Stats Week - Day 2!

The Normal Curve & Standardization

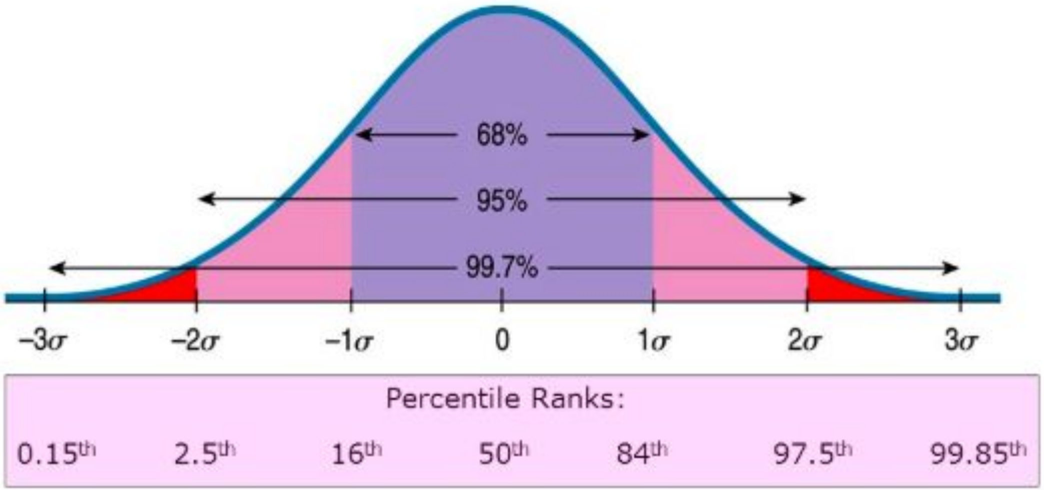
The Normal Curve

The Normal Curve

- You've probably seen "bell curves" before. In Statistics, we call them normal models or normal curves. Normal curves are **always unimodal and roughly symmetric**. They also have other special characteristics.
  - **Parameters:** A normal model is specified by two parameters:
    - a mean, $\boldsymbol{\mu}$, and a standard deviation, $\boldsymbol{\sigma}$
    - notation for the normal model looks like this:
      - ~N ($\boldsymbol{\mu}$, σ)

The Empirical Rule

- The empirical rule is a characteristics of the normal model and is also sometimes referred to as the 68-95-99.7 rule. These numbers relay what percentage of the data lies within one, two, and three standard deviations of the mean, respectively.



Example #1

The SAT Reasoning Test has three parts: Writing, Math, and Critical Reading(Verbal). Each part has a distribution that is roughly unimodal and symmetric and is designed to have an overall mean of about 500 and a standard deviation of 100 for all test takers. In any one year, the mean and standard deviation may differ from these target values by a small amount, but they are good overall approximations.

a. Draw the normal model for any one part of the SAT.
b. In what interval would you expect to find the central 68% of scores?
c. About what percent of students should have scores above 700?
d. About what percent of students should have scores between 600 and 800?
e. About what percent of students should have scores below 300?
f. What is the percentile rank of a student scoring a 400? A 600?

Group Work #1

In groups of 2-3, draw and label the normal model for the given scenarios.

See in-class activity handout.

The Empirical Rule

The Empirical Rule is handy, but what if we're interested in a particular value that doesn't neatly fall at one, two, or three standard deviations from the mean?...

Standardized Scores

Z-Scores

There are multiple types of standardized scores. We'll begin with "z-scores," as they are the standardized score associated with the normal model.

- **A z-score measures the number of standard deviations a data value is away from the mean.**

Group Work #2

In groups of 2-3, consider the two scenarios given and attempt to calculate the z-score for each given value, using the definition given.

See in-class activity handout.

Z-Score Formula

The formula to find a z-score is as follows:

$$z = \frac{x - \mu}{\sigma} \qquad z = \frac{x - \bar{x}}{s}$$

Okay. So what?

Group Work #3

Reference the normal curves you drew and labeled in Group Work #1 (child driving to school & IQ tests). Find the z-scores for the values provided. Then, draw a second normal curve, replacing the original values with the new z-score values.

See in-class activity handout.

Group Work #3

Note: Standardized scores and standard normal distributions essentially allows us to compare apples to oranges (or driving time to IQ scores). That is, comparing two seemingly unrelated observations, we can use their standardized scores to determine which is more unusual.

Standard Normal Distribution

Standard Normal Distribution

- A **standard normal distribution** or a **standard normal model** will *always* have a mean of 0 and a standard deviation of 1.
- Standardizing shifts the distribution up or down, making the mean zero. (This is the numerator of the z-score formula). Then, it scales the spread, making the standard deviation one. (This is the denominator).

