STA 3180 Statistical Modelling: Statistical Thinking

STA 3180 Statistical Modelling - Lecture Notes

Introduction to Statistical Thinking

Statistical thinking is the process of using data to make decisions and draw conclusions. It involves the use of data to identify patterns, trends, and relationships between different variables. Statistical thinking is an important skill for anyone working in the field of data science, as it allows them to make informed decisions and create meaningful insights from their data.

Key Concepts

Descriptive Statistics: Descriptive statistics are used to summarize data and describe its characteristics. This includes measures of central tendency (mean, median, and mode) and measures of variability (standard deviation, range, and interquartile range).

Inferential Statistics: Inferential statistics are used to make inferences about a population based on a sample. This includes tests of significance such as the t-test, ANOVA, and chi-square test.

Probability: Probability is the measure of how likely an event is to occur. It is expressed as a number between 0 and 1, with 0 indicating that the event is impossible and 1 indicating that the event is certain.

Random Variables: A random variable is a variable whose value is determined by chance. It can take on any value within a given range.

Definitions

Data: Data is information collected from observations or experiments. It can be numerical or categorical.

Population: A population is the entire set of individuals or items being studied.

Sample: A sample is a subset of the population that is used to make inferences about the population.

Parameter: A parameter is a numerical measure that describes a population.

Statistic: A statistic is a numerical measure that describes a sample.

Hypothesis Testing: Hypothesis testing is a method of making inferences about a population based on a sample. It involves formulating a null hypothesis and an alternative hypothesis, collecting data, and then testing whether the null hypothesis can be rejected.

Coding Examples

```
Start of Code
# Calculate mean of a vector
# Create a vector of numbers
x <- c(1,2,3,4,5)
# Calculate the mean of the vector
mean(x)
End of Code
```

Practice Multiple Choice Questions

Q: What is the purpose of descriptive statistics?

A) To make inferences about a population

B) To summarize data and describe its characteristics

C) To test hypotheses

D) To identify patterns and trends

Answer: B) To summarize data and describe its characteristics