Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Regression Analysis for Health and Social Research
Winter 2024

Instructor: James J. Dignam, Ph.D., AMB Building W-259, 834-3162, (`jdignam@uchicago.edu`).

Office hours: Monday. 11:00am–12:00pm or by appointment. e-mail is an effective way to reach me with requests for meetings, questions about the material, etc.

Department admin contact: Ms. Emma Collier, ecollier@uchicago.edu.

Course website: https://canvas.uchicago.edu/courses/39918
All homework assignments, class handouts, and other course documents will be posted here

Times and Place: Tuesday, Thursday 11:00am–12:20pm U.S. Central time. BSLC Room 218. Class will meet in person (unless otherwise notified).

Prerequisites: Introductory statistics course (PBHS 32100, STAT 22000, STAT 23400, or similar), ability to use statistical analysis software.

Format: The course will be conducted mostly via lectures. Lecture notes will be posted in Canvas immediately prior to class. During class, participation and input by all is welcome.

Text: (available online)

Recommended: Suárez E, Pérez CM, Rivera R, & Martínez MN *Application of Regression Models in Epidemiology*, 2017, John Wiley and Sons, Hoboken, NJ.

Reference Chatterjee & Hadi. *Regression Analysis by Example*, 5^{th} Edition 2005, Wiley Interscience, New York.

Website (note that this is for Edition 4):
https://onlinelibrary.wiley.com/doi/book/10.1002/0470055464

Other readings: Material of relevance may be identified for additional reading

Teaching Assistants :

Name	**e-mail:**	**Office Hrs**
Bowei Kang	**kbw@uchicago.edu**	**TBA**
Ryan O'Connell	**oconnellr@uchicago.edu**	**TBA**

Computer Software: Stata (Intercooled or SE) version 14 or later. Previous versions should largely be similar. All students can access Stata on USITE computers on campus (for example, at Regenstein level A, Crerar First floor, ect). You may also wish to purchase it (http://www.stata.com)

Other software such as R can be used if desired. A limited amount of R will be demonstrated in the course note examples; however the TAs are well-versed in R and there will be help available for both languages.

Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Regression Analysis for Health and Social Research - Winter 2023

Course Description: In this course, key fundamental elements of linear regression analysis will be addressed. Model estimation, hypothesis testing, model diagnostics, model interpretation, methods to improve models, and a brief introduction to alternative methods that are natural extensions of linear regression will be discussed. The course will present motivation from a statistical theory perspective but will concentrate more heavily on analysis and interpretation.

Course Objectives

Specifically, upon completion of the course the student will:

- Be able to conduct comprehensive linear regression analyses and interpret findings from a statistical science perspective
- Be fable to communicate linear regression analyses effectively to a non-statistician audience
- Understand the limitations of linear regression analyses and able to critique its application when encountered in any data context where it has been applied
- Recognize and be familiar with problems where extensions to standard linear regression are needed, and be prepared to study these methods

Master of Public Health (MPH) Core Competencies Addressed: The Council on Education for Public Health (CEPH) requires that the MPH curriculum include specific core competencies. The core competencies addressed in this course include:

- **MPH Public Health Knowledge (PHK) Competencies**
  - **3.** Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate
  - **4.** Interpret results of data analysis for public health research, policy or practice
- **MPH Concentration (PH-C) Competencies**

Concentration	Competencies
Epidemiology & Global Health (EGH)	3. Develop multivariable statistical models to quantify the relationships between risk factors and disease, potentially in the presence of confounding and effect modification
Health Policy (HP)	3. Develop multivariable statistical models to study the relationship between health care policy, health care services, and health outcomes
Data Science (DS)	3. Develop multivariable statistical models to quantify the relationship between variables and outcomes

Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Regression Analysis for Health and Social Research - Winter 2022

Evaluation: Consists of the following:

Homework (50%): Homework is a critical learning component of this course, and will consist mostly of data analysis and interpretation problems. There will be 6 graded homework assignments. Aside from mathematical notation, homework should be typed. Steps leading to solutions must be shown and computer output must be clearly annotated when included. Copying solutions from others or from past course materials is prohibited. Genuine collaboration and discussion is permitted and in fact encouraged, but everyone must provide their own complete and detailed homework responses. Please also see policy statement on homework on Canvas.

