Midterm Review Sheet

Descriptive Statistics and Visualization

- Locations (mean, median), spread (standard deviation, variance), skewness
    - Understand, interpret, and compare across groups
    - Understand how mean and median are different in skewed data
- Histograms, Quantile-Quantile plots, box plots
    - Understand and interpret; draw qualitative conclusions about the distribution sampled from

Tests for Normality

- Shapiro-Wilk
    - Null hypothesis: Data follows Normal/ Alternative hypothesis: Data does not follow Normal
    - Interpret the results and draw conclusions regarding normality of a population based on a sample

Location Tests

One-sided and two-sided alternatives

- One sample tests
    - T test when sampled data follows normal
    - Nonparametric test when data is not normal
    - Choose correct test based on sample distribution and draw correct conclusion for mean or median of the population
- Two sample tests
    - T test when both samples are not far from normal
        - Null: two groups have same mean; Alternative: two groups have different means
        - Test equal variance assumption and choose pooled or Satterthwaite t results based on variance conclusion
        - Correctly interpret results in terms of population means
    - Rank Sum test when one or both samples is far from normal
        - Null: two populations come from the same distribution; Alternative: one of populations tends to have larger values
        - Correctly interpret results in terms of population values
    - Check decision tree file for proper test for each situation

Analysis of Variance (ANOVA)

- Assumptions of ANOVA model
- One-way ANOVA; Null: group means are same; Alternative: at least one group has different mean
- Test for equal through Levene's test and adjust using Welch's adjustment (if necessary) in the one-way ANOVA case

- Null hypothesis for Levene's test: equal variance; Alternative hypothesis: at least one group has different variance
- Understand meaning of interaction and interaction plot
- Post-hoc test (multiple comparison or all pairwise test) to get more specified conclusion about group means
- Check Normality assumption through diagnostics plot (check straight line in Q-Q plot)

Balanced Case

- Equal number of observations in each cell
- Choose a best model (model selection) and explain selection process
- Interpret ANOVA tables and significance of model and terms in model
- Interpret R^2
- Determine and interpret significant differences between groups

Unbalanced Case

- Unequal number of observations in cells
- Choose a best model (model selection) and explain selection process
- Interpret ANOVA tables, type I and III sums of squares and significance of model and terms in model
- Interpret R^2
- Determine and interpret significant differences between groups