DATA 8

Spring 2022

Lecture 26

The Normal Distribution

Announcements

Review: Standard Units

- How many SDs above average?
- ***z* = (value - average)/SD**
  - Negative z: value below average
  - Positive z: value above average
  - z = 0: value equal to average
- When values are in standard units: average = 0, SD = 1
- Gives us a way to compare/understand data no matter what the original units

The SD and the Histogram

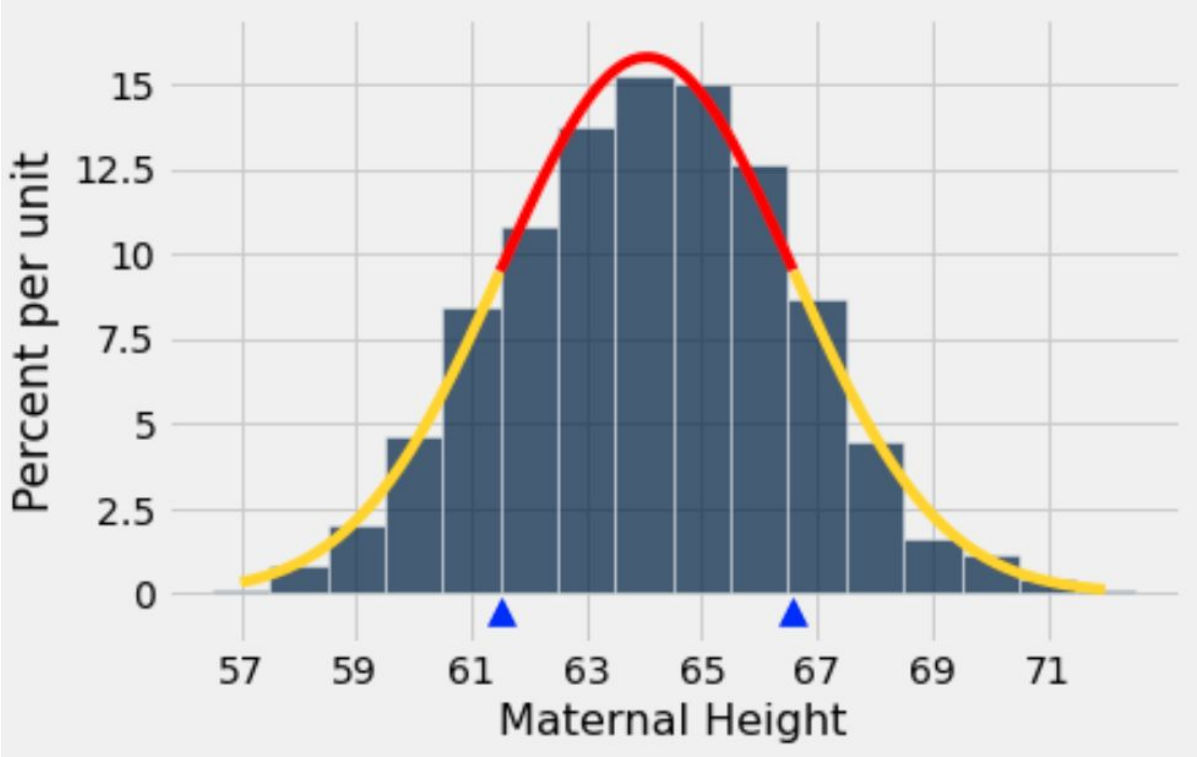
- Usually, it's not easy to estimate the SD by looking at a histogram.
- But if the histogram has a bell shape, then you can.

The SD and Bell-Shaped Curves

If a histogram is bell-shaped, then

- the average is at the center
- the SD is the distance between the average and the points of inflection on either side

Points of Inflection



(Demo)

