McDaniel Europe, Campus in Budapest

STA 2216 – Statistical Methods

Professors: Balazs Laki

Contact information
info@balazslaki.info

Availability
Balazs Laki is available by appointment, preferably before and after classes or via email.

Course description

This course gives students an opportunity to use the public domain and free software R to perform statistical computing. The R language provides a rich environment for working with data, especially for statistical modeling and graphics. The course emphasis is on data manipulation and basic statistical analysis. We will cover exploratory data analysis, programming, classical testing, categorical data analysis, and regression. Students will identify appropriate statistical methods for the data or problems and conduct their own analysis using the R environment. This is a hands-on, project-based course to enable students to develop skills and to solve statistical problems using R.

Course objectives:

The objective of this course is to provide students with an understanding of the concepts, the theory and the practice of statistical methodologies and to prepare students to conduct and critically evaluate statistical studies based on both large and small data sets of univariate and/or multivariate statistical analysis with a healthy scepticism. Emphasis is on underlying assumptions, limitations, adequate interpretations and practical applications of modern statistical methods. Drawing heavily from the methodologies of business administration and economics disciplines, the course focuses on the analytical tools that can be used for problem solving and showing the probabilistic way of thinking in decision making for a variety of applied fields in macroeconomics, microeconomics, industrial and international marketing, accounting, consumer's behaviour and quality planning & control. In addition, the course provides the students with the required knowledge and self-contained application skills needed for their further professional career.

Learning outcomes:

The goal of this course is to provide a supportive, hands-on environment for students to use R to conduct basic statistical analysis.

At the end of this course students will be able to:

- Create and modify R datasets
- Write their own R functions and use available packages in R
- Create figures and plots using R
- Perform and interpret parametric one- and two-sample tests using R
- Perform and interpret nonparametric one- and two-sample tests using R
- Perform and interpret chi-square analysis using R
- Perform and interpret multiple linear regression using R
- Perform and interpret one-way ANOVA using R
- Perform and interpret logistic and survival analyses using R

Required texts

- Class readers.
- (VS) An Introduction to R,
  URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Recommended texts

- (PD) P. Dalgaard. Introductory Statistics with R, 2nd Edition. (Springer 2008)
- Kleinman, Horton. R: Data Management, Statistical Analysis, Graphics (CRC 2009)
- Abedin. Data Manipulation with R. (Pact 2014)

Assignments & grading

Grading system -- 100 points total

*mid-term exam (40 points)
*final project (30 points for in-class presentation and 20 points for written report)
*class participation (10 points)

Standard scale: 100+ A+
93-100 A
90-92 A-
88-89 B+
83-87 B
80-82 B-
78-79 C+
73-78 C
70-72 C-
68-69 D+
63-67 D
60-62 D-
< 60 F

Project Paper
For the Spring STA 2216 term paper (2,500-3,000 words), (1) Students will work on final project individually. (2) Students will submit a proposal on their dataset and analysis plan by the fourth class period. (3) Students will analyze their own dataset. (4) Students will present their final projects in class and write a comprehensive report of the data analysis.

Honor Code

You are expected without question to adhere completely to the McDaniel College academic honor code. Any violation will result in a zero for the given assignment and other possible sanctions.

Course Policies

Informed, critical exchange of ideas forms the core of the McDaniel College learning experience. It should occur in every classroom. This is why in-class participation determines part of your grade. You are expected to share ideas during discussions; you are wholeheartedly encouraged to ask questions when you do not understand something; you will have chances to speak one-on-one with your fellow students as well as to speak before the entire class.
Creation of a proper classroom environment requires above all else respect for fellow students. We all ask that you don't be late; that you don't read newspapers, magazines, or otherwise distract everyone else during class; that you don't browse the Internet; that you turn off your cell phone and that you do not check for messages during class; that you don't pack up with five minutes left. Likewise, you can expect me to end class on time, to engage you in discussion and debate (that is, not to do all of the talking), and to be respectful of all points of view.
Participation implies attendance; absences will be noted and will adversely affect your final participation grade. In addition, in accordance with McDaniel Budapest's attendance policy, there will be an automatic deduction of one letter grade for every unexcused absence following your third unexcused absence.

Semester schedule/topics covered

Week 1 A Summary in Elementary Probability and Statistics
Data and Statistics, Descriptive Statistics, Set Theory, Elementary Probability Calculus, Basic Probability Distributions.

