BIOS 762: Theory and Applications of Linear and Generalized Linear Models
Fall, 2015

Texts:
1. Christensen, R. (2011). Plane Answers to Complex Questions: The Theory of Linear Models, 4th Edition, Springer-Verlag (required).
2. Agresti, A. (2012). Categorical Data Analysis, 3rd edition, New York: Wiley (required).
3. Stokes, M. E., Davis, C. S., and Koch, G. G. (2000) Categorical Data Analysis Using the SAS System (required).
4. Thompson, L. (2006). S-plus and R Manual to Accompany Agresti's Categorical Data Analysis (2002), Available at \http://www-stat.stanford.edu/~owen/courses/306a/Splusdiscrete2.pdf" (recommended).
5. Generalized Linear Models, 2nd Edition, by McCullagh and Nelder. Publisher: Chapman and Hall (recommended).

Class notes will also be distributed in class. The notes will cover a lot of material not contained in the textbooks.

Prerequisites: Bios 661, 663, Math 547, 417, or 577 and currently enrolled in Bios 760.

Homework: There will be 7 homework assignments. The problems will come from the books and separate handouts.

Computing: There will be a lot of computing in this course. We will primarily use the software packages SAS and R.

Exams: There will be one midterm exam (Wednesday Oct 8) and a comprehensive final exam on the date University specifies. Absolutely no makeup exams will be given.

Grading:
Homework 30%
Midterm 35%
Final exam 35%

Course Goals:
1. Propose, explain, and correctly use linear models and generalized linear models (GLMs) for various data from cross-sectional and longitudinal studies;
2. Know how to estimate parameters and make statistical inference using linear models, GLMs, and quasi-likelihood models;
3. Evaluate and compare different models and different statistical methods for the same data sets;
4. Get familiar with linear models and GLM procedures for some commercial packages, such as SAS and R;
5. Read statistical research publications, analyze data using linear models and GLMs, compare different linear models and GLMs and their extensions, and use these models to answer specific scientific questions.

Course Outlines:

1. Week 1
   - Motivating examples for linear models in biostatistics, including regression analysis, analysis of variance, longitudinal data and repeated measures
   - Vector spaces, subspaces, orthogonal bases, Gram-Schmidt decomposition, column space of a matrix. Review of matrix and vector operations and linear algebra, including trace, rank, inverse, determinant, null space, eigenvalues, orthogonal complement

2. Week 2
   - Matrix decompositions including spectral, singular value, QR, idempotent matrices, projection operators
   - Estimability in the univariate linear model, least squares, BLUE, maximum likelihood, Gauss-Markov theorem, applications

3. Week 3
   - Multivariate normal distribution, distribution of quadratic forms
   - UMVUE, sampling distributions of estimates, weighted least squares, normal equations

4. Week 4
   - Breaking up sums of squares, testing linear hypothesis and generalizations.
   - One-way anova, estimating and testing contrasts, two-way anova with and without interaction, contrasts, applications

5. Week 5
   - Unbalanced two-way anova, higher way anova, polynomial regression
   - Experimental design models, including completely randomized design, randomized complete block design

6. Week 6
   - The Exponential Family and introduction to generalized linear models (GLMs): definitions, estimation methods

7. Week 7
   - Introduction to generalized linear models (GLMs): likelihood theory, deviance
   - GLMs for continuous response: gamma regression model, inverse Gaussian regression

8. Week 8
   - Generalized Linear Models for Categorical Responses I: categorical response data, distributions of categorical variables, contingency tables and sampling methods, models for binary and binomial responses

9. Week 9
   - Generalized Linear Models for Categorical Responses II: models for polytomous and ordinal responses, Poisson regression, loglinear models for contingency tables

10. Week 10
    - Eliminating Nuisance Parameters: hypergeometric distributions, conditional logistic regression, conditional likelihood methods

11. Week 11

- Over- and Varying-Dispersion Models: models for over-dispersion, score tests for over-dispersion, the heteroscedastic linear model

12. Week 12
    - Theory of the multivariate linear model, multivariate analysis techniques.

13. Week 13
    - Generalized Estimating Equations (GEE), Quasi-likelihood, Z-estimators

14. Week 14
    - Diagnostics and Influence Analysis: diagnostic measures, generalized leverage, goodness of fit statistics.