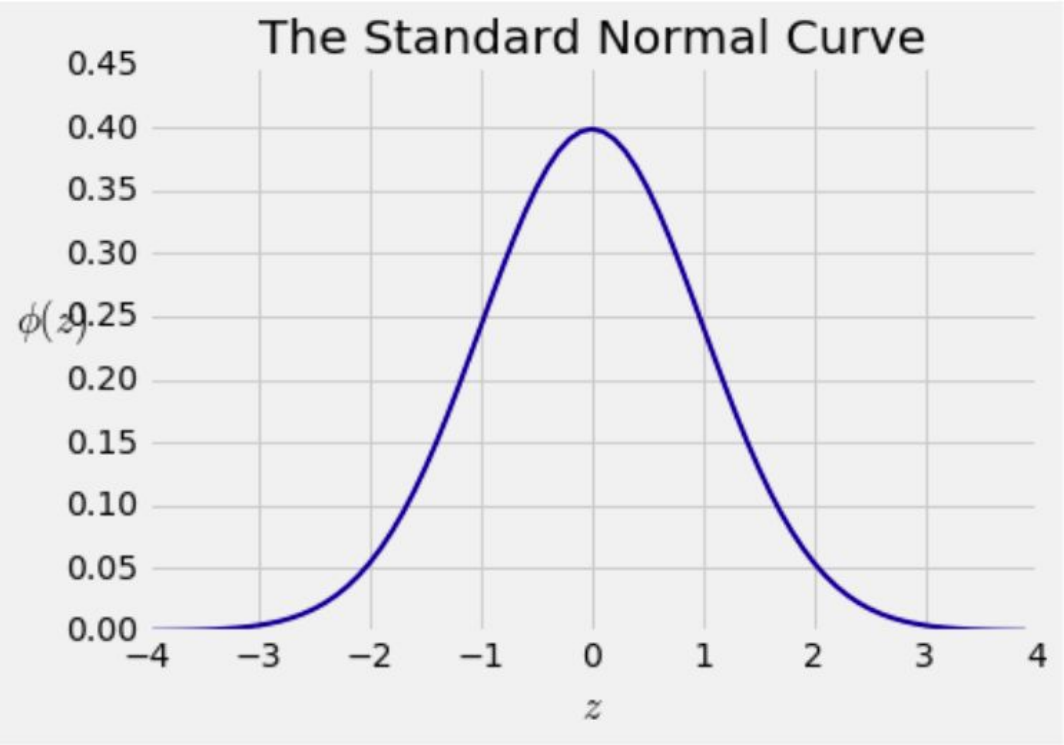
The Normal Distribution

The Standard Normal Curve

A beautiful formula that we won't use at all:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \qquad -\infty < z < \infty$$

Bell Curve


The Standard Normal Curve
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$\phi(z)$
−4
−3
−2
−1
0
1
2
3
4
z

Normal Proportions

How Big are Most of the Values?

No matter what the shape of the distribution,
the bulk of the data are in the range “average ± a few SDs”

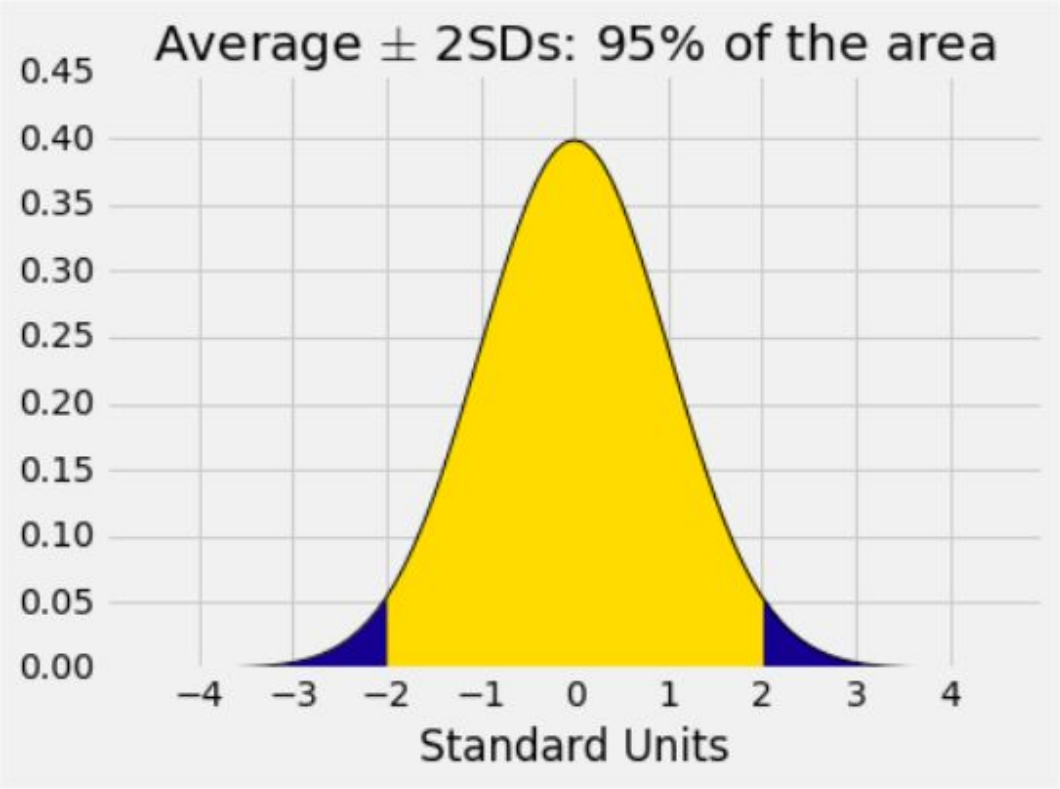
If a histogram is bell-shaped, then

- Almost all of the data are in the range “average ± 3 SDs”

Bounds and Normal Approximations

Percent in Range	All Distributions	Normal Distribution
average $\pm$ 1 SD	at least 0%	about 68%
average $\pm$ 2 SDs	at least 75%	about 95%
average $\pm$ 3 SDs	at least 88.888...%	about 99.73%

A “Central” Area



Central Limit Theorem

Sample Averages

- The Central Limit Theorem describes how the normal distribution (a bell-shaped curve) is connected to random sample averages.
- We care about sample averages because they estimate population averages.

(Demo)

Central Limit Theorem

If the sample is

- large, and
- drawn at random with replacement,

Then, *regardless of the distribution of the population,*

the probability distribution of the sample average
is roughly normal

Discussion Question

After rolling 1,000,000 fair 6-sided dice, which of these histograms would you expect to have a bell shape? Check all that apply.

1) The histogram of outcomes of these million rolls
2) The histogram that results from computing the average outcome of these million rolls
3) The histogram that results from splitting the outcomes into 1,000 groups of 1,000 (in the order they occurred) and computing the average outcome of each group