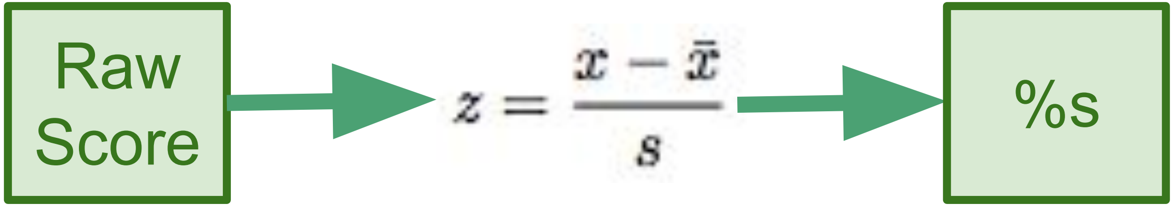
Standard Normal Distribution

- Standardization does not change the shape of the distribution. Further, when we use the standard normal model, we must *start* with a normal model. That is, we should have a unimodal approximately symmetric distribution *before* we standardize scores. Otherwise, we are making unreasonable assumptions.
- When we check to make sure we have a normal distribution, this is called checking **conditions**. Specifically, when we check to make sure we can use a normal model to represent our data, we're checking the **Nearly Normal Condition**. An easy way to check this is to make a histogram or dotplot and look at the shape of the distribution.

Finding Normal Percentiles

Finding Normal Percentiles

- Earlier, we could use the empirical rule to find some percentiles, but only if a value happened to fall neatly at one, two, or three standard deviations from the mean. However, we can find the percentile for *any* value using its z-score and the standard normal distribution.
- To find the percentile, we'll need to use technology or a z-score table.



Please find the z-score table under Day 6 in Git.

Example #2

Find the probability for the following:

1. $z > 1.5$
2. $z < -0.75$

Example #3

A distribution is approximately normal and has a mean of 500 and a standard deviation of 100 (SAT distribution). Use this information to solve the following:

1. What percentage of test-takers score below a 680?
2. What percentage of test-takers score above a 720?
3. What percentage of test-takers score between a 540 and 650?

Group Work #4

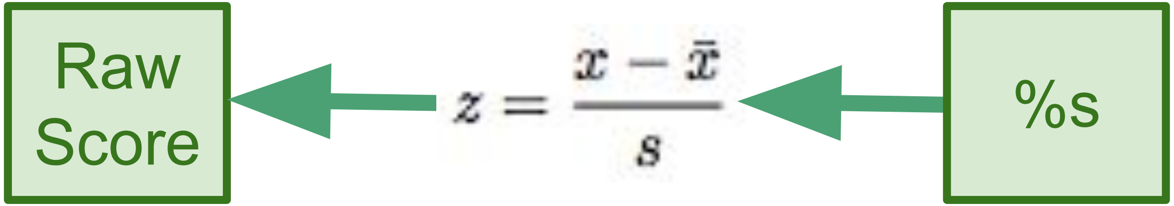
In groups of 2-3, find the percentiles requested using a standard normal distribution. ** Please use both the table and a website, so you are comfortable with both!

See in-class activity handout.

Reversing the Process: Going from Percentiles to Scores

Reversing the Process: Going from Percentiles to Scores

- We can also use the z-table or technology to go from percentages to raw scores.
- To do this, we'll first utilize the table and then solve for the raw score within the z-score formula.



Example #4

Find the raw score for the following:

1. What z-score cuts off the bottom 25% in a Normal model?
2. What z-score cuts off the top 10% in a Normal model?

Example #5

A distribution is approximately normal and has a mean of 500 and a standard deviation of 100 (SAT distribution). Use this information to solve the following:

a. How high of an SAT Verbal score does a student need to be in the top 10% of test takers?
b. What SAT Verbal score does a student have if she is at 20th percentile of test-takers?
c. What range of SAT Verbal scores can a student have to fall within the middle 50% of test-takers?

Group Work #5

In groups of 2-3, find the z-scores and/or raw scores that satisfy the given scenarios. Again, please use the z-tables AND technology so you are comfortable with both.

See in-class activity handout.

Student's or Gosset's *t*

Student’s or Gosset’s *t*

- Tomorrow, we'll consider when the normal model fails to well-represent unimodal, approximately symmetric data distributions. For now, it's enough to know that another family of distributions, called *t* distributions exist. They are particularly helpful when working with small sample sizes and/or when we don't know the population standard deviation (which is often!).
- There are different tables we'll need to use when working with the family of *t* distributions.