ST 553 - Practice Midterm Cheatsheet

Brian Cervantes Alvarez

Formulas and Key Concepts

1. Cell Means Model

- **Formula:** $y_{ij} = \mu_i + \epsilon_{ij}$
- **When to Use:** Employ this model when analyzing data from experiments with different treatment groups, particularly in agricultural or biological studies involving categorical independent variables.
- **Description:** This model is used to analyze the effects of categorical treatments on a continuous response variable. Each μ_i represents the mean response for treatment i, while ϵ_{ij} represents the random error associated with each observation, assumed to be normally distributed with mean zero and constant variance.

2. ANOVA F-test

- **Formula:** $F = \frac{MS_{treatment}}{MS_{error}}$
- **When to Use:** Utilize the ANOVA F-test to determine if there are significant differences among group means in a scenario where there are three or more groups.
- **Description:** This test compares the mean squares between the treatments (MS_treatment) with the mean squares within the treatments (MS_error) to assess if the group means are statistically significantly different. The F-value is calculated by dividing the variance attributed to the treatment by the variance within the treatments.

3. Tukey's Honestly Significant Difference (HSD) Test

- **Formula:** $q = \frac{\text{Mean}_i - \text{Mean}_j}{\text{SE}}$
- **When to Use:** After an ANOVA indicates significant differences, use Tukey's HSD to compare all possible pairs of means.

- **Description:** This method provides a single-step multiple comparison procedure that controls the family-wise error rate. It's particularly useful for pairwise comparisons and is based on the studentized range distribution. The test calculates a q-value, comparing the absolute difference between two means to the standard error of differences among means, adjusted for multiple comparisons.

4. Scheffé's Method

- **Formula:** $F = \frac{(SSC_i/c)}{MS_{error}}$
- **When to Use:** When exploring multiple linear comparisons among means, especially if the number of hypotheses to test is not predetermined.
- **Description:** Scheffé's method is useful for constructing confidence intervals around any linear combination of means and testing complex hypotheses about them. It provides the flexibility to formulate hypotheses post hoc without increasing Type I error. This method is generally less powerful for pairwise comparisons but advantageous for broader hypothesis testing.

5. Bonferroni Method

- **Formula:** Adjusted $\alpha = \frac{\alpha}{k}$ (where k is the number of comparisons)
- **When to Use:** When you need a stringent control of Type I errors across multiple comparisons.
- **Description:** This correction method adjusts the significance level α for the number of comparisons being made, reducing the chance of spurious findings by dividing the overall alpha level by the number of tests. It's widely used due to its simplicity and conservative nature.

6. Contrast Analysis

- **Formula:** $C = \sum c_i \mu_i$
- **When to Use:** To make specific, theoretically driven comparisons among treatment means, often specified before data collection.
- **Description:** This statistical technique allows the researcher to test for differences in combinations of means according to their theoretical importance. Coefficients c_i are weights assigned to each mean μ_i that specify the nature of the hypothesis being tested.

7. Distribution of the ANOVA F-statistic

- **Formula:** $F \sim F(df_{between}, df_{within})$

- **When to Use:** To determine the sampling distribution of the F-stat the null hypothesis for comparison with the observed F-value.
- **Description:** This concept is crucial for understanding the behavior of the F-statistic under the assumption that no treatment effects exist. The F-distribution is used to determine p-values for the ANOVA test, based on the degrees of freedom for the treatment (between groups) and error (within groups).