STA 3180 Statistical Modelling: Statistical Graphics

STA 3180 Statistical Modelling - Lecture Notes on Statistical Graphics

Introduction

Statistical graphics are visual representations of data that allow us to quickly and easily identify patterns, trends, and relationships in the data. They are an essential tool for any data analyst, as they can help to identify outliers, detect errors, and provide insights into the data that would otherwise be difficult to discern. In this lecture, we will discuss the different types of statistical graphics, how to create them in R, and how to interpret them.

Types of Statistical Graphics

There are several different types of statistical graphics, each of which has its own purpose and use. The most common types of statistical graphics include:

- Bar Charts: Bar charts are used to compare categorical data. They are useful for displaying the frequency or percentage of observations within each category.

- Histograms: Histograms are used to display the distribution of a continuous variable. They are useful for identifying the shape of the distribution and any outliers.

- Line Charts: Line charts are used to show the relationship between two variables. They are useful for identifying trends and correlations.

- Scatter Plots: Scatter plots are used to display the relationship between two continuous variables. They are useful for identifying outliers and correlations.

- Box Plots: Box plots are used to display the distribution of a continuous variable. They are useful for identifying outliers and the shape of the distribution.

Creating Statistical Graphics in R

In R, there are several different packages that can be used to create statistical graphics. The most commonly used packages are `ggplot2` and `lattice`. Both of these packages provide a wide range of plotting functions that can be used to create various types of statistical graphics.

ggplot2

The `ggplot2` package is a powerful and flexible plotting system for R. It provides a wide range of plotting functions, including functions for creating bar charts, histograms, line charts, scatter plots, and box plots. To create a plot using `ggplot2`, we first need to load the package:

```
Start of Code
library(ggplot2)
End of Code
```

Once the package is loaded, we can use the `ggplot()` function to create a plot. This function takes several arguments, including the data to be plotted, the aesthetics (i.e. the variables to be plotted), and the type of plot to be created. For example, to create a bar chart of the number of observations in each category of a categorical variable, we could use the following code:

```
Start of Code
ggplot(data, aes(x = variable, y = count)) +
    geom_bar(stat = "identity")
End of Code
```

lattice

The `lattice` package is another powerful and flexible plotting system for R. It provides a wide range of plotting functions, including functions for creating bar charts, histograms, line charts, scatter plots, and box plots. To create a plot using `lattice`, we first need to load the package:

```
Start of Code
library(lattice)
End of Code
```

Once the package is loaded, we can use the `xyplot()` function to create a plot. This function takes several arguments, including the data to be plotted, the aesthetics (i.e. the variables to be plotted), and the type of plot to be created. For example, to create a scatter plot of two continuous variables, we could use the following code:

```
Start of Code
xyplot(y ~ x, data = data)
End of Code
```

Interpreting Statistical Graphics

Once a statistical graphic has been created, it is important to interpret it correctly. To do this, we need to look for patterns, trends, and relationships in the data. For example, if we were looking at a bar chart of the number of observations in each category of a categorical variable, we would look for differences in the heights of the bars to identify which categories have more or fewer observations. If we were looking at a histogram of a continuous variable, we would look for the shape of the distribution and any outliers. If we were looking at a line chart of two variables, we would look for any trends or correlations.

Practice Multiple Choice Questions

Q1. Which of the following is NOT a type of statistical graphic?

A. Pie Chart

B. Bar Chart

C. Line Chart

D. Scatter Plot

Answer: A. Pie Chart