STA 3180 Statistical Modelling: ANOVA

STA 3180 Statistical Modelling - Lecture Notes on ANOVA

Introduction

Analysis of Variance (ANOVA) is a statistical technique used to compare the means of two or more groups. It is used to determine whether there are significant differences between the means of different groups. ANOVA can be used to compare the means of two or more independent groups, or to compare the means of two or more dependent groups.

Key Concepts

* Null hypothesis: The null hypothesis states that there is no difference between the means of the two or more groups being compared.

* Alternative hypothesis: The alternative hypothesis states that there is a difference between the means of the two or more 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: Degrees of freedom is the number of values in a dataset that are free to vary.

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 value and the mean, and dividing by the number of values.

* Sum of Squares: The sum of squares is the sum of the squared differences between each value and the mean.

* Mean: The mean is the average of a set of values. It is calculated by adding up all the values and dividing by the number of values.

Coding Examples

Calculating Variance

```
Start of Code
```

```
# Calculate the variance of a dataset
# Define the dataset
dataset = [1, 2, 3, 4, 5]
# Calculate the mean
mean = sum(dataset) / len(dataset)
```

```
# Calculate the sum of squares
sum_of_squares = 0
for value in dataset:
	sum_of_squares += (value - mean)**2
# Calculate the variance
variance = sum_of_squares / len(dataset)
print(variance)
```

End of Code

Calculating F-test

Start of Code

```
# Calculate the F-test for two datasets
# Define the datasets
dataset1 = [1, 2, 3, 4, 5]
dataset2 = [6, 7, 8, 9, 10]
# Calculate the means
mean1 = sum(dataset1) / len(dataset1)
mean2 = sum(dataset2) / len(dataset2)
# Calculate the sum of squares
sum_of_squares1 = 0
for value in dataset1:
	sum_of_squares1 += (value - mean1)**2
sum_of_squares2 = 0
for value in dataset2:
	sum_of_squares2 += (value - mean2)**2
# Calculate the variance between the groups
variance_between_groups = (sum_of_squares1 + sum_of_squares2) /
(len(dataset1) + len(dataset2))
# Calculate the variance within the groups
variance_within_groups = (sum_of_squares1 + sum_of_squares2) /
(len(dataset1) * len(dataset2))
# Calculate the F-test
F_test = variance_between_groups / variance_within_groups
print(F_test)
```

End of Code