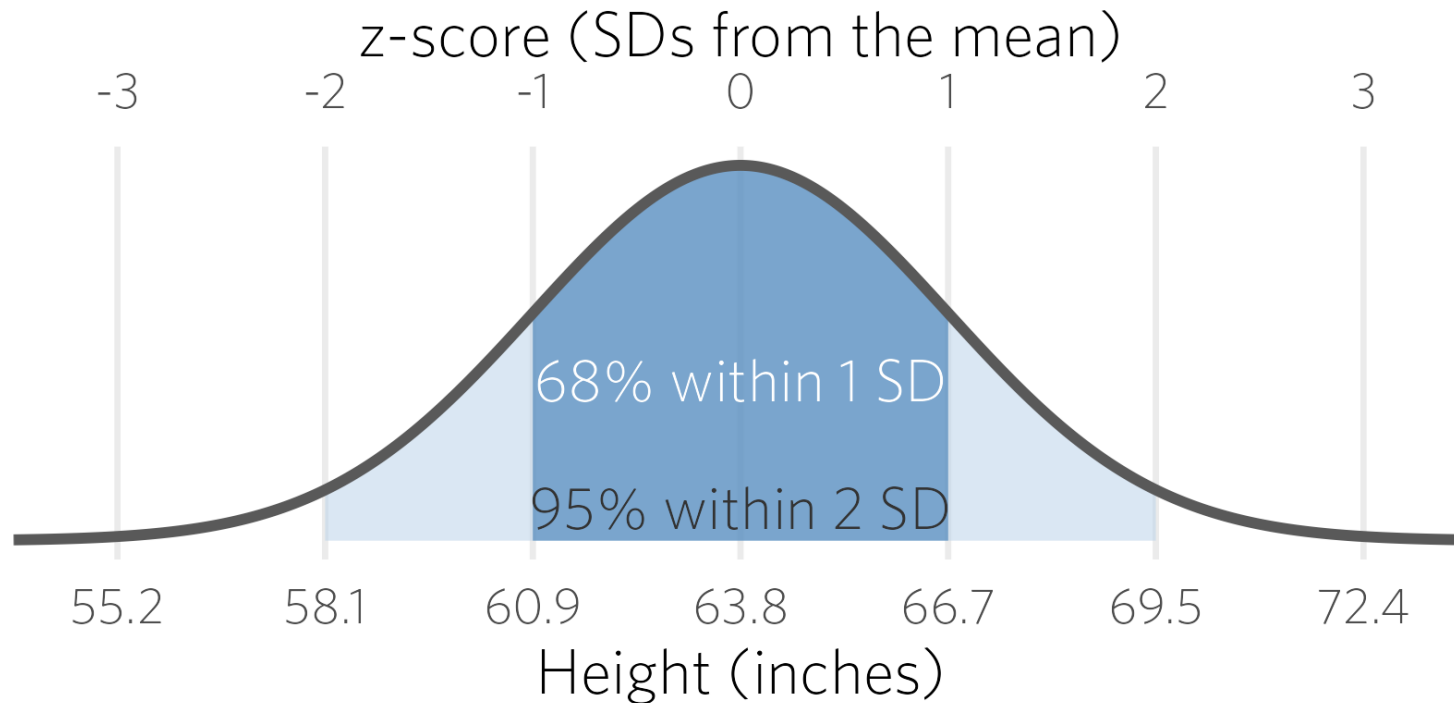
3,658 women in the NHANES survey: mean 63.8 inches, SD 2.9 inches;



The blue curve is the **normal distribution** with the same mean and SD as the data. It is a model in last Wednesday's sense: the curve stands for the distribution, and the departures of the histogram from the curve are the error.

Areas under the normal curve

The normal curve for women's heights: mean 63.8, SD 2.9



Under the normal curve, 68% of the area lies within 1 SD of the mean, 95% within 2 SDs, and 99.7% within 3 SDs (CourseKata's Zones 1, 2, and 3).

What counts as unlikely?

- Under the normal curve, about 5% of the area lies more than 2 SDs from the mean, 2.5% in each tail.
- By a common convention, a value more than 2 SDs from the mean is treated as unusual. The convention is a choice: the curve gives a probability, and whether that probability counts as small depends on the purpose at hand.
- The 5% threshold is widely used in psychology and widely debated; it is one reasonable cutoff among several, not a property of the distribution.

