STAT 3011 Discussion 015

Introduction to R: Week 3

Talha Hamza
University of Minnesota

Spring 2025

Setting Working Directory and Importing Data

- **Setting Working Directory**:
  - Select "Session" then "Set Working Directory" then "Choose Directory."
- **Checking Working Directory**:
  - Use `getwd()` to check the current directory.
- **Importing Data:**
  - `movie1 <- read.csv("movies2009.csv")`
  - `movie2 <- read.csv(file.choose(), header = TRUE)`
  - `movie3 <- read.csv("http://users.stat.umn.edu/ parky/movies2009.csv")`

Exploring the Data Set

Basic Commands:

```
# View dataset
View(movie1)

# Check structure
str(movie1)

# Display first and last rows
head(movie1)
tail(movie1)

# Frequency table for categorical variables
table(movie1$Rating)
```

Measures of Centre and Spread

Key Points:

- The mean represents the central tendency, but outliers can distort it.
- The median is another measure of central tendency, which is resistant to outliers
- The standard deviation measures how spread out the data is.
- Data points beyond 2 standard deviations from the mean are generally considered abnormal.

Interpreting Standard Deviation

What It Tells Us:

- A smaller standard deviation means data points are closer to the mean.
- A larger standard deviation means more variability in data.

Formula for Standard Deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \tag{1}$$

Where:

- σ = standard deviation
- x_i = each individual data point
- μ = mean of the data set
- N = total number of data points

Interpreting Standard Deviation

If the data is assumed to be normal:

- 68% of data falls within 1 standard deviation of the mean.
- 95% of data falls within 2 standard deviations of the mean.
- 99.7% of data falls within 3 standard deviations of the mean.

Boxplots and Histograms

When to Use Each:

- **Histograms**: Show the frequency distribution of a variable.
- **Boxplots**: Help visualize median, quartiles, and potential outliers.
- Outliers in boxplots appear as individual points beyond the whiskers.

Questions? Let's Discuss!