# Session 2 Presentation
# Statistical Methods in Research
# COSC 6323
# Spring 2018

Dinesh Majeti
Ioannis Thomas Pavlidis

Computational Physiology Lab

*dmajeti@uh.edu*
*ipavlidis@uh.edu*

January 26, 2018

# Overview

1. Variables
2. Location
   - Sample Mean
   - Sample Median
3. Measures of Dispersion
   - Range
   - Standard Deviation
   - Variance
   - Relative Standard Deviation
4. Degrees of Freedom
5. Confidence Interval
6. Plots
   - Histogram
   - Boxplot

# Variables

- **Integral / Quantitative variables** yield numerical measurements.
- **Categorical / Qualitative variables** yield non-numeric information.

| | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb | color |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mazda RX4 | 21.0 | 6 | 160 | 110 | 3.90 | 2.620 | 16.46 | 0 | 1 | 4 | 4 | Silver |
| Mazda RX4 Wag | 21.0 | 6 | 160 | 110 | 3.90 | 2.875 | 17.02 | 0 | 1 | 4 | 4 | Blue |
| Datsun 710 | 22.8 | 4 | 108 | 93 | 3.85 | 2.320 | 18.61 | 1 | 1 | 4 | 1 | Yellow |
| Hornet 4 Drive | 21.4 | 6 | 258 | 110 | 3.08 | 3.215 | 19.44 | 1 | 0 | 3 | 1 | White |

Figure 1: Sample cars dataset

# Sample Mean

- Mean of the sample from the population

$$\bar{x} = \frac{\sum x}{n} \tag{1}$$

where

$\bar{x}$ is the sample mean,

$\sum x$ is the total of the observations, and

n is the sample size (or number of observations)

# Sample median

- The **sample median** is the middle observation when all the observations are placed in order of their magnitute

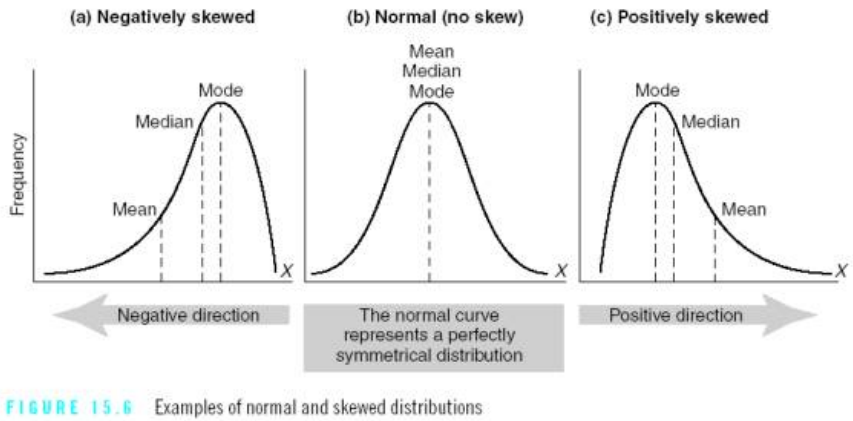


FIGURE 15.6 Examples of normal and skewed distributions

Figure 2: Normal and Skewed distributions

# Measures of dispersion

- The **range** is the difference between the largest and the smallest observations
- The **standard deviation** is the root mean square deviation from the mean.
- The **variance** is the square of the standard deviation.

# Relative Standard Deviation

- The **relative standard deviation (RSD)** represents the standard deviation expressed as a percentage of the mean.

# Degrees of freedom

- Degrees of freedom is the number of observations which can be varied independently under a constraint.

# Confidence Interval

- Confidence interval is an interval estimate of a population parameter.
- The confidence level indicates the probability that this range captures the true population parameter.
- A **confidence interval for the population mean** ($\mu$) is given by

$$\bar{x} \pm \frac{ts}{\sqrt{n}} \tag{2}$$

where

$\bar{x}$ is the sample mean,

s is the standard deviation,

n is the sample size (or number of observations), and

t is obtained using mathematical theory

# Histogram

- Enables us to see the shape of the distribution and make appropriate inferences