Worksheet - Part 2

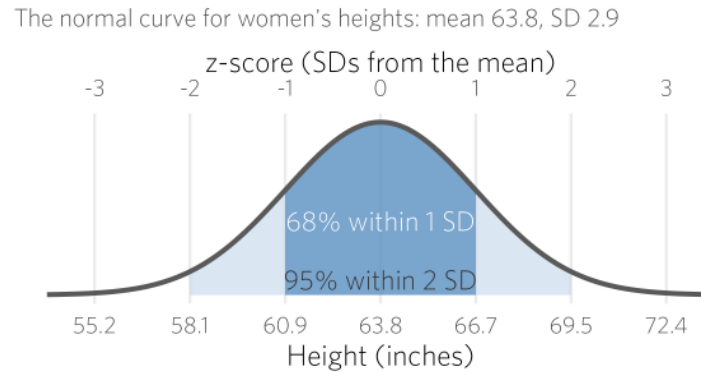
The curve over the histogram is the **normal curve** with the same mean and SD as the data. Under the normal curve, 68% of the area is within 1 SD of the mean, 95% within 2 SDs, and 99.7% within 3 SDs.

- a.** On the figure, shade the region within 1 SD of the mean, from 61 to 67 inches, and write the curve's percentage inside the shaded region.
- b.** According to the curve, about what percentage of women are taller than 67 inches?

Worksheet - Part 2, continued

- c.** Taylor Swift's height, 70 inches, is 2 SDs above the mean. According to the normal curve, about what percentage of women are that tall or taller?
- d.** In the survey, 71% of the women are between 61 and 67 inches. Is that close to the curve's figure for the same range?

Part 2a-c · Predictions from the normal curve



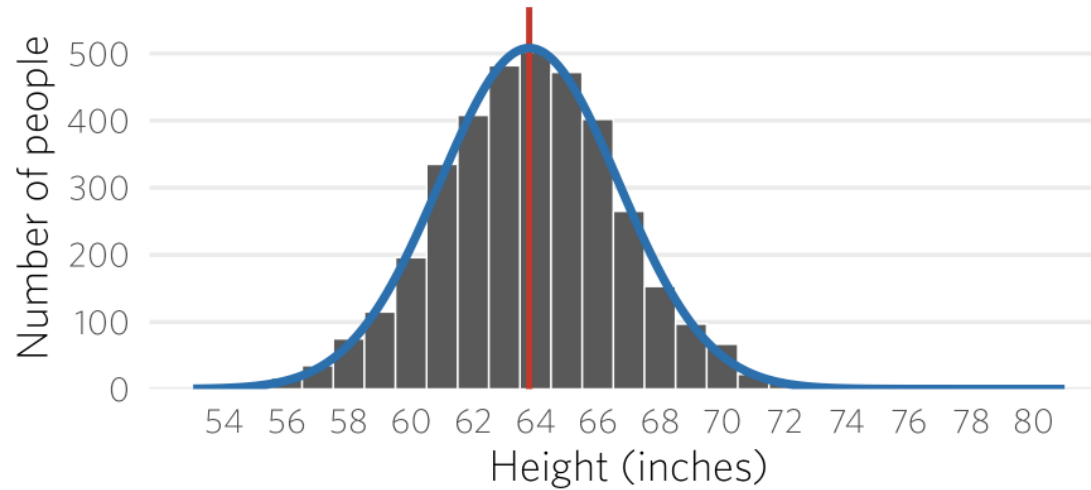
b. $\frac{100 - 68}{2} = 16\%$

c. $\frac{100 - 95}{2} = 2.5\%$

- **a-b.** Under the curve, 68% of women are within one SD of the mean, 61 to 67 inches, so about 16%, half of the remaining 32%, are taller than 67 inches.
- **c.** 70 inches is 2 SDs above the mean; 5% of the area lies beyond 2 SDs, half of it above, so the curve puts about 2.5% of women at 70 inches or taller.

Part 2d - Checked against the survey

3,658 women in the NHANES survey: mean 63.8 inches, SD 2.9 inches;



In the survey, 71% of the women are between 61 and 67 inches, against the curve's 68%, and 1.3% are 70 inches or taller, against 2.5% (1.5% with the unrounded mean and SD). For these heights the normal distribution is a close description of the data.

Today

z-scores and the normal distribution

1

How many SDs from the mean?

