Percentiles & Empirical Rule

Unit 5

Another way to compare values of different variables - like your ACT and SAT math scores - by determining it's RELATIVE value to other values of the same variable are ***percentiles.***

Measuring Position: Percentiles

One way to describe the location of a value in a distribution is to tell what percent of observations are less than it.

Definition:

The *p*th percentile of a distribution is the value with *p* percent of the observations less than it.

Jenny earned a score of 86 on her test. How did she perform relative to the rest of the class?

Calculate both her percentile and her z-score.

6	7
7	2334
7	5777899
8	00123334
8	569
9	03

The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2009.

5	9
6	2 4 5 5
7	0 0 4 5 5 5 8 9
8	0 3 4 5 6 6 7 7 7 8
9	1 2 3 5 5 7
10	3

Key: 5|9 represents a team with 59 wins.

Problem: Find the percentiles for the following teams:

(a) The Colorado Rockies, who won 92 games.

(b) The New York Yankees, who won 103 games.

(c) The Kansas City Royals and Cleveland Indians, who both won 65 games.

Suppose that Mitchell took the SAT last year and knows he scored in the 79th percentile. He also knows that approximately 1.66 million students took the SAT. How many students did he score better than?

Can you determine his score?

Normal Distributions

A class of distributions whose density curves are symmetric, uni-modal, and bell-shaped.

Normal distributions are VERY important in statistics.

Which numerical summary would we use to describe the center and spread of a Normal distribution?

Notation:

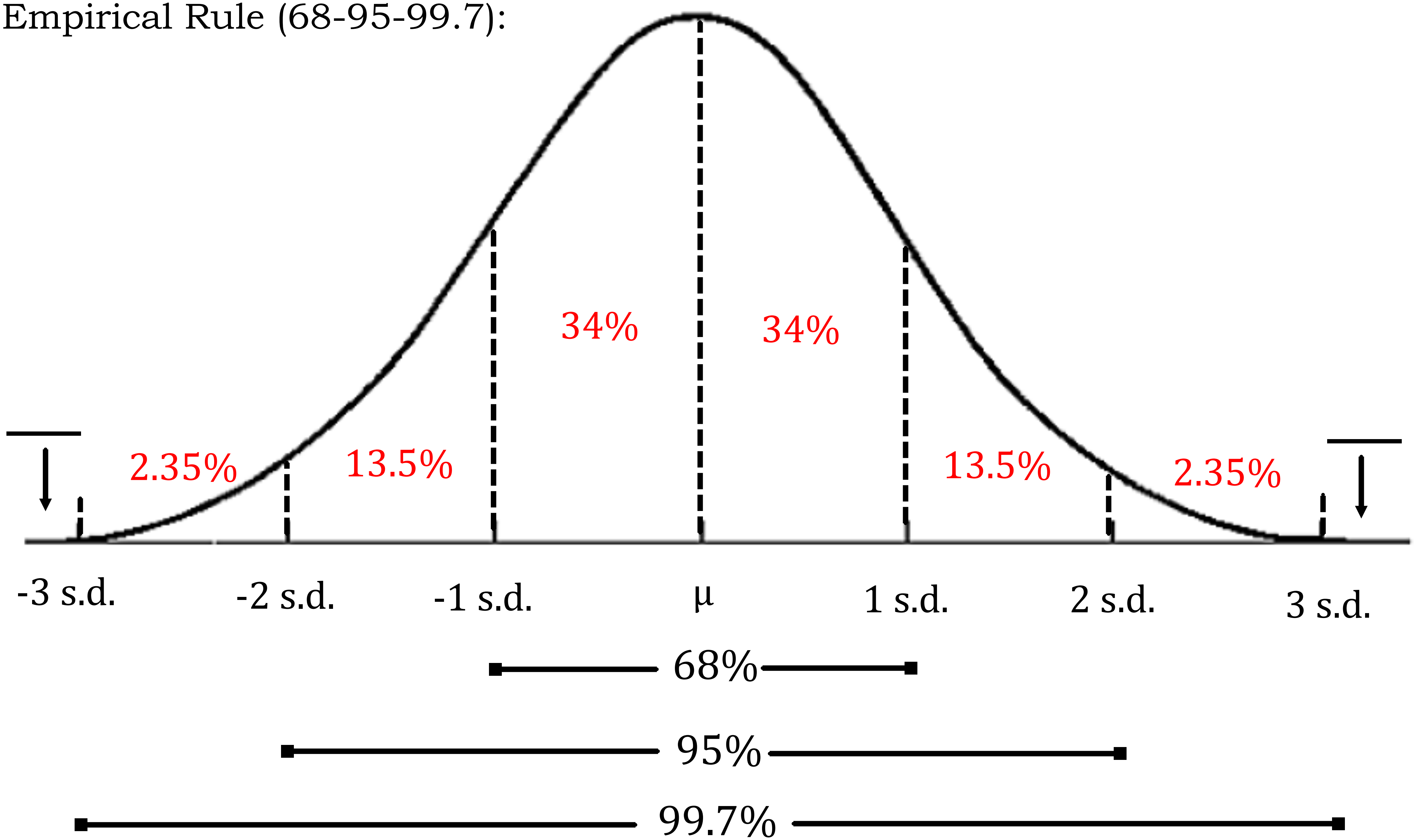
The 68-95-99.7 Rule (The Empirical Rule) - In the Normal distribution with mean μ and standard deviation σ:

