

Data Science 1

Analysis of Categorical Data

Introduction

Ann Maharaj

Analysis of Categorical Data: Introduction

1 Analysis of Categorical Data

2 Available materials

Analysis of Categorical Data: Introduction

- Categorical data which arise from one or more non-numeric variables, are frequency counts of categories from these variables.
- Hence the variables of interest are qualitative.
- For example, it is determined that in a particular company there are 16 people employed in Engineering services, 30 in Computer services, 13 in Finance services and 5 in other services.
 - The variable of interest is: Type of service area. This is a qualitative variable.
 - Categories of the variable: Engineering, Computer, Finance and Other services. These are on the nominal scale since they are labels.
 - The data are frequency counts of how many people work in each of these services.

Analysis of Categorical Data: Introduction

- In previous modules, we studied procedures for testing hypotheses and constructing confidence intervals concerning the proportion of a single population and the equality of proportions of two independent populations.
- We now consider alternative procedures for testing these hypotheses, using chi-square tests.
- This will enable us extend the test of equality of two independent population proportions to that of more than two independent populations as well as to tests of independence of two qualitative variables.
- We will also be working through categorical tests for two or more related populations.

Available materials

- 1 Short videos including demonstrations using R software
- 2 Accompanying slides
- 3 R scripts and relevant data sets
- 4 Exercises
- 5 Additional material
- 6 List of references