# Medical Biometry III
# Biostatistics 513
# Spring 2001

Lecture: MWF 9:30 - 10:20am, T439


**Instructor:**

Patrick J. Heagerty, PhD
Associate Professor
Department of Biostatistics

**Office:**

F667 (F-wing, 6th floor)

**Phone:**

(206) 616-2720

**e-mail:**

heagerty@u.washington.edu

**Course Objectives:**
This course introduces the principles and methods of statistical inference for categorical data and survival data. The major topics covered are: contingency table methods; logistic regression; Kaplan-Meier and log-rank methods; and Cox regression.

Upon entering this course the student is expected to be able to:

- use a statistical package (STATA) to enter data and calculate summary statistics.
- perform two-sample hypothesis tests for binary data (proportions).
- compute and interpret confidence intervals for one-sample and two-sample situations.
- perform "multivariable" (KKMN) linear regression and interpret model coefficients, compute confidence intervals for model coefficients, and test hypotheses using the partial F test.

Upon completing Biostat 513 students can expect to be able to:

- perform standard tests of homogeneity and trend with data from $2 \times C$ tables.
- compute odds ratio estimates and confidence intervals from $2 \times 2$ and stratified $2 \times 2$ tables.
- perform logistic regression to estimate model coefficients and odds ratios (adjusted), compute and interpret coefficient and odds ratio confidence intervals, and test hypotheses that one or more coefficients in the logistic regression model are zero.
- compute and interpret the product limit estimate (Kaplan-Meier) estimate of survival and associated confidence intervals.
- perform and interpret the log-rank test for differences between survival curves with right censored survival data.
- perform Cox regression to estimate proportional hazards model coefficients, interpret coefficient estimates and confidence intervals, and test hypotheses that one or more coefficients in the regression model are zero.

- interpret and critique the results of application of these statistical techniques as found in the health sciences literature.

**Lectures:**
Lecture notes will be distributed prior to each lecture with additional copies made available in the Biostatistics department (F-wing, 6th floor). Lectures will be videotaped and made available at the HSL reserve desk.

**Discussion Section:**
The purpose of discussion section is to provide feedback on the practice of the methods introduced in class. We will concentrate on a data analysis using logistic regression (3 weeks) and a data analysis using survival analysis techniques such as proportional hazards regression (3 weeks). The scheduled times are:

| | | | | | |
|---|---|---|---|---|---|
| (AA) | Thursday | 8:30pm | - | 9:20pm | T531 |
| (AB) | Wednesday | 11:30am | - | 12:20am | T531 |
| (AC) | Wednesday | 10:30am | - | 11:20am | T531 |

**Exercise Sets**:
Weekly exercise sets will be assigned, collected, and graded. These exercises are designed to provide practice and feedback. Late assignments will not be accepted.

**Text:** (from 511 & 512)
*Fundamentals of Biostatistics*, 4th Edition, Bernard Rosner, Duxbury

*Applied Regression Analysis and Multivariate Methods*, 3rd Edition, Kleinbaum, Kupper, Muller and Nizam, Duxbury

⋆ *Logistic Regression*, D.G. Kleinbaum, Springer.

⋆ *Survival Analysis*, D.G. Kleinbaum, Springer.

**Statistical Computing:**

STATA (student version available – see web page) either version 6.0 or 7.0 should be adequate.

**Computer Lab:**
Access is provided to the PCs in the Micro Computing Laboratory in the Health Sciences Library. The lab is acessible whenever the library is open, except the lab closes one-half hour earlier than the library.

There are no reserved times for the HSL computing room but a teaching assistant will be available for consultation at regular times (see TA schedule).

