

Categorical data analysis

Academic Year: (2022 / 2023)

Review date: 30-05-2022

Department assigned to the subject: Statistics Department

Coordinating teacher: CABRAS , STEFANO

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Statistical Inference Techniques I
Statistical Inference Techniques II
Regression Methods

OBJECTIVES

1. Understanding the basic techniques for analyzing categorical data.
2. Knowing and managing statistical programs for the analysis of categorical data.
3. Using the methodology for the analysis of real data.

1. Capacity for analysis and synthesis.
2. Modeling and resolution of problems.
3. Oral and written communication.

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction.
 - 1.1. Categorical Response Data.
 - 1.2. General approach to different statistical techniques.
 - 1.3. Examples.
2. Contingency Tables. Measures of relationship and association. Contrasts.
 - 2.1. Association Measures for categorical data.
 - 2.2. Statistical inference (parametric and nonparametric).
 - 2.3. Examples.
3. Simple and multiple correspondence analysis.
 - 3.1. Introduction: assumptions, estimation and interpretation.
 - 3.2. Simple correspondence analysis.
 - 3.3. Multiple correspondence analysis.
 - 3.4. Examples.
4. Decision trees.
 - 4.1. Introduction: assumptions, estimation and interpretation.
 - 4.2. Algorithms: CHAID, CART and QUEST.
 - 4.3. Examples.
5. Generalized Linear Models (GLM). Models for binary data (logistic regression) and multiple response.
 - 5.1. Introduction to GLM and comparison with other models.
 - 5.2. Limited dependent variable models: models for binary data. Binary logistic regression: assumptions, estimation and interpretation.
 - 5.3. Models for multinomial data. Multiple Logistic Regression: assumptions, fitting and interpretation.
 - 5.4. Examples.

LEARNING ACTIVITIES AND METHODOLOGY

Theory (4 ECTS). Theoretical classes with support material available on the Web. Practice (2 ECTS) problem-solving classes. Computing practices in computer lab.

ASSESSMENT SYSTEM

Specific tasks throughout the course and final assignment count 100% of the final note.

Students who get no less than 5 on continuous assessment will be released for the final exam. In this case, the note of continuous evaluation will be worth 100% of the final mark only if the continuous evaluation is higher or equal to 5. However, students that never attempt continuous evaluation tasks could have the final exam, which only counts 60% of the final note.

% end-of-term-examination:	0
% of continuous assessment (assignments, laboratory, practicals...):	100

BASIC BIBLIOGRAPHY

- Agresti, A Categorical Data Analysis, New York: John Wiley & Sons, 2013 (third Edition)
- Agresti, A. An introduction to Categorical data analysis, John Wiley & Sons,, 2007
- Andersen, E.B Introduction to the Statistical Analysis of Categorical Data, Springer, 1997
- Collett D. Analysis of Binary Data, Chapman & Hall., 2003
- Cox D.R. & Snell E.J. Analysis of Binary Data, Chapman & Hall, 1989
- Cox D.R. & Snell E.J. Analysis of Binary Data, Chapman & Hall, 2018
- Kateri, M Contingency Table: Analysis Methods and Implementation Using R, Birkhäuser, 2014
- Zelterman, D Models for Discrete Data, Oxford University Press, 2006 (revised edition)

ADDITIONAL BIBLIOGRAPHY

- Bishop, Y. M., Fienberg, S. E., Holland, Paul W. Discrete Multivariate Analysis: Theory and Practice, Springer (Originally published by MIT Press, 1975), 2007
- Hosmer, D.W. and Lemeshow, S. Applied Logistic regression, Willey, 2000
- McCullagh, P. and Nelder, J.A. Generalized Linear Models, Second Edition, London: Chapman & Hall, 1989
- Stokes, M.E., Davis, C.S. and Koch, G.G. Categorical Data Analysis Using The SAS System, Second Edition, NC: SAS Institute Inc., 2000

BASIC ELECTRONIC RESOURCES

- Alan Agresti . Website for CATEGORICAL DATA ANALYSIS, 3rd edition: <http://www.stat.ufl.edu/~aa/cda/cda.html>