2

The normal distribution

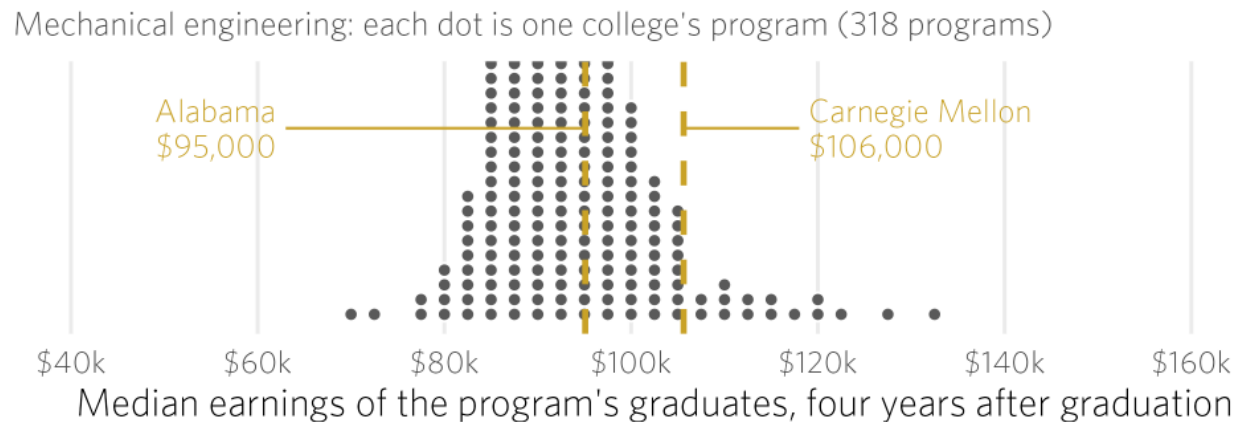
3

Does the curve fit?

For heights, the normal curve came close to the data. The last section asks whether it fits a distribution from last week's news.

What the Times published last Wednesday

A College-by-College Guide to Which Majors Pay Off, or Don't, by Arfa Momin and Ron Lieber, The New York Times, October 7, 2026. For every major at every college in the country, one number: the median earnings of the program's graduates four years after graduation. The charts show one major at a time, each dot one college's program.



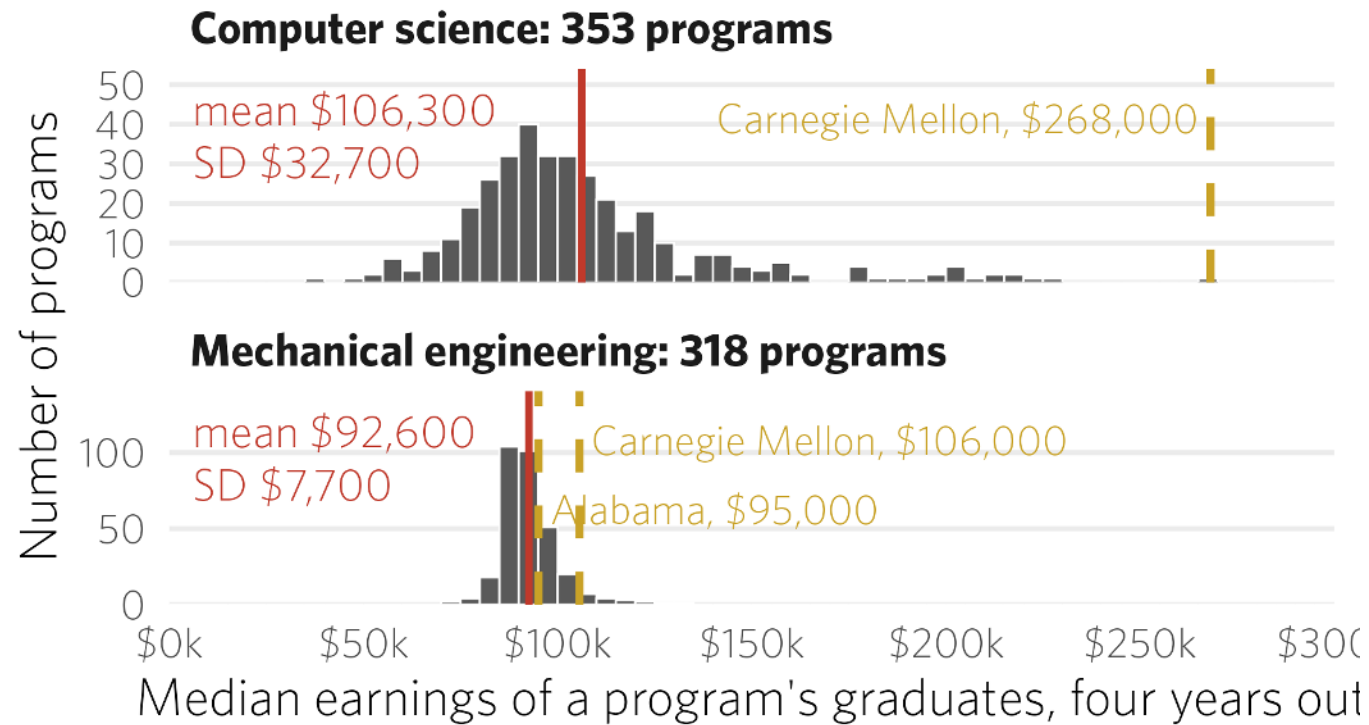
Most mechanical engineering programs lie within about \$10,000 of one another; the Times' example is that a Carnegie Mellon graduate earned roughly \$10,000 more than one from the University of Alabama.