**Course Web Page**:
A course web page will be maintained to provide access to data sets and other course materials.

**Office Hours:**

Patrick Heagerty:

- Wednesday 10:30am - 12:30pm, F-667
- Friday 10:30am - 11:30am, F-667
- ○ other times by appointment

**Teaching Assistants:**

| | office hours | room |
|---|---|---|
| Heidi Caron | | |
| London Fairborn | | |
| Holly Janes | | |
| Zheng Zhang | | |

**Grading:**
30% Midterm Exam
30% Final Exam
25% Exercise Sets
15% Data Analysis Project

• The midterm and final exams will be in-class exams. A list of key formulas will be provided with each exam. Exams are closed-book/notes.

• The exercise sets are crucial to mastering the material. Exercises provide an opportunity to obtain feedback from the instructor and teaching assistants.

• The data analysis project will be a 5-page maximum written summary of one aspect of analysis for the data sets discussed in the discussion section.

**Lecture Notes:**
The lecture notes are based on material prepared originally by Dean Patricia Wahl, later modified by myself, and then modified by Professor Norm Breslow. The lecture notes are prepared in advance and will either be made available at the copy center, or distributed in class.

**Tentative Course Outline**

| Date | Topics | Reading | Discussion Section |
|---|---|---|---|
| M 3/26 | Introduction / Organization / Goals / $R \times C$ tables | | none |
| W 3/28 | Review of categorical methods for $2 \times 2$ tables | R:10.1-10.8 | |
| F 3/30 | Paired Binary Data / Fisher's exact test | | |
| M 4/2 | Testing for trend / Stratification and confounding | R:10.8-10.9 | |
| W 4/4 | Mantel-Haenszel procedures | R:10.10 | $R \times C$ and |
| F 4/6 | Effect modification and homogeneity testing | R:10.11 | $2 \times 2$ tables |
| M 4/9 | Review of linear regression methods | R:11.1-11.13 | |
| W 4/11 | Introduction to logistic regression | R:11.14, HL:1 | Logistic regression |
| F 4/13 | Logistic regression with a single categorical predictor | S:7 | - descriptives |
| M 4/16 | Logistic regression with a single continuous predictor | HL:3 | - bivariate |
| W 4/18 | Logistic regression with two covariates: Interpretation & testing | S:8 | Logistic regression |
| F 4/20 | Logistic regression with two covariates: Confounding | K1:7 | - multivariate |
| M 4/23 | Logistic regression with two covariates: Interaction | KKMN:23 | - fit / interpret |
| W 4/25 | Logistic regression with multiple predictors | HL:2 | Logistic regression |
| F 4/27 | Methods for assessing model adequacy | HL:5 | - model checking |

## Tentative Course Outline

| **Date** | **Topics** | **Reading** | **Discussion Section** |
|---|---|---|---|
| M 4/30 | Issues with model building | HL:4 | - conclusions |
| W 5/2 | **EXAM** | | |
| F 5/4 | Logistic regression for case-control data | HL:6 | analysis |
| M 5/7 | Person-time data: rate estimation and comparison | R13.1-13.5 | Survival analysis |
| W 5/9 | Introduction to survival analysis: censoring, survival, hazard | R13.6-13.7 | - descriptives |
| F 5/11 | Nonparametric survival estimation: Kaplan-Meier | PM:2, K2:2 | - bivariate |
| M 5/14 | Comparing survival curves: log-rank test | R13.8, PM:4-5 | Survival analysis |
| W 5/16 | Proportional hazards model | R13.9 | - multivariate |
| F 5/18 | Cox regression with a single covariate | PM:6 | - fit / interpret |
| M 5/21 | Cox regression with multiple covariates | K2:3, MV:6 | Survival analysis |
| W 5/23 | Cox regression: confounding and interaction | PM:7 | - model checking |
| F 5/25 | Methods for assessing model adequacy | K2:4 | - conclusions |
| M 5/28 | HOLIDAY | | Review / |
| W 5/30 | Other approaches to survival data: parametric, additive | PM:3, MV:5 | Questions |
| F 6/1 | Review | MV:8 | |

Final Exam: Wednesday June 6, 1999 8:30am-10:20am