PBHS 32700/STAT 22700 Biostatistical Methods

Spring 2024

Lin Chen

Synopsis. This course is designed to provide students with tools for analyzing *categorical and time-to-event* data frequently encountered in medicine, public health and related biological and social sciences. The course will emphasize methods and application of methods rather than statistical theory, including recognition of the appropriate methods, interpretation and presentation of results. Methods covered include: contingency table analysis, logistic regression, regression models for ordinal data and nominal data, Poisson regression, Kaplan-Meier survival curves, parametric survival models, and the Cox proportional-hazards survival model, and others.

Time and Location. T, Th 12:30 pm - 01:50 pm, BSLC 324.

Instructor. Lin Chen, email: lchen4@bsd.uchicago.edu
Office hours: Thursday 4-5pm (zoom link will be provided on canvas) and by email appointment.

Teaching Assistant. Bowei Kang, email: kbw@uchicago.edu. Office hours: Wednesday 3:30-4:30 pm via zoom.

Prerequisites. A first course in regression analysis. PBHS 32400 (HSTD 32400)/STAT 22400 or STAT 24500 or equivalent or consent of instructor. Facility with the statistical package Stata is assumed, and this package will be used in the course.

Course Website. https://canvas.uchicago.edu/courses/55690
All lecture notes, homework assignments, solutions, and other course documents will be posted here.

Text Books.

- Collett, D. (2003).*Modelling Binary Data, Second Edition.* This book will be the primary resource for the course. It does not cover methods of survival analysis, however, which the author addresses in the companion book, below.
- Collett, D. (2014). *Modelling Survival Data in Medical Research, Third Edition.* This book deals with the basics of survival analysis. We shall cover material drawn primarily from chapters 1-7.

Course Requirements.

- *Homework.* There will be 8 homework assignments. Instructor and TA will use *Canvas/Discussions* to answer homework questions and provide guidance (maybe hints or Stata code).

- *Mid-term exam.* Tuesday, April 16th, 2024, in class. The midterm allows one page of notes. More info will follow when the date is closer.

- *Final exam.* TBD.

Grading and Feedback. The final grade will be based 60% on the problem sets and 40% on the exams. Two late submissions of homework within 48 hours will be allowed. Please contact us if you need an exception for late submissions.
No homework can be dropped. To provide timely feedback, TA/grader may not give detailed individual comments, but a set of solutions will be provided when the papers are returned with comments about major issues and areas of difficulty. Instructor/TA may provide hints or instructions for some problems.

Policies.

1. *Late assignments.* Problem sets are due at 11:59 pm on the assigned date (normally Thursdays and maybe Fridays), and **must be** submitted via Canvas site. No submissions will be accepted after 48 hours beyond deadline, since solutions will be posted. If you anticipate a later than late submission beyond the deadline, please contact the instructor for permission.

2. *Announcements.* We will post new announcements, changes to office hours, corrections and additions to the problem sets, as well as changes to the due dates or times on the course web site. These postings will appear in the "Announcements" section. You are responsible for checking the site regularly and setting your email to receive notifications.

3. *Non-Stata statistical programs.* The official statistical software for this course is Stata, which is available at the Crerar computing cluster and at other locations. You may use other programs at your own risk. In particular, there are things that are easy to do in Stata that are difficult or impossible to do using your favorite package, e.g. R. If you use another program, you are expected to know how that program differs from Stata (if it does), and you are still responsible for a full and complete solution to the problem.

4. *Collaboration.* We encourage you to collaborate on homework. *However, you must write up solutions on your own, neither copying solutions nor providing solutions to be copied.* This also applies to sets of computer commands for Stata and to Stata output. If you use sources beyond the course materials in one of your solutions, e.g., an "expert" consultant, another text, or material other than the text or handouts, be sure to include a proper scholarly citation of the source.
   We **forbid** use of materials from earlier years to which a student may have access. Use of such materials will be taken as a priori evidence of plagiarism subject to serious penalty. Plagiarism, cheating, and similar anti-intellectual behavior are serious violations of academic ethics and will be correspondingly penalized. If you are concerned about a possible violation of this kind, please talk with the instructor. We understand the pressure that students may experience, and we will try to help as best as we can. It is better if you take the initiative to contact us in such cases, rather than vice-versa.
   Collaboration of any sort on the exams is not permitted.

Course Outline and Reading List

Date	Topic(s)	Reading*	Important item@
Mar 19	Lecture 1. Course plan, review and overview		Week 1
Mar 21	Lecture 2. Basic methods for binary outcome data	B Ch. 2	no HW
Mar 26	Lecture 3. Introduction to logistic regression	B Ch. 3.1-3.7	Week 2
Mar 28	Lecture 4. Logistic regression continued	B Ch. 3.1-3.7	HW 1 due
Apr 2	Lecture 5. Goodness of fit	B Ch. 3.8-3.11	Week 3
Apr 4	Lecture 6. Model comparison	B Ch. 3.12-3.16	HW 2 due
Apr 9	Lecture 7. Model diagnostics	B Ch. 5	Week 4
Apr 11	Lecture 8. Overdispersion and Study design	B Ch. 6 and 7	HW 3 due
Apr 16	In-class midterm exam		Week 5
Apr 18	Lecture 9. Probit and Cloglog models	B Ch. 4	no HW
Apr 23	Lecture 10. Ordinal regression	B Ch. 10.1	Week 6
Apr 25	Lecture 11. Multinomial regression	Notes	HW4 due
Apr 30	Lecture 12. Poisson regression	Notes	Week 7
May 2	Lecture 13. Introduction to survival analysis	S Ch. 1	HW 5 due
May 7	Lecture 14. Nonparametric survival methods	S Ch. 2.1-2.4	Week 8
May 9	Lecture 15. Inference for survival data	S Ch. 2.5-2.9	HW 6 due
May 14	Lecture 16. Parametric survival models	S Ch. 5, 6	Week 9
May 16	Lecture 17. Cox proportional hazards model	S Ch. 3, 4	HW 7 due
			Final week
May 21	Final exam time 12:30-2:30pm		HW 8 due

* Readings from Collett binary data (B) or survival (S) books unless otherwise indicated.
@Homework due dates subject to change (deferral to later date).