Does which college a student attends matter more for some majors than others?

In mechanical engineering, the college accounts for about \$10,000 of pay. Computer science programs have a similar mean, \$106,300 against \$92,600 for mechanical engineering.

With your neighbor: in which of the two majors does the college you attend matter more for pay? That is, in which major do program earnings vary more from college to college?

Which major's pay varies more?

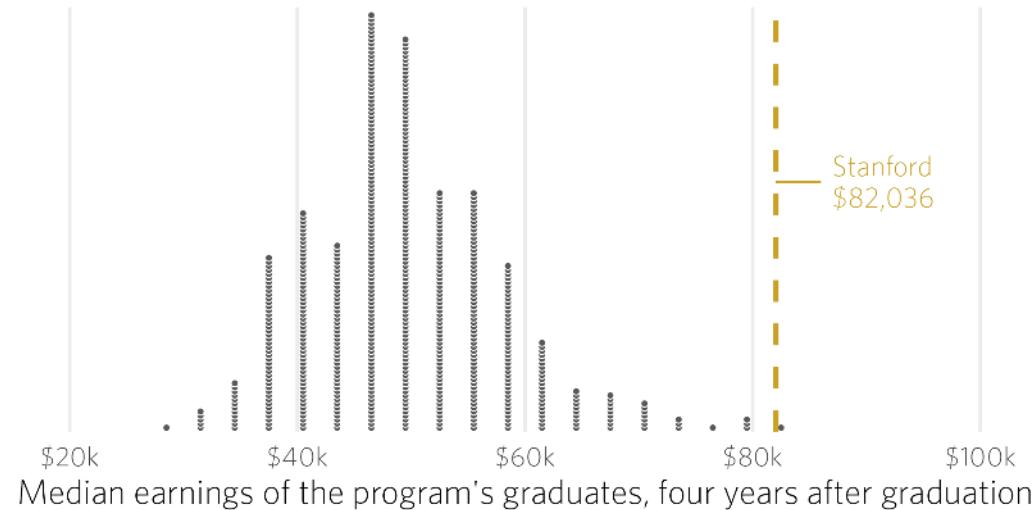


Across programs, the SD is \$32,700 for computer science and \$7,700 for mechanical engineering. The college matters about four times as much in computer science, and each SD states that spread as one number in dollars.

The Times' chart for English

The whole file behind the guide holds 27,247 bachelor's programs, one number each.

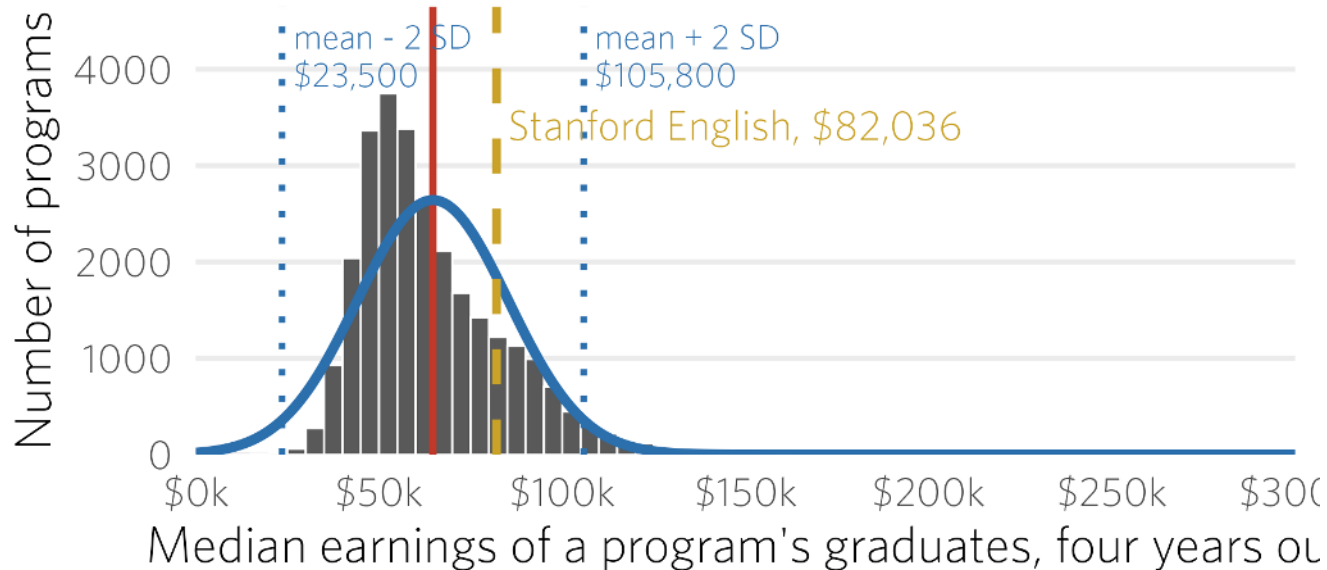
English: each dot is one college's program (576 programs)



