



## Applied Statistics and Probability

4317013 Course of Professor Gilbert Ritschard

Master in International Trading, Commodity Finance and Shipping

---

AppStat, Fall 2013, Info

The aim of this course is to provide the participants with the fundamentals of statistical reasoning as well as with know-how in statistical analysis. The course is end-user oriented and much time will be devoted to the practice of statistics with the help of Excel and R.

As for the covered statistical concepts, the course will, after a quick overview of descriptive statistics, mainly deal with inferential aspects, that is with how we can evaluate the statistical significance or reliability of empirical results. We will show that mastering this concept of statistical reliability is the central key for decision making from data.

Participants will learn how to conduct a full statistical analysis from the data collection and preparation to the presentation and interpretation of the statistical outcome.

Recommended textbook: [Levine et al. \(2008\)](#) and [Maindonald \(2008\)](#).

## Organization of the module

This is a 24 hours module given in 6 sessions of 4 hours each. Sessions are scheduled the Friday, September 20, October 4, 18, 26, November 15 and December 6, 2013. Class will be held at Uni Mail in classroom M-5342, 8:15 - 12:00.

## Evaluation

Homework: Each participant will have to produce a case study using her/his personal data. See the guidelines at the end of this syllabus.

## Contact

Prof. Gilbert Ritschard, Department of Economics, Uni Mail, room M-5232,  
e-mail: [gilbert.ritschard@unige.ch](mailto:gilbert.ritschard@unige.ch), web page: <http://mephisto.unige.ch>.

## COURSE OUTLINE

### 1 Introduction

- 1.1 Why do we need statistics?
- 1.2 Illustrative example: How is Crude Oil Price related to economic fundamentals?

### 2 Understanding data

- 2.1 The data
  - Sources and reliability, level of measurement
  - Databases: individual and aggregated data
- 2.2 Software
  - Basics of database management in Excel
  - Introduction to Visual Basic Programming in Excel (VBA)
  - Introduction to R
  - Hands-on: Retrieving, aggregating and merging data sets
- 2.3 Exploratory univariate data analysis
  - Displaying empirical distributions: summary tables and graphics
  - Summary numbers: central values, dispersion, skewness and kurtosis
  - Detecting and filtering atypical data
  - Modeling data distributions: the normal distribution
  - Comparing empirical and theoretical distributions (qq-plot)
- 2.4 Bivariate data
  - Cross tabulation
  - Displaying bivariate data: the scatter plot
  - Measuring association: covariance, correlation, Pearson Chi-square
- 2.5 Linear Regression
  - Measuring the impact of a covariate: Simple versus multiple regression
  - Interpreting coefficients of a linear relation
  - Making predictions
  - About non linear relations
- 2.6 Hands-on: Exploring real world data series

### 3 Inferential analysis

- 3.1 Sampling and probability
  - Relationship between sample and population
  - Elements of probability
  - Randomness of sample mean and variance
  - Expected value and variance of sample statistics
- 3.2 Three main concepts for inferential purposes
  - Standard error
  - Margin of error
  - Statistical significance
- 3.3 Hands-on: assessing the reliability of sample results

## 4 Assessing regression results

- 4.1 Regression analysis
  - Reliability of parameter estimations
  - Goodness of fit and other validation statistics
  - Residual Analysis
  - Multicollinearity issues
- 4.2 Hands-on: fitting linear models, seeking for most significant factors, assessing model quality (goodness of fit, residuals, collinearity, ...)

## References

- Albright, S. C., W. Winston, and C. Zappe (2008). *Data Analysis and Decision Making with Microsoft Excel (3rd Edition)* (3rd ed.). South-Western College Publishing.
- Levine, D. M., D. F. Stephan, T. C. Krehbiel, and M. L. Berenson (2008). *Statistics for Managers: Using Microsoft Excel* (5th ed.). Upper Saddle River, NJ, USA: Prentice Hall.
- Maindonald, J. and J. Brown (2010). *Data Analysis and Graphics Using R: An Example-based Approach* (3rd ed.). Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
- Maindonald, J. H. (2008). Using R for data analysis and graphics: Introduction, code and commentary. Manual, Centre for Mathematics and Its Applications, Australian National University.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Watson, C. J., P. Billingsley, D. J. Croft, and D. V. Huntsberger (1993). *Statistics for Management and Economics* (Fifth ed.). Boston: Allyn and Bacon.
- Wonnacott, T. H. and R. J. Wonnacott (1990). *Introductory statistics for business and economics* (Fourth ed.). New York: Wiley.

## Guidelines for the Homework

The Home Work can be done alone or by group of two.

Deadline: The report is due January 13, 2014 (midnight). The report should be provided as a PDF file with a name starting with author's name(s).

The report length should be between 10 and 16 pages, plus possibly appendices with extended outputs. Your capacity to synthesize results into easily readable tables and graphics is part of the evaluation. (No copy-pasted text output in the main text!)

A tentative structure could look out as:

1. Introduction (1-2 pages). Description of the considered issue (which variable would you like to explain or predict?), of the hypotheses that you intend to test (provide bibliographical references), and short enumeration of the data and methods that you will use. It is also good practice to announce in the introduction the main findings of your study.
2. Data (1-2 page). You should here give detailed information about your data: sources, concerned population (statistical units, covered period, ...), number of cases, time granularity, precise definition of the variables and their values and, when applicable, applied recoding and filters. A third person should be able to reproduce your analysis and results.
3. Exploratory analysis (2-3 pages). Plots and descriptive statistics of univariate distributions. Comments about these distributions and indication about possible missing values. Descriptive analysis (cross table, scatter plot, measures of association) of the bivariate association between your dependent variable and potential explanatory factors. Comments.
4. Inferential analysis and regression models (2-4 pages). This should be the core of your study. We expect you to run statistical tests (e.g. on difference between mean values, proportion, independence, ...) and fit some regression models (simple and multiple). Explain your modeling strategy (succession of tested models). Presentation of the quantitative results and their statistical validation. Outcomes from the used statistical software should be synthesized. All your regression results should for instance be summarized in a single table.
5. Interpretation and discussion (1-2 pages) of the results (coefficients, statistical significance, plots of residuals, etc.) in regard of the objective of your study.
6. Conclusion (1 page). Summary of the approach followed and of the main findings.
7. Bibliography. Alphabetical list of consulted references.

This structure is indicative, and it may be judicious in some cases to group some sections such as 4 and 5 for example into a same section.

Do not forget to give all necessary labels in tables and plots to avoid any ambiguity. When necessary, you may shortly explain how the plot or table should be read.