The University of British Columbia  
Faculty of Forestry  
Forestry 530 -- Advanced Regression Analysis  
Course Outline for 2011

**Calendar Description:**  
**NOTE:** unlike the description, we will cover both least squares and maximum likelihood methods.

**Instructor:** Dr. Val LeMay  
FSC 2039  
822-4770  
Valerie.LeMay@ubc.ca

**Graduate Teaching Assistant:** Suborna Ahmed

**Meeting Times:**  
**Classes:** Monday, Wednesday, Friday  
1200 to 1300  
FSC 1003

**Drop in Instructor Office Hours:** Monday and Wednesday, 2:00 to 4:30 pm  
(will be a limited days where these are cancelled)

**Weeks of March 21 and April 4:** project appointments

**Classes Cancelled:**  
February 16 – 20 (Reading Week)  
February 23 and 25 (meeting off-campus)  
March 28, 30 and April 1 (project time)

**Evaluation:**  
Assignments: 35  
Project: 25  
Project Outline due Friday, March 11; Project is due Friday, April 15  
[subject to changes due to scheduled exam]  
Final Exam: 40  
(During April Exam Period)

**AUDITORS:** Doing the assignments is an important part of the learning process. Please hand in each assignment, and we will provide some comments on these.

**Course Content** *(NOTE: content and schedule may vary depending upon interest and previous backgrounds of students—to be discussed)*

**Part I: Background**

1. Overview of models, introduction to notation [January 5]

2. Quick review of some elements of statistics, with emphasis on probability distributions and small versus large properties of statistics (consistency, accuracy, precision, bias, maximum likelihood, and asymptotic properties). [January 5, 7]

3. **SAS:** introduction  
SAS for bringing in and manipulating data, and for statistics calculations  
[Computer lab, FSC 1404, January 10 and 12]
Part II. Least Squares Methods

4. Linear regression using least squares (Ordinary Least Squares, OLS, includes simple and multiple linear regression) [January 14, 17, 19, 21, 24, 26]
   a. Notation. Principles and assumptions. Effects of not meeting the assumptions.
   b. Measures of fit, hypothesis testing, and confidence intervals.
   c. Ways to correct for problems using transformations.
   d. PROC REG and GLM in SAS.
   e. Using class variables as predictors
   f. Conditioned (restricted) regression

Assignment 1 (10 marks): Ordinary least squares using SAS; Regression using class variables. Due Monday January, 24.


5. OLS using matrices: Matrix algebra review (practices exercises); Matrices for OLS (example using two predictor variables). [January 28, 31, Feb 2]

6. Generalized and estimated generalized least squares (linear models; GLS and EGLS): [February 2, 4, 7]
   a. Principles of GLS and EGLS
   b. The error covariance matrix under:
      i. No restrictions
      ii. OLS assumptions
      iii. Heteroskedasticity only: Small and large sample properties, weighted least squares in PROC REG

Assignment 3 (10 marks): Fitting equations when error terms are heteroskedastic, beginning with an OLS fit, tests of assumptions, and using weighted least squares under PROC REG. Due Friday, February 11

   iv. temporal correlation (serial (also called auto-) correlation in time): Small and large sample properties, first order autocorrelation and other time series patterns, PROC AUTOREG for first order autocorrelation [examples in class]
   v. heteroskedasticity and serial correlation for clustered objects – grouped by subject (called cross-sectional, longitudinal, panel, repeated measures, etc, data depending upon the discipline): small and large sample properties, examples.


Assignment 4 (10 marks): Fitting nonlinear equations using PROC NLIN [Due Monday, February 28]

[NOTE: February 16 to 20 – no classes – Reading Week]
Part III. Maximum Likelihood Methods

8. **Maximum Likelihood:** Simple examples to illustrate the method, small and large sample properties. [February 21]

    [NOTE: February 23 and 25 – no classes – readings assigned]