English: 576 programs, with Stanford's first at \$82,036. Their mean is about \$50,000 and their SD about \$8,000, so Stanford's is 4 SDs above the mean, a value the normal curve would put at about 3 in 100,000.

Every program, with its normal curve

The normal curve with the same mean and SD as the 27,247 programs



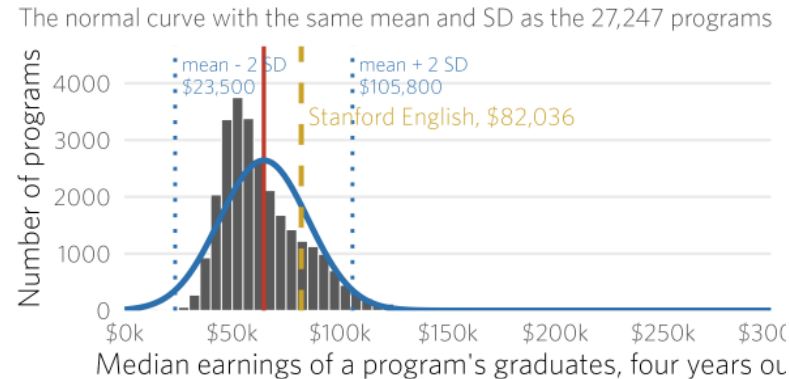
All 27,247 programs, with the normal curve that has their mean, \$64,672, and SD, \$20,571. Part 3 asks how well the curve fits these data.

Worksheet - Part 3

Last Wednesday's New York Times guide listed, for every major at every college, the median earnings of the program's graduates four years after graduation: 27,247 programs in all. The histogram shows all of them, with the normal curve that has the same mean and SD, about \$65,000 and \$20,000.

- a.** For women's heights, the normal curve followed the histogram. Describe one way this histogram differs from its curve.
- b.** The normal curve puts 2.5% of programs above \$105,000 (two SDs above the mean) and 2.5% below \$25,000 (two SDs below). In the data, about 1,100 of the 27,247 programs are above \$105,000 and about 25 are below \$25,000. What percentage of the programs is in each tail, and is each larger or smaller than the curve's 2.5%?
- c.** Across the 27,247 programs, the mean is about \$65,000 and the median about \$60,000. Why are they \$5,000 apart here, when for women's heights the mean and the median are both 64 inches?

Part 3a-b · Does the curve fit?



$$\text{above } \$105,000 : \frac{1,084}{27,247} = 4.0\%$$

$$\text{below } \$25,000 : \frac{27}{27,247} = 0.1\%$$

- **a.** Most programs lie between \$40,000 and \$70,000, with a long right tail (21 programs above \$200,000; Duke mathematics at $z = 11$). The symmetric curve puts programs in a left tail that is nearly empty.
- **b.** The data have 4% of programs more than two SDs above the mean against the curve's 2.5%, and 0.1% below against 2.5%. For heights both tails matched the curve within about a point; here the curve misses in both directions. The first question to ask of a model is whether its shape matches the data.

Part 3c • Mean or median?

- The long right tail pulls the mean up while the median stays with the bulk of the programs, as one high salary did to last Wednesday's seven. Women's heights have no such tail, so their mean and median are both 64 inches.
- The Times' search tool reports one number for all programs, \$64,500. That is the mean: unrounded, the mean of the 27,247 programs is \$64,672 and the median is \$59,659.
- One number for 27,247 programs is a model. The shape of the distribution decides whether the mean or the median is the better summary, and here the Times reported the summary most affected by the tail.

Before Friday

- Finish CourseKata **Chapter 6**.
- Friday: **The Forged Average**. Check your discussion section's Slack channel for your time slot.
- Milestone 3 is due Sunday.

Hand in your worksheet on the way out

On the last line, write down one thing you learned today.

Thanks!