Due date will be specified when the homework is posted. See the website for additional details.

Exams (50%): There will be two in-class exams. More information will be provided early in the course.

Master of Public Health (MPH) Core Competencies: The core competencies addressed in this course are addressed in the evaluations as follows:

Activity	Description	Competencies
Homework 1	simple linear regression (SLR) facts and hypothesis testing, correlation, SLR analysis	PHK-3
Homework 2	more SLR, MLR with continuous variables	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Homework 3	model building, diagnostics	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Exam 1	material to date	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Homework 4	interaction effects, transformations	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Homework 5	Poisson regression, logistic regression	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Homework 6	logistic regression	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3
Exam 2	all material (mainly since Exam 1)	PHK-3, PHK-4 EGH PH-C-3 HP PH-C-3 DS PH-C-3

Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Course Outline and Reading List

Date	Topic(s)	Reading*	Important item@
Jan 4	Course plan, discuss course objectives, review of basic statistics background Perspectives, Overview	Ch. 1	
Jan 9	Introduction to simple linear regression, correlation	Ch. 1, 2, 6.1-6.3	
Jan 11	Simple linear regression (continued) Least squares estimation, testing	Ch. 2	
Jan 16	Matrix representation, Multiple linear regression (MLR)	Ch. 3, 4	HW 1 due
Jan 18	Multiple linear regression (cont.) Indicator variables	Ch. 4	
Jan 23	Interaction effects	Ch. 4	HW 2 due
Jan 25	Testing in MLR	Ch. 5	
Jan 30	Main Effects, Interactions, Testing review Variable selection		
Feb 1	Variable selection, model building Brief review	Ch. 5	HW 3 due
Feb 6	**Exam 1**		

* Reading from main text (SPRM) for this meeting. @Homework will be distributed regularly throughout course, at least one week before due date.

Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Course Outline and Reading List (cont.)

Date	Topic(s)	Reading*	Important item@
Feb 8	Residuals, regression diagnostics Transformations	Ch. 7	
Feb 13	Adjusted estimates, confounding	Notes	
Feb 15	TBA		HW 4 due
Feb 20	Generalized Linear Models Poisson regression	Ch. 10	
Feb 22	Poisson regression (cont.) Binary endpoints - logistic regression	Ch. 10, 11	HW 5 due
Feb 27	Logistic regression Estimation and interpretation	Ch. 11, 12	
Feb 29	Logistic regression (cont) Model fit, prediction Review	Ch. 11, 12	HW 6 due
Mar 6 week	**Final Exam - date TBD**		

* Reading from main text (SPRM) for this meeting. @Homework will be distributed regularly throughout course, at least one week before due date.

Public Health Sciences 32410 (PBHS 32410/STAT 22401)
Master of Public Health (MPH) Core Competencies

Date	**Topic(s)**	Public Health Knowledge **(PHK)**	Concentration **(PH-C)**
Jan 4	Course plan, discuss course objectives Perspectives, Overview	PHK-4	
Jan 9	Introduction to simple linear regression	PHK-3, PHK-4	
Jan 11	Simple linear regression (continued) Least squares estimation, testing	PHK-3, PHK-4	
Jan 16	Matrices,Multiple linear regression (MLR)	PHK-3, PHK-4	EGH PH-C-3, HP PH-C-3, DS PH-C-3
Jan 18	MLR (cont.), Indicator variables	PHK-3, PHK-4	EGH PH-C-3, HP PH-C-3, DS PH-C-3
Jan 23	Interaction effects	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Jan 25	Testing in MLR	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Jan 30	Variable selection, Model building	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 1	Residuals, regression diagnostics, review	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 6	**Exam 1**	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 8	Diagnostics (cont.), Transformations	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 13	Adjusted estimates, confounding	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS-PH-C-3
Feb 15	Generalized Linear Models	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 20	Poisson regression	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 22	Binary endpoints - logistic regression	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 27	Logistic regression Estimation and interpretation	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Feb 29	Logistic regression (cont) Model fit, prediction	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3
Mar 5 week	**Final Exam - date TBD**	PHK-3, PHK-4	EGH PH-C-3 HP PH-C-3, DS PH-C-3