9. **Generalized linear models:** Error terms follow normal or other distributions.
   a. Principles of Generalized linear models: [February 28]
      i. Model, distribution, and link function.
      ii. Small and large sample properties
   b. **Logistic (logit) analysis** to predict the probability of an event using PROC LOGISTIC and PROC GENMOD in SAS. [February 28, March 2, 4]
      i. Maximum likelihood, dist=binomial, link=logit.
      ii. Hypothesis tests
      iii. Fit statistics including percent concordance, discordance, Pseudo-R square, etc.
      iv. The classification table.
   c. **Poisson Regression for count data** and other examples using different distributions and transformations (link functions) using PROC GENMOD in SAS. [March 4]

Assignment 5 (10 marks): Fitting logistic and Poisson models and interpreting results [Due Friday, Wednesday, March 9]

10. **Mixed linear models:** Models with fixed and random components using PROC MIXED of SAS [March 7, 9, 11, 14]
    a. Principles of linear mixed models: General model.
    b. Small and large sample properties
    c. Fitting using maximum likelihood and residual estimation maximum likelihood.
    d. Examples of use:
       i. Mixed-effects factors in experimental design
       ii. heteroskedasticity;
       iii. serial correlation: temporal or spatial correlation
       iv. hierarchical models: Partitioning the Error Covariance Matrix.
       v. Observations grouped by subject (clustered data)
          1) Heteroskedasticity and/or autocorrelation
          2) Random coefficients modeling

Assignment 6 (10 marks): Fitting mixed models and interpreting results [Due Friday, March 18]

11. **Mixed generalized linear and nonlinear models** Mixed models where the base model is generalized linear (PROC GLIMMIX in SAS) or nonlinear (PROC NLMIXED in SAS). [March 16, 18, 21]

Assignment 7 (10 marks): Fitting mixed generalized linear and nonlinear models and interpreting results [Due Friday, March 25]
**Part IV. Systems of Equations**

12. **Systems of Equations** [March 23, 25]: (PROC MODEL in SAS) Overview only
   a. Least squares (seemingly unrelated regressions; two stage and three stage least squares, and multi-stage least squares)
   b. Maximum likelihood (limited information maximum likelihood and full information maximum likelihood).

   **[NOTE: March 28 through April 3 – no classes – Project Time]**

**Part V. Review**

Review: Exam questions practice [April 4, 6 and 8]

**Textbooks and other references**

Kutner, and others. Applied linear statistical models, 5th edition. Available from the UBC bookstore for about $200 CDN plus tax. Excellent reference for about half the course [There are a few copies in the UBC library system]

Littell, et al. SAS for Mixed Models, 2nd edition. [you can purchase this on-line from SAS or other bookstores such as Amazon.com]. This is a VERY good book with both SAS code and theory of mixed models.

Pinheiro, J.D. and D. M. Bates. 2000. Mixed-effects models in S and S-Plus. Springer, New York. This is THE classic text on mixed-effects model – heavy reading, but the basis for most mixed-models theory, and their examples are the ones used in R code.


**e-books**


You will find other e-books texts in the “Springer” and “StatsNetBase” indexes, on any kind of modeling you would like to learn about.
Project Guidelines

Score: 5 (outline) + 15 (project) = 20 (maximum score)

Objective: The project part of the requirements for completion of this course is intended to give you the opportunity to use the tools presented in the course on a problem that is of interest to you.

Length: The project should be from seven to 12 typewritten pages.

Topic: The subject of the project should be a research project that involves analysis of some data of interest to you using the techniques learned in the course.

Format: The project must be written as a formal research report with:
- a title page,
- introduction (including relevant background literature and the objectives for the paper),
- materials and methods,
- results,
- discussion (connect your results to previous research),
- conclusion,
- references cited (any format you wish, but MUST be consistently used), and
- any Appendices

Relevant SAS outputs should be brought in as formatted figures and tables, with captions (at the bottom for figures and at the top for tables and these must be referenced in the text. Large outputs that are relevant should be put into formatted, labeled Appendices. This must be original work, with due credit to any reference material you have used.

Project Outline: To obtain some early feedback on your chosen project, an outline for the project is due by Friday, March 11. In the outline, indicate:
- The objective of the research
- The data you will use
- The methods you might use

I will give you feedback on the project and let you know if this is too much for the project. The outline is worth 5 points towards a total maximum project score of 20 points.

PROJECT OUTLINE: Due Friday, March 11
PROJECTS DUE: Friday, April 15 (subject to exam schedule)