Course Syllabus and References

Overview of Modelling

***Lecture 0*:** Overview of experiments and studies; Goals of modelling; General form of statistical models; Types of response variables and classes of models

References: General Statistics, Modelling, and Software Guides

Agresti, A. and Finlay, B. (1997). *Statistical Methods for the Social Sciences (3rd ed.)*. New Jersey: Prentice Hall. (A popular, clear, and comprehensive introduction to probability and statistics, requiring only a basic grounding in mathematics.)

George, D. and Mallery, P. (2001). *SPSS for Windows Step by Step: A Simple Guide and Reference, 10.0 Update (3rd ed.)*. Boston: Allyn and Bacon. (Provides a substantial and straightforward introduction to statistics and various data analysis methods and clearly explains how to implement those methods in SPSS. The first 16 chapters cover basic methods and require no knowledge of statistics but some grounding in mathematics; the last 12 chapters present more advanced methods.)

Moore, D. and McCabe, G. (1998). *Introduction to the Practice of Statistics (3rd ed.)*. Freeman. (A straightforward and example filled introduction to probability and statistics/data analysis, requiring only a basic grounding in mathematics.)

Norusis, M. and Norusis, J. (2000). *SPSS 10.0 Guide to Data Analysis*. New Jersey: Prentice Hall. (A good guide to using SPSS to perform basic data analysis. Provides overviews of various statistical methods, but is best used as a supplement to a good statistics text.)

Ramsey, F. and Schafer, D. (2001). *The Statistical Sleuth: a Course in Methods of Data Analysis (2nd ed.)*. Belmont, CA: Duxbury. (A comprehensive (800+ pp.) and example-filled guide to modelling topics from linear models to generalized linear models. Requires only a basic knowledge of statistics and mathematics.)

Rice, J. (1995). *Mathematical Statistics and Data Analysis (2nd ed.)*. Belmont, CA: Duxbury. (Very highly-regarded mathematical introduction to probability and statistics. Full comprehension requires knowledge of multivariable calculus.)

Venables, W. and Ripley, B. (1999). *Modern Applied Statistics with S-Plus (3rd ed.)*. New York: Springer-Verlag. (A comprehensive and extremely popular, but fairly high level, guide to S-Plus's capabilities and its (command line) usage. Includes a basic introduction to S-Plus; subsequent chapters cover using S-Plus for tasks ranging from linear models to survival analysis. Comprehension requires a solid background in statistics and some familiarity with programming.)

Linear Models

***Lecture 1 [Overview of Linear Regression]*:** General concept; Basic model setup and assumptions; Least squares estimation

***Lecture 2 [More Details of Linear Regression]*:** Residuals; R^2, total variance, and conditional variance; Coefficients and their properties; Predicted values and their properties; Checking assumptions and goodness of fit (overall diagnostic plots and statistics); Outliers and influential points (case statistics); Comparing models

***Lecture 3 [Problems with Linear Regression and Extensions/Alternatives]*:** Transforming and adding variables; Generalized (weighted) least squares regression; Robust regression; Resistant regression; Multicollinearity and solutions; Non-parametric regression

References: Linear Models

Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: John Wiley. (A popular, comprehensive (600+ pp.) text on regression: covers linear and nonlinear regression, non-parametric and semi-parametric methods, and GLMs. Applications oriented: focuses on seeing results through graphs and on computer implementation using ARC, a freeware regression package. Text requires only a basic grounding in statistics.)

Draper, N. and Smith, H. (1998). *Applied Regression Analysis (3rd ed.)*. New York: John Wiley. (An elementary to intermediate level, very comprehensive (700+ pp.), popular text on regression. Covers linear and nonlinear regression, including recent advances. Requires only a basic grounding in statistics.)

Hamilton, L. (1992). *Regression with Graphics: a Second Course in Applied Statistics.* Belmont, CA: Duxbury. (An accessible and short regression text that covers the basics of regression and extensions such as curve fitting, robust regression, and logistic regression. Comprehension requires only a basic grounding in statistics.)

Rawlings, J. (1988). *Applied Regression Analysis: A Research Tool*. Pacific Grove, CA: Wadsworth and Brooks. (A more mathematically advanced, traditional text on regression. Covers regression basics and diagnostics, transformation of variables, and collinearity. Focuses on applications: does a very good job of presenting examples and describing how to interpret results and diagnostics. Understanding the presentation of methods requires a good background in mathematics (i.e., knowledge of matrix algebra) and basic knowledge of statistics; however, the presentation of applications is very accessible.)

ANOVA

***Lecture 4*:** One-way ANOVA and Kruskal-Wallis; Two-way ANOVA; Multiple comparisons and contrasts; Analysis of Covariance; Repeated measures ANOVA (MANOVA)

References: ANOVA

Box, G., Hunter, J. S., and Hunter, W. (1978). *Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley. (The extremely popular, canonical text for anyone designing and analysing scientific experiments. Thought to be very accessible for non-mathematicians.)

Christensen, R. (1996). *Analysis of Variance, Design and Regression: Applied statistical methods*. New York: Chapman & Hall. (A modern applied look at the application of ANOVA techniques. Reasonably accessible for non-mathematicians.)

Cox, D. R. (1958). *Planning of Experiments*. New York: John Wiley.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods: Kendall's Library of Statistics 5*. London: Arnold.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. New York: Prentice Hall.

Scheffé, H. (1959). *The Analysis of Variance*. New York: John Wiley. (A classic but very theoretical text that requires a solid mathematical background. Not well suited for an introduction into the subject.)

Generalized Linear Models

***Lecture 5 [The GLM Framework]*:** Motivation; Basic GLM setup and assumptions; Coefficients and their properties; Scale parameter and overdispersion; Goodness of fit and comparing models; Residuals, case statistics, and diagnostic plots

***Lecture 6 [Frequently used GLMs]*:** Logistic regression; Poisson/Log-linear models; Gamma model; Multiple Logistic Regression and Proportional Odds Model

References: GLM

Dobson, A. (1990). *An Introduction to Generalized Linear Models,* (*2nd ed.*). Boca Raton, FL: Chapman & Hall/CRC. (A good introductory text on the subject).

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (2nd ed.)*. London: Chapman and Hall. (Considered the canonical GLM text by statisticians. Requires a very solid understand of mathematics, statistics, and linear modelling; not particularly accessible or application-oriented. A good reference, but not introductory, text.)

Multivariate Response Modelling (esp. Longitudinal Data)

***Lecture 7*:** Fixed effects; Random effects

Lecture 8: Generalized linear mixed models; Generalized estimating equations

References: Longitudinal/Multivariate Data Modelling

Diggle, P., Liang, K., and Zeger, S. (1994). *The Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Pinheiro, J. and Bates, D. (2000). *Mixed Effects Models in S and S-Plus*. New York: Springer-Verlag.

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling.* London: Sage.