

CLRES 2021 Syllabus

Course Description and Objectives

This course is designed for medical researchers who are not biostatistics majors. Topics covered include multiple linear regression, regression diagnostics, analysis of variance, analysis of covariance, confounding, mediation, moderation, and model selection. At the completion of the course, trainees should be able to understand the appropriate uses of ANOVA and linear regressions, to assess their appropriateness and adequacy, to analyze simple data sets taken from the fields of medicine and public health, and to summarize results from regression models via written communication.

The course objectives are to:

- Demonstrate an understanding of the basic principles of ANOVA and linear regressions
- Develop the skills to appropriately use ANOVA and linear regressions
- Be able to perform and interpret regression analysis using Stata
- Be better able to read and understand professional articles in clinical research that have used linear regression techniques
- To recognize situations that require more advanced knowledge and methods

Instructors

Charity G. Moore, PhD, MSPH

Associate Professor of Medicine
Center for Research on Health Care Data Center
University of Pittsburgh
200 Meyran Ave., Suite 300
Pittsburgh, PA 15213
Tel: 412-246-6961
Fax: 412-586-9672
Email: moorecg@upmc.edu

Teaching Fellow

Kaleab Abebe, MA

Department of Statistics
University of Pittsburgh
Email: kza3@pitt.edu
Office hours: Tuesday and Thursday, hours TBA

Meeting times and location

September 3 – September 29, 2008 MW 1:00pm – 3pm
200 Meyran Ave., Suite 222 (Parkvale Building Computing Lab)

Credits

1 credit (16 contact hours; 4 contact hours/week for 4 weeks)

Grading

Letter grade based on

60%	Homework assignments
20%	Take-home final project
10%	Final exam
5%	Class participation
5%	Attendance

TEXTBOOK

Required

Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied Regression Analysis and Multivariable Methods*, 4th Edition. Pacific Grove: Duxbury Press (2007).

Additional text (optional)

Vittinghoff E, Glidden DV, Shiboski SC, McCulloh CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer (2005).

Both books are available at the CRTP library in Parkvale 302.

COMPUTING PACKAGE

Stata 10, Stata Press, College Station, Texas (www.stata.com)

PROBLEM SETS

The homework assignments are essential in learning the concepts of linear regression and ANOVA. Trainees are encouraged to discuss and work together about difficult problems, but all trainees must write their own solutions.

Homework 1 (assigned 9/3/08): due 9/10/08
Homework 2 (assigned 9/10/08): due 9/17/08
Homework 3 (assigned 9/17/08): due 9/24/08
Homework 4 (assigned 9/24/08): due 10/1/08

LATE HOMEWORK ASSIGNMENTS: Students will be assigned written exercises that will be graded. All homework assignments will be assigned with a due date. Late homework assignments will be penalized 10% per day over the due date unless prior arrangements have been made with the instructors.

Final Exam: There will be a final exam on the last day of class consisting of short answer and multiple choice questions.

Final Project (assigned September 29, 2008; due October 13, 2008): There will be a final project distributed on the last day of class. Students will be asked to analyze a data set and answer specific questions pertaining to this analysis and its results. Trainees must work on the final project independently. Discussion, consulting, or working with other people are not permitted.

Academic Integrity: Students in this course will be expected to comply with the [University of Pittsburgh's Policy on Academic Integrity \(http://www.provost.pitt.edu/info/ai1.html\)](http://www.provost.pitt.edu/info/ai1.html). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

Disabilities: If you have a disability that requires special testing accommodations or other classroom modifications, you need to notify both the instructor and the [Disability Resources and Services \(http://www.dr.s.pitt.edu/policies.html\)](http://www.dr.s.pitt.edu/policies.html) no later than the 2nd week of the term. You may be asked to provide documentation of your disability to determine the appropriateness of accommodations. To notify Disability Resources and Services, call 648-7890 (Voice or TTD) to schedule an appointment. The Office is located in 216 William Pitt Union.

Incomplete grades: Students who are unable to complete the course for any reason must contact the course instructor as soon as possible to discuss grades and remediation [course reasons ("I" incomplete), extenuating personal reasons ("G" incomplete), withdrawal ("W")]. Students will have one calendar year from the start of the course to complete the course requirements, otherwise an "I" or "G" grade will remain on the transcript.

