STA 3100 Programming With Data in R: Hypothesis Testing

Hypothesis Testing in R

Introduction

Hypothesis testing is a statistical procedure used to determine if there is a significant difference between two sets of data. It is used to test the validity of a hypothesis by using sample data. In this lecture, we will be discussing the basics of hypothesis testing and how to conduct it in R.

Definition

A hypothesis is an educated guess or prediction about the relationship between two variables. Hypothesis testing is a process of testing the validity of a hypothesis by using sample data. The goal of hypothesis testing is to determine if the observed data is statistically significant enough to reject the null hypothesis.

Key Concepts

1. Null Hypothesis: The null hypothesis is a statement that there is no relationship between two variables. This is the hypothesis that is tested in a hypothesis test.

2. Alternative Hypothesis: The alternative hypothesis is a statement that there is a relationship between two variables. This is the hypothesis that is accepted if the null hypothesis is rejected.

3. Type I Error: A type I error occurs when the null hypothesis is rejected when it is actually true.

4. Type II Error: A type II error occurs when the null hypothesis is accepted when it is actually false.

Coding Example

We will use the following example to illustrate how to conduct a hypothesis test in R.

Suppose we want to test the hypothesis that the mean height of adult males is 5 feet 8 inches. We have a sample of 10 adult males with the following heights:

5'6", 5'9", 5'7", 5'10", 5'8", 5'11", 6'0", 5'9", 5'10", 5'9".

```
Start of Code
# Set up the null and alternative hypotheses
null_hypothesis <- "The mean height of adult males is 5 feet 8 inches"
alternative_hypothesis <- "The mean height of adult males is not 5 feet 8 inches"
# Create a vector of the heights
heights <- c(5.6, 5.9, 5.7, 5.10, 5.8, 5.11, 6.0, 5.9, 5.10, 5.9)
# Calculate the mean
mean_height <- mean(heights)
# Calculate the standard deviation
```

```
std_dev <- sd(heights)
# Calculate the t-statistic
t_stat <- (mean_height - 5.8)/(std_dev/sqrt(length(heights)))
# Calculate the p-value
p_value <- 2*pt(-abs(t_stat), df = length(heights)-1)
# Print the results
print(paste("The t-statistic is", t_stat))
print(paste("The p-value is", p_value))
End of Code
```

Practice Multiple Choice Questions

Q1. What is the goal of hypothesis testing?

A. To determine if the observed data is statistically significant enough to accept the null hypothesis.

B. To determine if the observed data is statistically significant enough to reject the null hypothesis.

C. To determine if the observed data is statistically significant enough to accept the alternative hypothesis.

D. To determine if the observed data is statistically significant enough to reject the alternative hypothesis.

Answer: B. To determine if the observed data is statistically significant enough to reject the null hypothesis.