

Health Studies 32100
Introduction to Biostatistics
Summer 2007

Instructor: James J. Dignam, Ph.D., AMB Building W259, 834-3162, (jdignam@health.bsd.uchicago.edu).
Office hours: Tues. 10:30am–12:00pm or by appointment. E-mail is an effective way to reach me with requests for meetings, questions about the material, etc.

Course assistant (CA): Ms. Jiyoung Kim (jiyoungk@uchicago.edu). You are welcome to e-mail her with questions.

Department contact: Ms. Jessica Miotk, Department Project Assistant, 834-4056

Course website: From the Health Studies website (<http://health2.bsd.uchicago.edu/>), a link to the course website can be found via the instructor's homepage or the list of courses.

Times and Place:

Lectures: Tues-Wed-Thurs 3:00–4:20pm, AMB Building W-229

(optional) Computer lab session: Wed 10:00–11:30am, Harper 406

CA office hours:

Note: By arrangement, the CA can also come to your office to help with computer problems if necessary. Also, if you have a laptop that you do your work on, please feel free to bring it to meetings.

Prerequisites: College algebra and ability to use a personal computer.

Description: This course will provide an introduction to the basic concepts of statistics as applied to the biomedical and public health sciences. Emphasis is on the use and interpretation of statistical tools for data analysis. Topics include (i) descriptive statistics; (ii) probability and sampling; (iii) the methods of statistical inference; and (iv) an introduction to linear and logistic regression.

Texts: (University Bookstore)

Required: Pagano M and Gauvreau K. *Principles of Biostatistics*, 2nd Ed., 2000, Duxbury Press, Pacific Grove, CA.

References:

Daniel WW. *Biostatistics: A Foundation for Analysis in the Health Sciences.*, 7th ed. 1999, John Wiley and Sons, New York.

Glantz SA. *Primer of Biostatistics*, 4th ed. 1997, McGraw-Hill, New York.

Moore DS and McCabe GP. *Introduction to the Practice of Statistics*, 2nd ed. 1993, W.H. Freeman and Co., New York.

Fisher L and van Belle G. *Biostatistics: A Methodology for the Health Sciences*, 1993, John Wiley and Sons, New York. and Co., New York.

Rosner B. *Fundamentals of Biostatistics*, 4th ed. 1995, Duxbury Press, Belmont, CA.

Petrie A, Sabin C. *Medical Statistics at a Glance*, 2000, Blackwell Science, Oxford, UK.

Computer Software:

Stata (Student or Intercooled), version 9. Can be purchased at Campus Computer Store (? it appears they will send you to <http://www.stata.com/order/new/edu/gradplans/gp-direct.html>). Please see me if you do not have access to Stata.

Helpful web page: <http://www.stata.com/support/>

Helpful book: Hamilton, L.C. *Statistics with Stata*, Duxbury Press

Netscape v.4.x or higher for downloading course materials.

Homework: There will be 5 homework assignments, generally due one week after they are assigned. You are encouraged to work together on and discuss the homework, if doing so will help you learn the material. However, each person should carry out all analyses and write up the homework his/her self.

I will try to indicate after each lecture which problems can be done at that time. If I forget to do this, please do not hesitate to remind me.

Midterm: Tuesday, 31 July. The midterm exam will be closed-book and will cover material through “Relationships among two categorical variables” (July 25).

Final: Thursday, 23 August, and take-home part/final homework due Monday, 27 August. The in-class part of the final will be closed book and will concentrate on material in the second half of the course, but not including the material on linear and logistic regression.

The final homework will cover analysis techniques from the entire course, including logistic regression. The questions largely resemble a typical homework assignment.

You will be allowed to work together on this part; the same rules as those for the homework will apply.

Grading: Homework: 60%, Midterm: 20%, Final: 20%

Objectives: The student will learn to do the following:

- Identify and give examples of nominal, ordinal, interval-scale and ratio-scale variables
- By hand and using statistical software, present and interpret empirical distribution functions for nominal, ordinal or continuous data.
- Using statistical software, graphically present the joint empirical distribution of two variables, perhaps of different types
- Using statistical software, compute and interpret measures of association of two variables, perhaps of different types
- Using the lexicon of probability, define event, probability, conditional probability, independence, mutual exclusivity, and random variable
- In a biomedical or epidemiologic research study, identify the population and the sample
- Correctly interpret tests of simple hypotheses and confidence intervals
- Carry out simple power analyses and sample size calculations for one- and two-sample study designs
- Perform basic tests on association measures arising from cross-classified nominal variables
- Perform basic analyses on outcomes expressed as proportions
- Estimate and interpret the parameters in a linear regression model using statistical software
- Estimate and interpret the parameters in a logistic regression model using statistical software
- Interpret linear or logistic regression models presented in biomedical or epidemiologic research reports from scientific journals

SYLLABUS AND READING LIST

Date(s)	Topic	Reading Assignments	Due
July 10	Introduction Probability: basic concepts and definitions	Ch. 1: 1.1, 1.2 Ch. 6: 6.1–6.4, 6.6	
July 11	Probability (cont.) Random variables and probability distributions	Ch. 7: 7.1	
July 12	Binomial (Bernoulli) distribution Poisson and Normal (Gaussian) distributions	Ch. 7: 7.2 Ch. 7: 7.3, 7.4	
July 17	Poisson and Normal (Gaussian) distributions (cont.)	Ch. 7: 7.3, 7.4	
July 18	Data types, exploring and summarizing data Descriptive statistics for a single variable	Ch. 2: 2.1 Ch. 2: 2.2, 2.3; Ch. 3: 3.1–3.4	HW 1
July 19	Descriptive statistics for a single variable (cont.) Relationships among two continuous variables	Ch. 17: 17.1, 17.2	
July 24	Relationships among two continuous variables (cont.) Relationships among two categorical variables	Ch. 17: 17.1, 17.2 Ch. 6: 6.5; Ch. 15: 15.3	HW 2
July 25	Relationships among two categorical variables (cont.)	Ch. 6: 6.5; Ch. 15: 15.3	
July 26	Review and “Q & A” Session		HW 3
July 31	MID-TERM EXAM: in class		

(cont.)

Date(s)	Topic	Reading Assignments	Due
August 1	free day	Ch. 8: 8.1–8.4	
August 2	Populations, studies, samples (data) and sampling distributions	Ch. 8: 8.1–8.4	
August 7	One sample hypothesis tests using Z and t distributions	Ch. 10: 10.1–10.4	
	Two sample hypothesis tests using Z and t distributions	Ch. 11: 11.1, 11.2	
August 8	hypothesis tests (cont.)	Ch. 11: 11.1, 11.2	
	Estimation and confidence intervals: Use and interpretation	Ch. 9: 9.1–9.3	
August 9	Nonparametric tests	Ch. 13: 13.1–13.5	HW 4
August 14	Proportions: Estimation and hypothesis tests	Ch. 14: 14.1–14.4, 14.6	
August 15	Chi-square (χ^2) distribution and χ^2 tests	Ch. 15: 15.1	
August 16	Power and sample size calculation	Ch. 10: 10.5–10.7	HW 5
August 21	Simple linear regression	Ch. 18: 18.2, 18.2	
	Multiple linear regression	Ch. 19: 19.1	
August 22	Logistic regression	Ch. 20: 20.1, 20.2, 20.3	
August 23	Review and “Q & A” Session		
August 24	FINAL EXAM: in class part		
August 27			final homework set due