STA 3100 Programming With Data in R: Regression

Regression in R

Overview

Regression is a statistical method used to predict the value of a dependent variable (Y) based on the values of one or more independent variables (X). It is used to analyze the relationship between variables and to identify patterns in data. In R, there are several packages available for performing regression analysis.

Key Concepts

- Dependent Variable: The variable that is being predicted.

- Independent Variable: The variable(s) used to predict the dependent variable.

- Linear Regression: A type of regression where the relationship between the dependent and independent variables is assumed to be linear.

- Non-Linear Regression: A type of regression where the relationship between the dependent and independent variables is not assumed to be linear.

Coding Examples

Simple Linear Regression
Start of Code
```R
# Load the necessary packages
library(tidyverse)

# Read in the data
data <- read.csv("data.csv")

# Fit the linear model
model <- lm(Y ~ X, data = data)

# Print the summary of the model
summary(model)
```
End of Code

Multiple Linear Regression
Start of Code
```R
# Load the necessary packages
library(tidyverse)

# Read in the data
data <- read.csv("data.csv")
```

```
# Fit the linear model
model <- lm(Y ~ X1 + X2 + X3, data = data)
# Print the summary of the model
summary(model)
```
End of Code

Practice Questions

Q: What type of regression is used when the relationship between the dependent and independent variables is assumed to be linear?

A: Linear regression.