STATISTICS 333, Fall 2003

APPLIED REGRESSION ANALYSIS

INSTRUCTOR: Greg Reinsel Office: 4361 Comp. Sci.-Stat. Phone: 262-0087,
E-mail: reinsel@stat.wisc.edu http://www.stat.wisc.edu/∼reinsel

TEXTBOOK: Draper and Smith, Applied Regression Analysis, 3rd edition, Wiley, 1997

OTHER BOOKS: Myers, Classical and Modern Regression with Applications, Duxbury, 1986
Ryan, Modern Regression Methods, Wiley, 1996

COURSE REQUIREMENTS:

The grade in the course will be determined by two midterm exams each counting 25% toward the final grade, homework assignments and projects counting about 20%, and a final exam counting about 30%. Course content based on lectures, handouts, and textbook. Assignments and projects will consist of problems taken from the above textbook, as well as additional assigned exercises including the analysis of various data sets. Minitab (also Splus and R) computer package will be discussed and used in course, but a student may use any package they are familiar with in doing required assignments and project work.

Tentative Outline and Schedule of Topics

1. Simple Linear Regression (7 lectures)
   Basic concepts of linear regression relationships, linear regression model for response with a single predictor variable, fitting a straight line by least squares, estimation of error variance, confidence intervals and tests for slope and intercept coefficients; correlation; assessment of the model, examining residuals, lack of fit and pure error.

   Draper and Smith, Sections 1.0–1.7, 2.1–2.6, 3.0–3.2
   Myers, Sections 1.1–1.2, 2.1–2.8

2. Regression Model and Analysis in Matrix Terms (10 lectures)
   Matrix formulation and approach to fitting a straight line, extension to general linear regression models with several predictor variables; basic statistical properties of least squares (LS) estimators, standard errors and t-statistics, correlations among LS estimates, partitioning of variability and analysis of variance, sequential and partial sums of squares of the regression, assessing contributions of additional predictor variables, testing linear hypotheses and the "extra sum of squares" principle, F-statistics, confidence intervals, prediction intervals.

   Draper and Smith, Sections 4.1–4.7, 5.1–5.4, 6.1–6.3, 6A
   Myers, Sections 3.1–3.8

   MIDTERM EXAM 1

3. More Special Topics on the Multiple Linear Regression Model (6 lectures)
   Orthogonality of regressor variables, results on introducing further regressors; issues relating to increasing number of predictors (e.g., multicollinearity, variance inflation factors); testing general linear hypotheses, reduced models and their residual sums of squares; bias in LS estimates; expected mean squares; weighted least squares and generalized least squares.

   Draper and Smith, Sections 9.1–9.5, 10.1–10.4
   Myers, Sections 4.3, 6.1

4. Checking Adequacy of Regression Model Fit (3 lectures)
   Methods of examination of residuals for model adequacy, various residual plots, normal probability plot; standardized and studentized residuals, leverage and influence diagnostics, detection of outliers, Cook's statistics; tests for autocorrelation, Durbin–Watson statistic.

   Draper and Smith, Sections 8.1–8.4, 7.1–7.2
   Myers, Sections 5.1–5.8, 6.4, 8.1–8.5

5. Polynomial and Other More Complicated Regression Models (5 lectures)
   Polynomial models with one or more predictor variables, response surface models, retaining terms in polynomial models (e.g., origin-shift criterion); Box–Cox power transformation for response variable, variance-stabilizing transformations, intrinsically linear versus nonlinear models.

   Draper and Smith, Sections 12.1–12.3, 13.1–13.2, 13.6, 24.1
   Myers, Sections 6.2–6.3

6. Use of Indicator Variables in Regression Models (3 lectures)
   Indicator variables to represent "block" effects and main effects of levels of other categorical variables, interaction terms involving indicator variables; straight line segmented models, two-phase change-point models.

   Draper and Smith, Sections 14.1–14.3

   MIDTERM EXAM 2

7. Model Selection Techniques (3 lectures)
   Criteria for selection of best regression model, R^2, adjusted-R^2, Mallow's C_p, mean square error (MSE); regression model selection procedures: all subset regressions, best subsets, backward elimination, forward selection, stepwise regression.

   Draper and Smith, Sections 15.0–15.6
   Myers, Sections 4.1–4.5

8. Issues of Ill-Conditioning and Multicollinearity (2 lectures) – as time permits
   Measuring multicollinearity, centering and scaling data, correlation matrix of predictor variables, variance inflation factors (VIFs), ridge regression.

   Draper and Smith, Sections 16.1–16.5, 17.1–17.3
   Myers, Sections 7.1–7.4

9. Linear Regression Applied to Experimental Design Problems (3 lectures) – as time permits
   One-way classification, two-way classification with equal or unequal numbers of observations in the cells, main effects and interaction effects of factors, (orthogonal) linear contrasts in factor effects.

   Draper and Smith, Sections 23.1–23.10

   FINAL EXAM