Week 2 A Summary in Elementary Statistics (continued) and Intro to R
Sampling and Sampling Distributions, Statistical Inference Making: Statistical Estimation and Hypothesis Testing, Chi Square Tests for Goodness-of-Fit and for Independence.
Introduction to the use of the R (Computer Lab Session)

Week 3 Analysis of Variance
One Factor Completely Randomized Design: Underlying Assumptions, Model Building, Among (Between) Treatment Estimate of the Variance, Within Treatment Estimate (Error) of the Variance, Expectations of the Estimators, F statistic, One-Way Analysis of Variance (ANOVA) Table. Interpretation of the p-values. The Principle of Multiple Comparisons.

Week 4 Multiple-Factor (Two-factor) Analysis of Variance
Experimental Design: Factorial Experiments, Two factor Completely Randomized Design, Assumptions, Model Building, Interaction Effects, Analysis of the Multiple-Factor Analysis of Variance Table.

Week 5 Simple Linear Regression/Correlation
Simple Linear Regression and Correlation: Underlying Assumptions, Model Building, Fitting of the Regression Line (Curve), Gauss Markov Theorem, Parameter Estimation, Least-Squares Criterion, Normal Equations, Residuals and their Properties, Estimator of the Conditional Probability Distribution Variance, The Standard Error of the Estimator: Measure of the Absolute Fit, Coefficient of Determination. Measure of the Linear Association (Strength of a Linear Relationship) of a Bivariate Normal Distribution: Pearson's Product Moment Correlation Coefficient.

Week 6 Inferences in Simple Linear Regression/Correlation
Inferences Concerning the Population Y-intercept, the Population Slope, the Mean of the Conditional Probability Distribution Variance, Estimation and Prediction of the Mean,

Week 7 Examples and Exercises.
Inferences Concerning the Population Correlation Coefficient.

Week 8 MIDTERM TEST

Week 9 Model Building in Simple Linear Regression
Steps of Model Building in Simple Linear Regression. A Summary and Interpretation of the Terms used in Simple Linear Regression and Correlation.

Week 10 Easter Holiday

Week 11 Multiple Linear Regression
Multiple Linear Regression; Introduction of more than one Independent (Predictor) Variables, Assumptions, Model Building, Parameter Estimation, Hypothesis Testing in Multiple Linear Regression: Two-Tailed t-tests, Testing the Significance of the Model: Joint F-test, Model Validation: Residual Plots, Durbin-Watson test for First-Order Auto-correlation, NSCORES test for Normality, Tests for Equal Variances.

Week 12 Multiple Linear Regression/Correlation
Multiple Linear Regression and Correlation: Detection and Correction for Multicollinearity, Formal and Informal Methods, Variance Inflation Factor, Coefficient of Multiple Correlation, Coefficient of Multiple Determination, Coefficient of Partial Correlation, Coefficient of Partial Determination.

Week 13 Model Building in Multiple Linear Regression
Steps of Model Building in Multiple Linear Regression. Variable Selection Procedures: Enter, Forward, Stepwise and Backward Regression.

Week 14 Time Series Analysis
Components and Models of a Time Series, Multiplicative, Additive, and Mixed Models of Time Series, Patterns of Secular Trend, Classical Decomposition: Measuring the Trend Component, Measuring the Seasonal Component, Measuring the Cyclical Component, Measuring the Irregular Component (Noise) of a Time Series.

Week 15 Forecasting methods
Time Horizons in Forecasting: Classification of Forecasting Models and Techniques with Respect to Their Properties and Data Patterns, Models for Stationary Processes: The Simple Moving Average Model, the Single Exponential Smoothing, Selection of a Proper Forecasting Model: Forecasting Errors, Mean Absolute Deviation, Mean Square Error, Tracking Signal, Impulse Response and Noise Dampening Abilities. Qualitative Approaching to Forecasting: Delphi, Scenario Writing, Intuitive Approaches.

Week 16 Statistical Decision Theory and Analysis
Structuring a Statistical Decision Problem: Payoff Tables and Regret Tables, Opportunity Loss, Decision Trees, Decision Making under Uncertainty: Optimizing Decision Criteria, Decision Making under Risk: Expected Monetary Value (EMV), Expected Value of Perfect Information (EVPI), Expected Profit under Certainty (EPPP).

Week 17 FINAL PROJECT PRESENTATION