# Histogram

```
hist(mtcars$mpg, xlab="Miles per gallon",
 main="Histogram of mileage of cars")
```

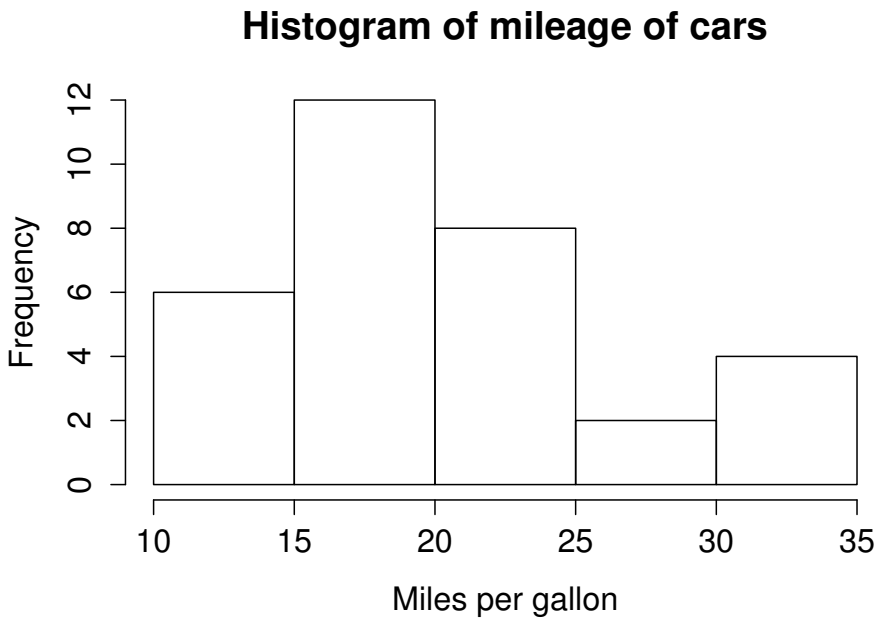


Figure 3: Sample histogram in R

# Box Plot

- Helpful to look at trends in several data sets
- Gives a clear graphical display of the spread of the data
- Features
  - the median
  - quartiles
  - middle 50% of the population
  - interquartile range
  - whiskers
  - outliers

# Boxplot

```
x <- c(1,2,3,3,4,5,5,7,9,9,25)
boxplot(x, xlab="x label", ylab="y label")
title("Sample Boxplot")
```

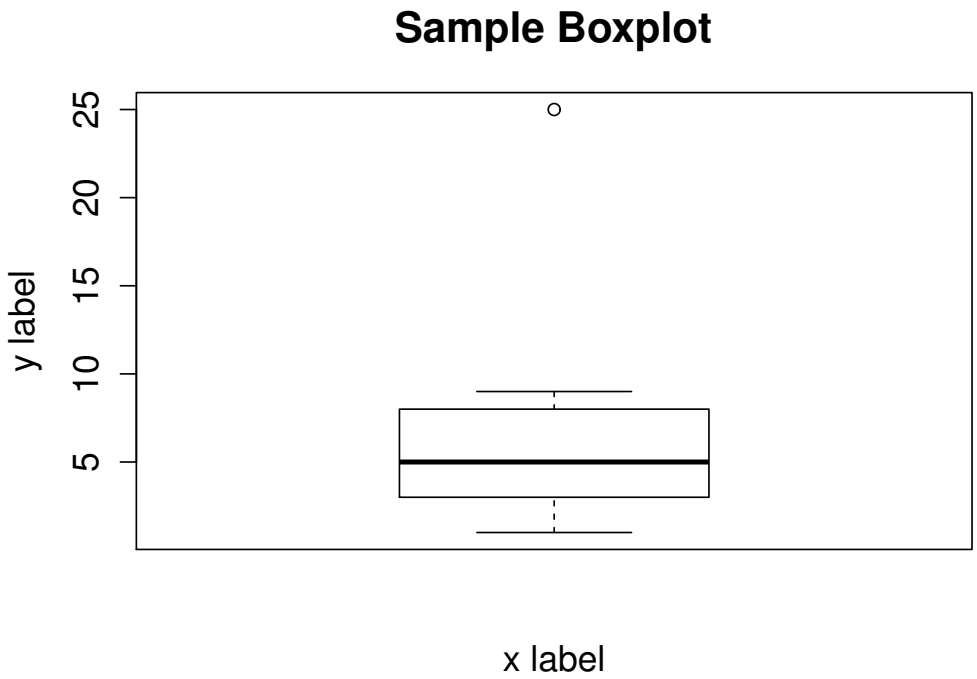


Figure 4: Sample boxplot in R