Regression analysis and the general linear model

M.S. in Personality and Behavior

Alberto Maydeu Olivares, amaydeu@ub.edu

OBJECTIVES

The main objective of the course is to provide students with a solid statistics-knowledge-net. In Statistics everything is related to everything. By the end of the course students are expected to understand how each technique relates to any other technique, why, and which one to use. Because knowledge nets are most effective when discovered rather than taught, the course is organized as a discovery trip, where each student is to build his or her own net. In this course we also learn how to let the computer do most of the work. Modern statistics use computers and students will learn how to use the statistical package SPSS to perform their statistical analyses. However, statistical packages only do as they are told and students need to know what to ask, how to ask it, and how to interpret the results.

METHODOLOGY

Statistics is a language. Therefore, research questions (which are formulated in words) must be translated into statistics language. Never the other way around. One does not answer only those questions for which has statistical knowledge.

Statistics is a live science, in constant progression. Therefore, when one formulates a question it may be that there is no "right" statistical translation for a it and that new procedure is needed. Even if it exists, it need not be a "standard" technique. Never hesitate to look around and ask around. The sequence of courses in our program does not cover everything. Finally, even if it is a "standard" technique it need not be implemented in the chosen statistical package, but it may have been implemented in other packages.

What it is critical is to understand the basic concepts correctly. For this reason, in this course we will devote our efforts to understand the key concepts in mathematical statistics (e.g., what is a p-value exactly).

This course aims at equating the statistics level of incoming students. Most likely, students come to the program in with a diverse level of statistical skills. For those of you with a low statistical level, this course will be a lot of work as we cover a lot of material.

Yet, the course will focus on concepts. Students are not expected to memorize any formula as these can be looked up from textbooks.

Each week we shall have two types of sessions: 1) Seminars/lectures and 2) computer laboratory. For the seminars, students must read the material covered for each week before class. When reading the materials focus on issues such as

What's the use of this material for psychological research? (what's most/least useful)

How the different procedures relate to each other? And to the previous material?
What's shocking about this? (things I did not know, things I do not know and I'd like to know the answer)
Pitfalls to avoid

Also, bear in mind that the material covered each week is extensive. Be synthetical. Concentrate on the basics (this course's lemma).

During the seminars I will summarize the theory to be covered each week. Be as participative as you can. During seminars some use of the statistical software (SPSS) will be made and exercises will be assigned each week. Then, during the computer lab, students will have a further opportunity to use the software and ask questions. Weekly exercises are to be turned during the next seminar.

Students are encouraged to bring laptops to reproduce what we do in class.

BASIC TEXT AND MATERIALS

REQUIRED TEXT

Norusis, M. J. (2006). *SPSS 15. Statistical procedures companion*. Upper Saddle River, NJ: Prentice-Hall.

(any version of the book will do 12 to 17)

RECOMMENDED TEXTS

Depending on your prior statistical level we recommend that you also use either

McClave, J.T. & Sincich, T. (2006). *Statistics* (10th ed). Upper Saddle River, NJ: Prentice Hall.
(introductory text)
The course assumes you know chapters 1 to 9.

Fox, J. (2008). *Applied Regression Analysis and generalized linear models* (2nd ed) Thousand Oaks, CA: Sage
(more advanced text)

PRE-REQUISITES

⇒ Chapters 1 to 8 of the Norusis book.

These materials are not part of the program. We will assume you have mastered them before the course begins. Also, we will assume you know how to use SPSS.

Also, should you decide to use McClave and Sincich's book, I urge you to read Chapters 1 to 9 before the course starts. The course assumes you know these materials. Nevertheless, we will cover them briefly in the first week.

FURTHER REFERENCES

A favourite of students instead of McCabe or Fox is

Field, A.(2005). *Discovering statistics using SPSS* (2nd ed). London: Sage.
(irreverent but innacurate at times)

For your applied work, you may need a more in-depth coverage of the material. To that aim, the following list of textbooks is given. Within each topic, the books are arranged in order of difficulty:

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for behavioral sciences* (3rd edition). Mahwah, NJ: Lawrence Erlbaum Associates.
(basic and very comprehensive, few formulae, very good source of interpretation of interaction models)

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.
(a must read)

Kirk, R.E. (1994). *Experimental design: Procedures for behavioral sciences*. (3rd ed) Wadsworth Publishing.
(everything you need to know about designing experiments)

Kutner, M.H., Nachtstein, C.J., Neter, J. & Li, W. (2005). *Applied linear statistical models*. (5th ed) Boston: McGraw Hill.
(very comprehensive)

Draper, N.R. & Smith, H. (1998). *Applied regression analysis*.(3rd ed) New York: Wiley.
(a little more advanced)

PROGRAM

Week 1
Gettting started with SPSS
Descriptive statistics: Means, medians, percentiles, confidence intervals for means, crosstabulations
Graphs: Histograms, bar charts, scatterplots, box and whisker plots
Obtaining z-scores

Week 2
Assessing normality
Correlations
Working with SPSS: Adding files, matching files, selecting cases, transforming variables, recoding categories, importing and exporting data

Week 3
Some terminology: parameters, statistics, …
Estimation methods
    Small sample properties and large sample properties
Confidence intervals
Testing hypotheses
Testing means
    1 population
    2 populations
    testing proportions
    2 dependent samples

Week 4
Regression analysis
    Estimation
    Parameter interpretation
    Using regression analysis for prediction
    R^2 and goodness of fit
    Assumptions and model diagnostics
Multiple regression
    F tests
    Parameter interpretation
        Standardized coefficients
        Bonferroni corrections
Hierarchical regression

Week 5
Regression with qualitative predictor variables
    Dummy coding
    Interactions with quantitative predictors
    Relationship to ANOVA and ANCOVA
    Indicator variables vs. allocated codes
Model selection
    All possible regressions
    Backward

Forward and stepwise
Model selection
R^2 criterion
R_a^2 criterion

Week 6
Polynomial regression
Regression with interactions among quantitative predictors
Multicollinearity
Mean centering and standardization
Advanced topics
Constrained regression
Piecewise regression

Week 7
Assumptions
Residual plots
Partial residual plots
Variance stabilizing transformations
Logarithmic transformations
What to do when assumptions are blatantly violated
Outliers and influential observations
Identifying outlying y observations
Identifying outlying x values
Identifying influential cases
Masking
Advanced topic: logarithmic transformations

Week 8
What's the difference between the analysis of variance and regression models?
ANOVA terminology
Fixed effects vs. random effects
Parameterizations: Regression, cell means, factor means
Implementation
Contrasts, multiple comparisons
Two-way ANOVA
MANOVA

Week 9
Binary logistic regression
Likelihood ratio test and Wald test
Parameter interpretation
Classification
Multinomial logistic regression
Parameter interpretation

Week 10
Exam / Buffer

EVALUATION CRITERIA

Grades will be assigned as follows:

Class discussions and assigned exercises	50%
Final exam	50%

Every 2 weeks you will be assigned exercises due the next week. At the end of the course there will be a short in-class exam where we will provide you some with some questions and SPSS output and you will be asked to answer some questions. The exame is open book, open notes, open everything.