Introduction to the R Statistical Environment

Sarah Hunter

sghunter@umich.edu

ICPSR Day 4

Overview

1. Review

2. Descriptive Statistics: A Review

3. Making Figures in R

Loading Data into R

1. Identify your file format
2. Set your working directory
3. Install/Load package needed to load data (if applicable)
4. Load the data into R, creating an object
5. Check to make sure the data were loaded correctly

Important Tidyverse Commands

- %>%
- select
- filter
- mutate
- pivot

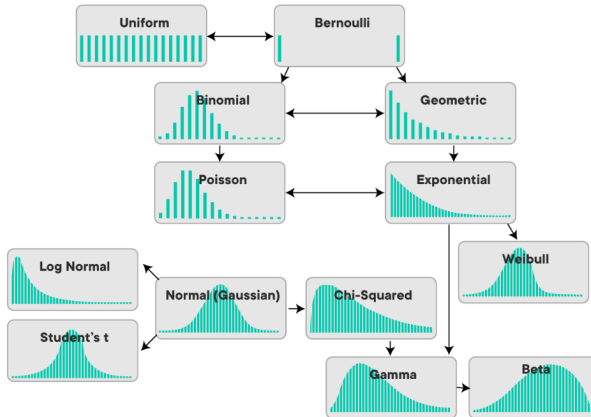Today's Topic: Descriptive Stats and Base Plots in R

Statistical Moments

Defined: "as summary statistic of a distribution" (p. 86)

Most commonly used moments:

1. Central Tendency (i.e. the mean)
2. Variation (i.e. standard deviation/variance)
3. Skewness
4. Kurtosis

Distributions

Distributions can show us a snapshot of the data:

Measures of Central Tendency

Defined

Where the data tend to be; the expected value of the variable

Mean

$$\bar{y} = \frac{\sum y_i}{n}$$

Median

The observation in the middle of an ordered sample

Mode

The most frequent value

Variation

Defined

The spread of the data; how much the data tend to stray from the center

Standard Deviation

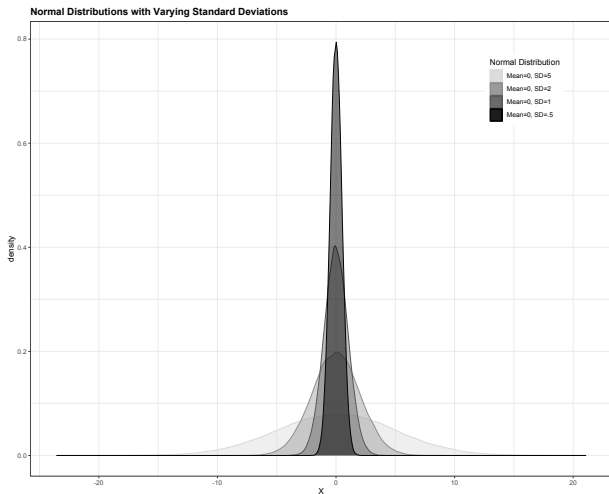$$s = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}$$

Variance

$$s^2 = \frac{\sum(y_i - \bar{y})^2}{n-1}$$

The Range

Defined

The difference between the largest and smallest observations; how much total variation is possible.
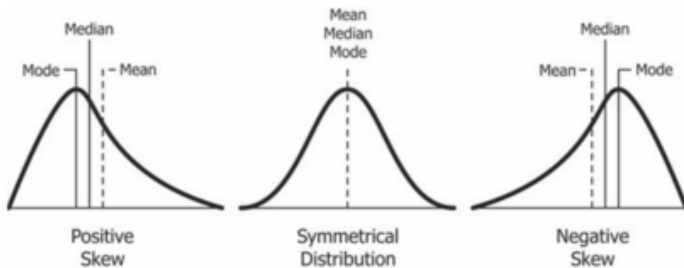
Why Descriptive Statistics Matter

Skewness

Defined

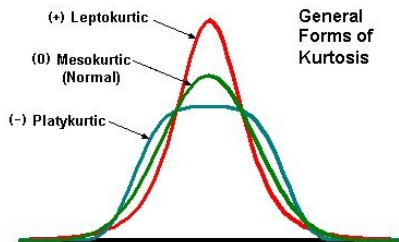The symmetry of the distribution in question; measure of the shape of data

$skew = \frac{mean - median}{standard deviation}$

Kurtosis

Defined

How tall/fat the distribution; a measure of the shape of data

Why Moments Matter

Mathematical models require use to make assumptions about the data.

- Moments can help us understand when the data do, or do not, follow those assumptions.
- Moments can also help us know how to address those assumption violations.

Useful R Commands

To Visually Inspect Data:

- plot(density(x))
- plot(density(na.omit(x)))
- plot(density(x, na.rm=TRUE))

To get descriptive statistics:

- summary(data)
- mean(x)
- median(x)
- sd(x)

From the "Moments" library

- skewness(x)
- kurtosis(x)

Some Guidelines for Effective Figures in R

1. Clean data make clean figures
2. Always add a descriptive title/labels
3. Plots should be easily read and understood
4. When using color, remember that some people cannot see well in color.
5. Don't do too much on one figure

And Now...

To R!

Until Next Time