CEU

Political Science MA

Multivariate Statistics

Winter 2020

Instructor: Tamás Rudas

Instructor e-mail: rudast@ceu.edu

Teaching assistant: TBA

The class meets: 15:30-17:10 Mondays and Thursdays

Instructor office hours: 14:30-15:30 Mondays and Thursdays Please send an e-mail if you wish to come

TA office hours: TBA

Learning goals

The course develops, on the one hand, a theoretical understanding of why and how multivariate statistical methods are applied in political science and, on the other hand, skills, including the application of statistical packages, to carry out such analyses and the interpretation of their results. The course prepares the students to critical application of multivariate statistical methods and requires oral and written presentation of the results. Precise interpretation of the results of the analyses will be emphasized, throughout. The presentation of the statistical methods will concentrate on the research question they may answer, and variations depending on characteristics of the data, including the level of measurement, will be discussed as a second level of differentiation.

Material to be covered (subject to revision)

Univariate regression

One-way analysis of variance

Multivariate regression

Two-way analysis of variance

Logistic regression

Log-linear modeling

Factor analysis

Cluster analysis

Tentative class schedule

1.09. Introduction: the role of multivariate statistical analysis, questions which may be answered with statistical analysis, data collection, experiments and observational studies, levels of measurement

1.13. Review of univariate regression analysis, the statistical question, level of measurement of the variables, model fit, parameter estimates

1.16. One-way analysis of variance: levels of measurement, operationalization of effect, sums of squares

1.20. Model fit, test of effect, the role of normality, nonparametric variants

Students report research group membership

1.23. Two-way analysis of variance, model fit, test of interaction

1.27. Multivariate regression analysis: model fit, parameter estimates, interpretation, limitations

1.30. Stepwise model selection, sources of misfit, ANOVA around the predicted values

Students report research topics and data files

2.03. Logistic regression (categorical response and numerical explanatory variables)

2.06. Review class

2.10. Log-linear models for categorical data: conditional and higher order odds ratios, association and simplicity, log-linear model interpretation, regression type models

Students report plan of data analysis

2.13. Parameter estimates, maximum likelihood estimates, interpretation of parameters

2.17. Testing model fit

2.20. Graphical models of multivariate data

2.24. Reduction of dimensionality for numerical variables: principal component and factor analysis

2.27. Cluster analysis, the k-means method

3.02. Review class

3.05. Midterm test

3.09. Student work in progress presentations

3.12. Student work in progress presentations

3.16. Discussion of the final projects, and time reserve for material not covered

3.19. Student final presentations

3.23. Student final presentations

3.26. Student final presentations

Date to be announced: submit research report

Texts

Free on the internet

www.rcommander.com

http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf

http://little-book-of-r-for-multivariate-analysis.readthedocs.org/en/latest/

http://cran.r-project.org/doc/contrib/Karp-Rcommander-intro.pdf

The library has the following relevant books:

Applied MANOVA and Discriminant Analysis / Huberty, Carl J.

Applied Multivariate Data Analysis / J.D. Jobson

Categorical Data Analysis / Alan Agresti

Graphical Models in Applied Multivariate Statistics / Joe Whittaker

Handbook of Probability :Theory and Applications / Tamas Rudas

Lectures on Categorical Data Analysis / Tamas Rudas

Log-linear Models and Logistic Regression / Ronald Christensen

Multivariate Analysis / W.J. Krzanowski and F.H.C. Marriott

Odds Ratios in the Analysis of Contingency Tables / Tamas Rudas

Probability Theory : A Primer / Tamas Rudas

Reading and Understanding Multivariate Statistics / Laurence G. Grimm and Paul R. Yarnold

Reading and Understanding More Multivariate Statistics / L. G. Grimm and P. R. Yarn

The Statistical Analysis of Discrete Data / Thomas J. Santner, Diane E. Duffy

Using Multivariate Statistics / Barbara G. Tabachnick, Linda S. Fidell

Students are not expected to read these books. Rather, the books may be consulted on particular topics.

Software

The class is software neutral. This means that students are welcome to run the statistical analyses discussed on any software of their choice. The instructor will use R for classroom presentations, with the graphical user interface R Commander. Support will be provided to students to implement these free pieces of software and making their own data readable by R.

Class procedures

The class relies on active student participation to make sure each student gains a clear theoretical knowledge and hands-on experience. The in-class presentations by the instructor will cover theoretical aspects illustrated by simple applications. After each class, a very brief summary of the material covered will be posted on the e-learning site. Students will write a midterm test, submit assignments and work on a project. Students are required to form groups to help each other in completing their projects. Each group will work on the analysis of a data set of their choice. While joint work is encouraged, each student will have to have an identifiable contribution to it. In particular, each student will have to choose an analytical problem, contribute to a work-in-progress presentation and give a final presentation in class and submit a final written research report.

Grading and student requirements

To earn credit in this class, each student will have to

- write a midterm test (35%)
- contribute a work-in-progress presentation of their project (5%)
- give a final presentation of their project (25%)
- submit a research report at the end of the term (35%)

Students who audit the class will have to successfully write the midterm test