

Multivariate Categorical Data Analysis

Introductory Multivariate Analysis Part 2 Course Reader 2023

Dr Retha Luus

Table of Contents

1	Introduction.....	2
2	Probability Distributions for Categorical Data.....	2
2.1	Binomial Distribution.....	2
2.2	Multinomial Distribution.....	3
3	Characterizing and Displaying Univariate Categorical Data.....	3
3.1	Numeric descriptive statistics.....	3
3.2	Graphical descriptive statistics.....	7
3.2.1	Bar chart.....	7
3.2.2	Pie Chart.....	8
3.2.3	Inspect Df.....	9
4	Characterizing and Displaying Bivariate Categorical Data.....	11
4.1	Graphical descriptive statistics.....	14
4.1.1	Grouped bar plots.....	14
4.1.2	Stacked bar plots.....	16
4.2	Comparing Proportions in 2x2 Contingency Tables.....	18
4.2.1	Difference of Proportions.....	18
4.2.2	Relative Risk.....	20
4.2.3	Odds Ratio.....	21
4.3	Tests of Independence: Nominal Variables.....	24
4.3.1	Pearson's Chi-squared Statistic.....	24
4.3.2	Standardized Residuals.....	27
4.3.3	Limitations of Chi-Squared Tests.....	28
4.3.4	Fisher's Exact Test.....	28

4.4	Correspondence Analysis.....	30
4.4.1	Balloon Plot.....	33
4.4.2	Row variables.....	34
4.4.3	Column variables.....	37
4.4.4	Association between row and column variables.....	39
4.4.5	Correspondence analysis plot.....	40
4.5	Tests of Independence: Ordinal Variables.....	41
4.6	Comparing Categorical and Numerical Variables.....	43
5	Summary.....	52
6	References.....	52

Multivariate Categorical Data Analysis

Introductory Multivariate Analysis Part 2 Course Reader 2024

Dr Retha Luus

Table of Contents

1.	Introduction.....	3
2.	Probability Distributions for Categorical Data.....	6
2.1	Binomial Distribution.....	7
2.2	Multinomial Distribution.....	7
3.	Characterizing and Displaying Univariate Categorical Data.....	8
3.1	Numeric Descriptive Statistics.....	8
3.2	Graphical Descriptive Statistics.....	12
3.2.1	Bar Chart.....	12
3.2.2	Pie Chart.....	14
3.2.3	Inspect Df.....	15
4.	Characterizing and Displaying Bivariate Categorical Data.....	16
4.1	Graphical Descriptive Statistics.....	21
4.1.1	Grouped Bar Plots.....	21
4.1.2	Stacked Bar Plots.....	23
4.2	Comparing Proportions in 2x2 Contingency Tables.....	26
4.2.1	Difference of Proportions.....	27
4.2.2	Relative Risk.....	29
4.2.3	Odds Ratio.....	31
4.3	Tests of Independence: Nominal Variables.....	34
4.3.1	Pearson's Chi-squared Statistic.....	34
4.3.2	Standardized Residuals.....	37
4.3.3	Limitations of Chi-Squared Tests.....	37

4.3.4	Fisher's Exact Test.....	38
4.4	Correspondence Analysis.....	40
4.4.1	Balloon Plot.....	43
4.4.2	Row Variables.....	45
4.4.3	Column Variables.....	48
4.4.4	Association between Row and Column Variables.....	51
4.4.5	Correspondence Analysis Plot.....	53
4.4.6	Correspondence Analysis with FactoMineR.....	54
4.5	Tests of Independence: Ordinal Variables.....	57
4.6	Comparing Categorical and Numerical Variables.....	59
4.6.1	Numeric Descriptive Analysis.....	59
4.6.2	Graphical Descriptive Analysis.....	60
4.6.3	Statistical Significance: Numeric and Nominal/Ordinal (2 levels).....	62
4.6.4	Statistical Significance: Numeric and Nominal/Ordinal (more than 2 levels).....	66
5.	Summary.....	69
6.	Bibliography.....	70

Multivariate Categorical Data Analysis

Introductory Multivariate Analysis Part 2 Course Reader 2025

Dr Retha Luus

Table of Contents

List of Tables.....	3
List of Figures.....	4
1. Introduction.....	5
2. Probability Distributions for Categorical Data.....	9
2.1 Binomial Distribution.....	10
2.2 Multinomial Distribution.....	10
3. Characterising and Displaying Univariate Categorical Data.....	11
3.1 Numeric Descriptive Statistics.....	11
3.2 Graphical Descriptive Statistics.....	15
3.2.1 Bar Chart.....	15
3.2.2 Pie Chart.....	17
3.2.3 Inspect Df.....	18
4. Characterising and Displaying Bivariate Categorical Data.....	19
4.1 Graphical Descriptive Statistics.....	24
4.1.1 Grouped Bar Plots.....	24
4.1.2 Stacked Bar Plots.....	26
4.2 Comparing Proportions in 2x2 Contingency Tables.....	29
4.2.1 Difference of Proportions.....	30
4.2.2 Relative Risk.....	32
4.2.3 Odds Ratio.....	34
4.3 Tests of Independence: Nominal Variables.....	38
4.3.1 Pearson's Chi-squared Statistic.....	38

4.3.2	Standardised Residuals.....	41
4.3.3	Limitations of Chi-Squared Tests.....	42
4.3.4	Fisher’s Exact Test.....	42
4.4	Effect Size.....	44
4.4.1	Cramer’s V.....	45
4.5	Correspondence Analysis.....	47
4.5.1	Contingency Table Visualisation.....	51
4.5.2	Row Variables.....	52
4.5.3	Column Variables.....	55
4.5.4	Association between Row and Column Variables.....	58
4.5.5	Correspondence Analysis Plot.....	60
4.5.6	Correspondence Analysis with FactoMineR.....	61
4.6	Tests of Independence: Ordinal Variables.....	64
4.7	Comparing Categorical and Numerical Variables.....	66
4.7.1	Numeric Descriptive Analysis.....	66
4.7.2	Graphical Descriptive Analysis.....	67
4.7.3	Statistical Significance: Numeric and nominal/ordinal (2 levels).....	69
4.7.4	Point-Biserial Correlation.....	74
4.7.5	Statistical Significance: Numeric and nominal/ordinal (more than 2 levels).....	76
5.	Summary.....	79
6.	Bibliography.....	80