Sessions

Wednesday, September 3, 2008

Lecture 1: Correlation and Simple Linear Regression

- Course description
- Levels of measurement and types of data
- Review of correlation
- Regression methods (preface to other courses)
- review of simple linear regression (form, assumptions, estimation, hypothesis testing, R^2)

Reading: Kleinbaum Ch 4 -7 (Intro Regression, SLR, Correlation, ANOVA table)

Supplemental reading: Vittinghoff Ch 2 (Explor & Descr) and Ch 3 pgs 36-44 (SLR)

Homework 1: Correlation, simple linear regression, and multiple linear regression

Data set: Klein3ed52.dta

Monday, September 8, 2008

Lecture 2: Multiple Regression

- Multiple (multivariable) regression model
- Estimation
- Evaluating model and predictor significance
- Correlations (multiple, partial, and multiple partial)
- Introduction to dummy coding (indicator variables) in linear regression

Reading: Kleinbaum Ch 8-10 & Sections 12.1-12.3 (MLR, Correlations, Intro to dummy vars)

Supplemental reading: Vittinghoff Chapter 4 69-82 (Linear regression and categorical predictors)

Homework: (see homework 1)

Wednesday, September 10, 2008

Lecture 3 Confounding, Mediation and Interaction in Regression

- Covariates
- Confounding
- Interaction and testing for equal slopes
- Mediation vs moderation

Reading: Kleinbaum Chapters 11 and 12 (Confounding and Interaction)

Supplemental reading: Vittinghoff Chapter 4 83-108 (Confounding, Mediation, and Interaction)

Homework 2: Confounding, interaction, and regression diagnostics

Data set: protein.dta, chol.dta

Monday, September 15, 2008

Lecture 4 Regression diagnostics I

- Simple approaches to diagnosing problems
- Residual analysis and types of residuals
- Detecting outliers (leverages, jackknife residuals, Cook's d)
- Assessing model assumptions
- Alternate strategies (if LINE assumptions are not met)
- Summary of how to validate assumptions

Reading: Kleinbaum Ch 14

Supplemental reading: Vittinghoff Chapter 4 109-127 (Model Assumptions & Fit) and Ch 5 147-149 (Collinearity & number of predictors)

Homework: see homework 2

Wednesday, September 17, 2008

Lecture 5: Regression diagnostics II and Model Selection

- Collinearity
- Model selection
- Best prediction of Y
- Steps in selecting the best regression equation
- Criteria for selecting a model
- Backward, forward, and stepwise procedures

- Evaluating a primary predictor and multiple important predictors
- Introduction to hierarchical regression

Reading: Kleinbaum Ch 14 and 16 (Diagnostics and Selecting the Best Regression Equation)

Supplemental reading: Vittinghoff Chapter 5 (Predictor Selection) and Ch 8 259-261 (Hierarchical Data)

Homework 3: Regression diagnostics, model selection, and ANOVA

Data set: chol.dta, fishermen.dta, wcfgs.dta

Monday, September 22, 2008

Lecture 6: ANOVA and Analysis of Covariance

- Review of ANOVA
- ANOVA vs Dummy Regression
- Reference cell coding
- Multiple comparisons
- ANCOVA and linear regression
- Adjusted means example

Reading: Kleinbaum Ch 13 and Ch 17 (One-way ANOVA and ANCOVA)

Supplemental reading: Vittinghoff Chapter 3 pgs 29-34 (t-test and ANOVA)

Homework: See homework 3

Wednesday, September 24, 2008

Lecture 7: Analysis of Correlated Data and Review of Topics

- Examples of correlated data
- Mixed models (fixed and random effects)
- Correlation structures
- Missing data in longitudinal analysis
- Review of class topics

Suggested reading: Kleinbaum Ch 25-26 (Analysis of Correlated Data) and Sections 27.1-27.3

Note about reading: some of the reading in Ch's 25-26 is statistically dense; the examples will be most helpful but feel free to use the lecture as your primary source for information

Supplemental reading: Vittinghoff Chapter 8 (Repeated Measures Analysis)

Homework 4: Correlated data

Data set: TBD

Monday, September 29, 2008

Sample size planning, review of project, In-class exam and case studies from literature