

# STA 3180 Statistical Modelling: ANOVA

## LECTURE NOTES ON ANOVA FOR STA 3180 STATISTICAL MODELLING

What is ANOVA?

ANOVA stands for Analysis of Variance. It is a statistical method used to compare the means of two or more groups. ANOVA can be used to determine if there is a statistically significant difference between the means of two or more groups.

Key Concepts:

- **Null Hypothesis:** The null hypothesis states that there is no difference between the means of the groups being compared.
- **Alternative Hypothesis:** The alternative hypothesis states that there is a difference between the means of the groups being compared.
- **F-Test:** The F-test is used to test the null hypothesis. It is calculated by dividing the variance between the groups by the variance within the groups.
- **Degrees of Freedom:** The degrees of freedom are used to calculate the F-test. It is the number of observations in the sample minus the number of parameters estimated.
- **Type I Error:** A type I error occurs when the null hypothesis is rejected even though it is true.
- **Type II Error:** A type II error occurs when the null hypothesis is accepted even though it is false.

Definitions:

- **Variance:** Variance is a measure of how spread out the data is. It is calculated by taking the sum of the squared differences between each observation and the mean, and then dividing by the number of observations minus one.
- **Sum of Squares:** The sum of squares is used to calculate the variance. It is the sum of the squared differences between each observation and the mean.

Rules:

- The null hypothesis should be stated before conducting an ANOVA test.
- The alternative hypothesis should be stated before conducting an ANOVA test.
- The F-test should be calculated using the degrees of freedom.

- The F-test should be compared to a critical value to determine if the null hypothesis should be rejected or accepted.

Examples:

Suppose we want to compare the mean heights of three different groups of people: men, women, and children. We can use ANOVA to determine if there is a statistically significant difference between the mean heights of the three groups.

First, we state the null hypothesis: There is no difference between the mean heights of the three groups.

Next, we state the alternative hypothesis: There is a difference between the mean heights of the three groups.

We then calculate the F-test using the degrees of freedom. The degrees of freedom is equal to the number of observations in the sample minus the number of parameters estimated (in this case, three).

Finally, we compare the F-test to a critical value to determine if the null hypothesis should be rejected or accepted. If the F-test is greater than the critical value, then we can reject the null hypothesis and conclude that there is a statistically significant difference between the mean heights of the three groups.