Lesson 12: What are some characteristics of a normal distribution? What does the empirical rule tell you about data spread around the mean?

Q: Does standardizing into z-scores change the shape of the distribution?

$$z = \frac{(y - \bar{y})}{s}$$

Think in terms of shifting and rescaling.

Standardizing into z-scores does not change the shape of the distribution.

Standardizing into z-scores changes the center by making the mean 0.

Standardizing into z-scores changes the spread by making the standard deviation 1.

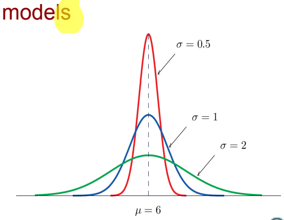
When is a z-score BIG?

- How far from 0 does a z-score have to be to be interesting or unusual?
- There is no universal standard, but the larger a z-score is (negative or positive), the more unusual it is.
- Often we consider a z-score greater than 2 or less than -2 to be roughly an indication of unusualness. But every data set is different, so be careful!

z-score of 3 is unusual ....7 more so

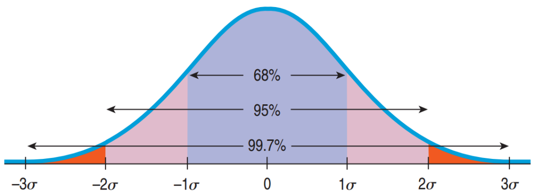
The normal models

"All models are wrong—but some are useful."
—George Box, famous statistician

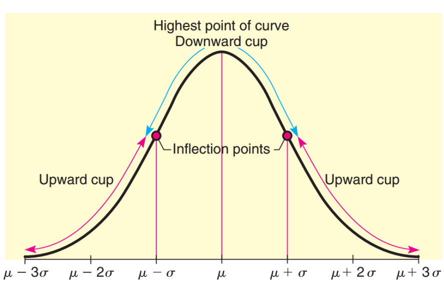


- There is a Normal model for every possible combination of mean and standard deviation. These models are appropriate for unimodal and symmetric distributions.
- We write N(μ,σ) to represent a Normal model with a mean of μ and a standard deviation of σ.
- We use Greek letters because this mean and standard deviation are not numerical summaries of the data. They are part of the model. They don't come from the data. They are numbers that we choose to help specify the model. Such numbers are called parameters of the model.

The 68–95–99.7 Rule



Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean.



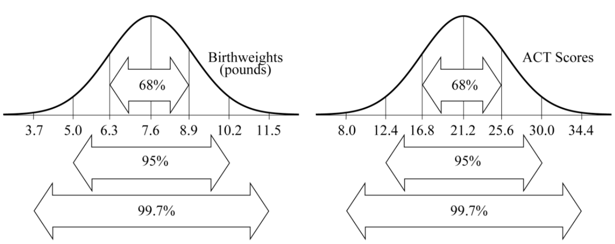
A Normal Curve

Important properties of a normal curve

1. The curve is bell-shaped, with the highest point over the mean μ.
2. The curve is symmetrical about a vertical line through μ.
3. The curve approaches the horizontal axis but never touches or crosses it.
4. The inflection (transition) points between cupping upward and downward occur above $\mu + \sigma$ and $\mu - \sigma$.
5. The area under the entire curve is 1.

Sketch Normal models using the 68-95-99.7 Rule:

- Birth weights of babies, N(7.6 lb, 1.3 lb)
- ACT scores at a certain college, N(21.2, 4.4)



Dutch are known to be tall

As a group, the Dutch are among the tallest people in the world. The average Dutch man is 184 cm tall—just over 6 feet (and the average Dutch woman is 170.8 cm tall—just over 5'7"). If a Normal model is appropriate and the standard deviation for men is about 8 cm, what percentage of all Dutch men will be over 2 meters (6'6") tall?

Draw the normal model

Nearly Normal Condition: The shape of the data's distribution is unimodal and symmetric. This condition can be checked with a histogram.

$$z = \frac{y - \mu}{\sigma}$$

Once we have standardized, we need only one model:

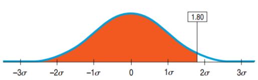
The N(0,1) model is called the standard Normal model (or the standard Normal distribution).

Three rules for working with normal models

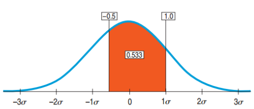
1. Make a picture.
2. Make a picture.
3. Make a picture.

actually a sketch is all you need

Finding normal percentiles with a calculator



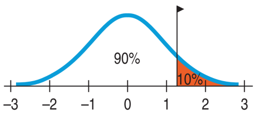
normalcdf(-5,1.80) ← why!!

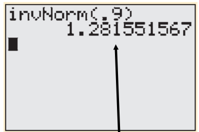


normalCdf(-.5,1)

normalcdf(finds the proportion of area under the curve between two *z*-score cut points, by specifying normalcdf(zLeft,zRight)

From Percentiles to Scores: z in Reverse

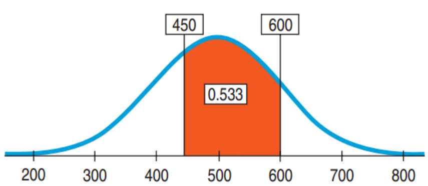




Since we want the cut point for the *highest* 10%, we know that the other 90% must be *below* that z-score.

The Normal model is our first model for data. It's the first in a series of modeling situations where we step away from the data at hand to make more general statements about the world. We'll become more practiced in thinking about and learning the details of models as we progress through the book. To give you some practice in thinking about the Normal model, here are several problems that ask you to find percentiles in detail.

Question: What proportion of SAT scores fall between 450 and 600?



Standardizing the two scores, I find that

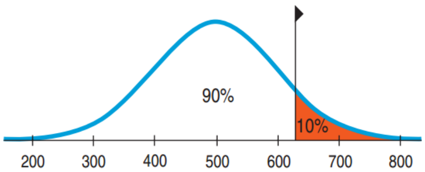
$$z = \frac{(y - \mu)}{\sigma} = \frac{(600 - 500)}{100} = 1.00$$

and

$$z = \frac{(450 - 500)}{100} = -0.50$$

The Normal model estimates that about 53.3% of SAT scores fall between 450 and 600.

Question: Suppose a college says it admits only people with SAT Verbal test scores among the top 10%. How high a score does it take to be eligible?



The cut point is $z = 1.28$.

A z-score of 1.28 is 1.28 standard deviations above the mean. Since the SD is 100, that's 128 SAT points. The cutoff is 128 points above the mean of 500, or 628.