Analysis of Variance (ANOVA)

BUS 735: Business Decision Making and Research

1

1.1 Goals

Goals

- Specific goals:
  - Learn how to compare means or medians among more than two populations.
- Learning objectives:
  - LO1: Be able to construct and test hypotheses using a variety of bivariate statistical methods to compare characteristics between two populations.
  - LO3: Be able to construct and use analysis of variance and analysis of covariance models to construct and test hypotheses considering complex relationships among multiple variables.
  - LO6: Be able to use standard computer packages such as R to conduct the quantitative analyses described in the learning objectives above.

2 Selecting the Right Method

Selecting Right Method

- Parametric Methods:
  - Only for *interval or ratio data*.
  - Make sure assumptions of CLT hold:
    - Large sample size *or*..
    - Normal distributed *population*.
- Non-parametric methods using ranks
  - Ordinal data *and/or*...
  - Central limit theorem does not apply.

2.1 Single Population

Single Population

- Examine a proportion
  - Parametric: treat data as 0s and 1s, T-test for a single mean.
  - There is a non-parametric test: Binomial distribution.
- Examine the "average" (measure of center) of a single population.
  - Parametric method: T-test for a single mean.
  - Non-parametric methods: Wilcoxon Signed Rank test for a single median.

2.2 Differences in Two Populations

Differences in Two Populations

- Independent Samples
  - Parametric: T-test for difference in means.
  - Non-parametric: Mann-Whitney U-Test - tests whether two populations have the same median.
- Paired samples (Dependent Samples)
  - Parametric: Paired samples T-Test for a difference in means.
  - Non-parametric: Wilcoxon signed rank test for a difference in medians.

2.3 Relationships Between Two Variables

Relationships Between Two Variables

- Parametric method: Pearson linear correlation coefficient.
- Non-parametric method: Spearman correlation.
- Two categorical variables: Chi-squared test of independence.

2.4 Differences in More than Two Populations

Differences in More than Two Populations

- Parametric method: Analysis of Variance (ANOVA)
  - Compares the means of two or more populations.
  - Null hypothesis: all populations have the same mean.

  - Alternative hypothesis: at least one population has a mean different than the others.

- Non-parametric method: Kruskal-Wallis test.
  - Compares the medians of two or more populations.
  - Null hypothesis: all populations have the same median.
  - Alternative hypothesis: at least one population has a median different than the others.

3 Analysis of Variance

3.1 Variance Decomposition

One-Way ANOVA

- Method for testing for significant differences among means from two or more groups.
- Essentially an extension of independent samples test for differences in means
- Uses measures of *variance* to measure for differences in *means*.
- Total variation in your data is decomposed into two components:
  - **Among-group variation**: variability that is due to differences among groups, also called **explained** variation.
  - **Within-group variation**: total variability within each of the groups, this is **unexplained** variation.

3.2 Assumptions

Assumptions behind One-way ANOVA F-test

- Randomness: random assignment to groups independently of the outcome
- Independence: individuals in each group are independent from individuals other groups
- Sufficiently large (?) sample size, or else population has normal distribution.
- Homogeneity of variance: the variances of each of the K groups must be equal ($\sigma_1^2 = \sigma_2^2 = ...\sigma_K^2$).
  - Levene test for homogeneity of variance can be used to test for this.

4 Kruskal-Wallis Test: Non-parametric Test

4.1 Non-parametric "ANOVA"

Non-parametric One-way ANOVA

- Kruskal-Wallis Rank Test: non-parametric technique for testing for differences in the *medians* among two or more groups.
- Like the Mann-Whitney U-test, uses information about the ranks of the observations.
- Null hypothesis: All groups have the same median.
- Alternative hypothesis: At least one of the medians differs.

4.2 Assumptions

Assumptions for Kruskal-Wallis Test

- Random assignment: random assignment to groups is independent of the outcome variable
- Independence: individuals in each group are independent from individuals in another group
- Only the location (i.e. the center) of the distributions differ among the groups, distributions otherwise have the same shape