- **68%** of all the observations fall within one standard deviation (σ) of the mean μ (in both directions)

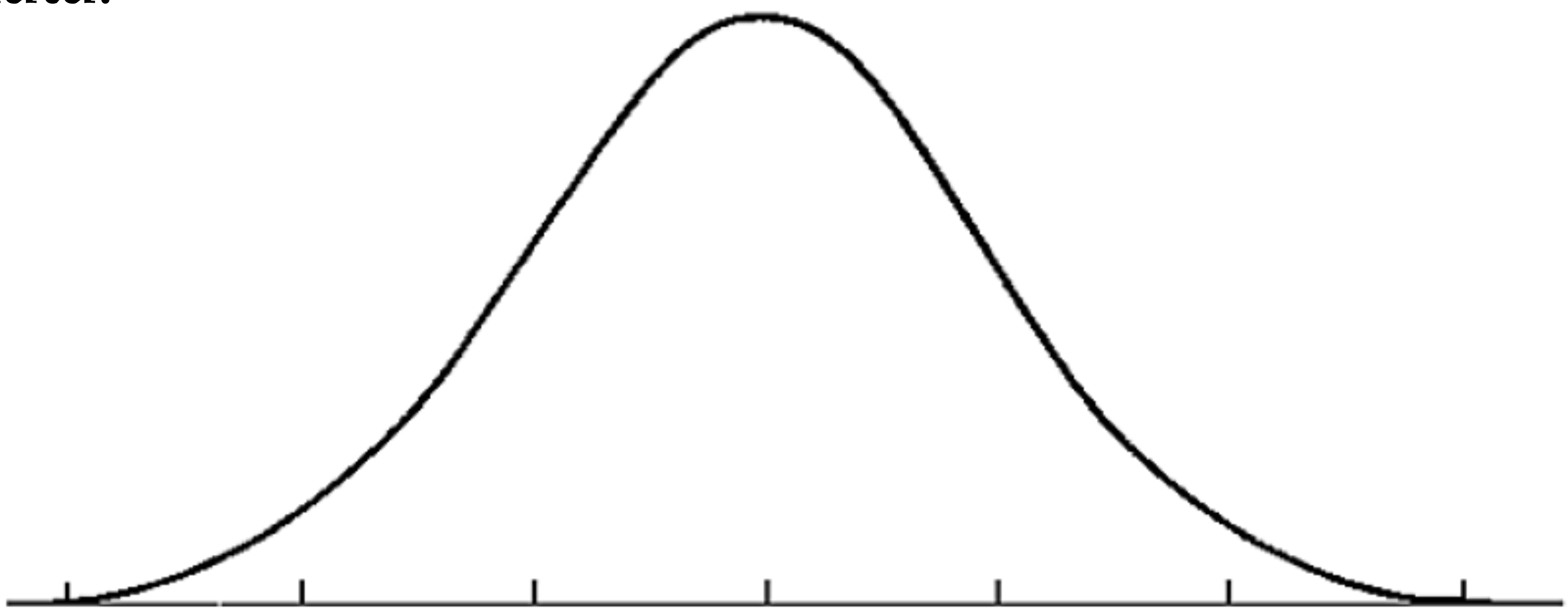
- **95%** of all the observations fall within two standard deviations (2σ) of the mean μ (in both directions)

- **99.7%** of all the observations fall within three standard deviations (3σ) of the mean μ (in both directions)

Empirical Rule (68-95-99.7):



Suppose we have a normal distribution with a mean of 52 and a standard deviation of 6. Draw a model of this data.



The distribution of heights of women aged 20 to 29 is approximately Normal with mean 64 inches and standard deviation 2.7 inches. Use the Empirical rule to answer the following questions.

(a) Between what heights do the middle 95% of young women fall?

(b) What percent of young women are taller than 61.3 inches?

You try:
The length of human pregnancies from conception to birth varies according to a distribution that is approximately Normal with mean 266 days and standard deviation 16 days. Use the Empirical rule to answer the following questions.

(a) Between what values do the lengths of the middle 68% of all pregnancies fall?

(b) How short are the shortest 2.5% of all pregnancies?

(c) What percent of pregnancies are longer than 314 days?

In 2009, 432 Major League Baseball players had a mean batting average of 0.261 with a standard deviation of 0.034. Suppose that the distribution is exactly Normal with $\mu = 0.261$ and $\sigma = 0.034$.

Problem:

(a) Sketch a Normal density curve for this distribution of batting averages. Label the points that are 1, 2, and 3 standard deviations from the mean.

(b) What percent of the batting averages are above 0.329? Show your work.

(c) What percent of the batting averages are between 0.227 and .295? Show your work.

Can we use the Empirical rule to determine if a set of data is approximately Normal?

Use the data on median household income for the 50 states in 2009-2010 and 2011-2012 to determine if these household incomes are normally distributed for either period of time.