Marketing Research

Regression methods and structural equation models (a computational R based approach)

Professor: Albert Satorra

Office hours: (by appointment)

Course Description

The course focuses on classical and modern regression methods useful for business, marketing, and behaviour studies. Taking advantage that "R is the amazing, free, open-access software package for scientific graphs and calculations used by scientists worldwide" the course will be computationally based on R. Using the "learning by doing" method, the course will address fundamental models for analysing the interrelationship among variables, including simultaneous regressions, possibly with latent/unmeasured variables. From the building blocks of multiple regression, the course will address logistic and generalised linear models, mixed-effects regression, exploratory and confirmatory factor analysis, and simultaneous equations. The SEM approach will be used to encompass a large and versatile family of models.

Objectives

At the end of the course, the students should be familiar with the basic concepts and the practice of the generalised regression methods, including structural equation models (SEM). They should also be fluent in the use of the free software R in their daily practice of statistics. One goal of the course also is to make the students be able to evaluate the use of statistical modelling by others' work critically. In particular, at the end of the course, the students should be familiar with:

- Fit linear, binary response regression models
- Adjust for measurement-error in predictor variables
- Estimate mixed-effects regression models
- Test complex causal theories with multiple pathways
- Estimate simultaneous equations with reciprocal effects.
- Incorporate latent variables with multiple indicators.
- Analyse longitudinal data.
- To be able to apply and discussing in their data, and work of others, basic aspects of the SEM approach.

Methodology

This is an applied course that will exploit the method of “learning by doing”. It will discuss examples and case studies involving marketing and survey data. Passing the course requires each student to undertake a Final Project that involves empirical data using the methods discussed in the course, the project culminating with a presentation and a final report. Whole classes can be devoted exclusively to hands-on use of the software of the course, R, and the assistance with individual course projects. Real-life marketing, behavioural and business data will be promoted to be used in the course projects

Prerequisites

The course advances the regression-based technique, so a one-quarter introductory graduate course in statistics (or equivalent) is required.

Evaluation criteria

Indicate all elements of evaluation and their weight in the final grade.

Assessment is composed of the following imputs:

1. Continuous evaluation: class discussion + homeworks (20%)
2. Final Project (40%)
3. Final Exam (40%).

Failing to achieve the 50% of any of the three components, will require a retake composed by the Main project (40%) and a final exam (60%)

Students are required to attend 80% of classes. Failing to do so without justified reason will imply a Zero grade in the participation/attendance evaluation item and may lead to suspension from the program

Students who fail the course during the regular evaluation are allowed ONE re-take of the evaluation, in the conditions specified above. If the course is again failed after the retake, the student will have to register again for the course the following year.

In case of a justified no-show to an exam, the student must inform the corresponding faculty member and the director(s) of the program so that they study the possibility of rescheduling the exam (one possibility being during the “Retake” period). In the meantime, the student will get an “incomplete”, which will be replaced by the actual grade after the final exam is taken. The “incomplete” will not be reflected on the student’s Academic Transcript.

Plagiarism is to use another's work and to present it as one's own without acknowledging the sources in the correct way. All essays, reports or projects handed in by a student must be original work completed by the student. By enrolling at any UPF BSM Master of Science and signing the "Honor Code," students acknowledge that they understand the schools' policy on plagiarism and certify that all course assignments will be their own work, except where indicated by correct referencing. Failing to do so may result in automatic expulsion from the program."

Final Project

Because the primary goal of the class is to prepare the student for the practice of the methods studied, a final project involving statistical modeling is required, ideally using students' data or with data of a secondary analysis (the data of a published work). In the last week of the course, time will be allocated for a 10-minute conference-style presentation of each of the projects, with discussion. The final project will be a written APA-style of "Method" and "Results and Discussion" (that should include both the statistical findings and some interpretations). The report should be about four typed pages (not counting appendices). Presentation and writing of the story should give the student the practice with (and feedback on) presenting the results of the statistical analysis. The Final Project written report due at the end of the Final Exam. Projects can be undertaken individually or in pairs.

Calendar and Contents

1. A review of data analysis and probability distributions in R

An example of data analysis with R, probability distributions, basic data analysis graphs (parallel boxplots, scatterplot, bar plots), comparing variation of continuous variable across groups, anova inference, resampling methods

2. Simple and multiple regression models

The linear regression model, marginal and partial regression "effects", diagnostic for linearity, outliers influential cases, dummy variables and ancova, prediction vs explanation, transformation of variables, warning on causality interpretation

3. Generalized linear models: logistic and Poisson regression

Motivation and frame of a generalized linear model, three basic examples: linear, logistic and Poisson regression. Conditional effect plots, the practice of glm()

4. Multivariate methods and the Exploratory Factor Model

Multivariate analysis and dimension reduction, PCA, multiple indicators of latent construct, Exploratory factor model, estimation methods, rotation, the practice of factanal( ) and FactoMineR

5. Measurement error and regression with errors-in-variables

Reliability and validity of behavioural variables, the impact of measurement error on regression analysis, accounting for measurement error, examples of the distortion due to measurement error.

6. Multivariate regression and simultaneous equations

Why multivariate regression, simultaneous equations, the path models, identification of parameters, full information estimation, lavaan SEM software, practice of simultaneous equations with no latent variables

7.. Basic SEM models.

Why latent variables in the model, multiple indicators model, the MIMIC model, the CFA model, regression with error in variables, models for measurement error (MTMM), other basic models, general SEM model, estimation using lavaan(), Models for longitudinal data.

8. Testing theories and model specification in SEM

Full information estimation methods, the chi-square goodness of fit test, diagnostic for model specification (Lagrange multiplier statistics, and expected parameter change), testing nested models, protection for non-normality, examples of application.

9. Wrap up of the course & Final Project presentations (with discussion)

Reading Materials/ Bibliography/Resources

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley & Sons.

Chapman, C., E. McDonnell Feit (2015), R for Marketing Research and Analytics, Springer

Hoyle, R. H (Ed.) (2012). Handbook of structural equation modeling. New York: Guilford Press.

Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4th edition). New York: Guilford Press.

Navarro, D. (2019) Learning statistics with R: A tutorial for psychology students and other beginners, Creative Commons BY-SA license (CC BY-SA) version 4.0. https://learningstatisticswithr.com/lsr-0.6.pdf

Shmueli, G, P.C. Bruce, I. Yahav, N.R. Patel, K. C. Lichtendahl Jr. (2018), Data Mining for Business Analytics, concepts, techniques, and applications in R, Wiley

Materials and technical support

Academic support for the course is provided via the web of the course that will be maintained by the instructor. The web of the course will list all recommended readings, presentations of lectures, replication code and data for empirical examples.