Tufts UNIVERSITY | Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

NUTR 307: Regression Analysis for Nutrition Policy
Spring 2023

Welcome to NUTR 307!

Regression analysis opens a world of insight into practical issues in nutrition science and policy. Quantitative analysis is not merely an academic curiosity but can help society achieve substantial public interest goals. These goals may include advancing food security, dietary quality, thriving local economies, environmental sustainability, scientific understanding, and social justice. A key contribution of statistics broadly is to facilitate sound generalizations, using samples to understand the wider world around us. A contribution of regression analysis is to help in holding "all else equal," permitting more rigorous interpretations of cause and effect in some circumstances.

You are welcome and wanted here. With quantitative courses, students sometimes may feel that instructors or colleagues want to teach just those with the strongest prior preparation or the greatest affinity for mathematics. Not so in this course. As noted in the prerequisites section below, you will find that, with some reasonable effort, the math content is within your grasp. The instructors and TAs enjoy the challenge of making this quantitative course relevant for students who care about real-world issues.

One implication of this approach is that we welcome your full self, meaning your own interpretation of your identity, including your race, ethnicity, class, gender, sexual orientation, country of origin, ability or disability, and religious or political affiliation. You don't need to leave the real you at the classroom door. Another implication is that the importance of your own physical and psychological wellness is recognized here. With any concern or stress, you can talk to us. In many cases, we may recognize your struggle and have resources or remedies ready to share. In other cases, we may learn something new from your question and go looking for additional support. Our goal is to make our topic fit your personal program, not to make you fit into our mold.

Important Information:

Class Meetings: *Mondays/Wednesdays, 1:30-3pm. Behrakis auditorium except when noted. Zoom Links are posted to Canvas. Email a TA if not in-person.*

Instructor(s): *Parke Wilde (he/him)*

339-368-2975

parke.wilde@tufts.edu (include "NUTR 307" in email subject line).

Teaching Asst.: *Yu-Hsiang Chiu (he/him)*

857-437-9697

Yu-Hsiang.Chiu@tufts.edu (include "NUTR 307" in email subject line).

Ananyaa Mohan(She/Her)

781-957-8419

Ananyaa.Mohan@tufts.edu (include "NUTR 307" in email subject line).

Semester Hour Units: *3.*

Prerequisites: Biostatistics. NUTR 207 or NUTR 209 or permission of the instructor. For a review of material covered in NUTR 207, see the text used by Prof. Sean Cash, Statistical Methods for the Social Sciences, 5th edition, A. Agresti, 2017.

Students should be familiar with mathematics at the pre-calculus level. As suggested for NUTR 207, see the Columbia University primer or the Khan Academy video series recommended by the Educational Testing Service. Calculus and matrix algebra are not required.

Course Communications:

There are many good ways to communicate with us. We use the "announcement" function on Canvas occasionally for major news (your email should receive announcement notifications). We use the "discussion" function more frequently (your email should notify you of new discussion threads, unless you choose a setting to stop receiving email). As much as possible, share questions and comments using the discussion page, so students can benefit from each other's questions and comments. In discussion board subject lines, we use the prefix "content:" for information or questions about course content (how to compute a standard error) and the prefix "mechanics:" for course mechanics (such as an adjustment to the due date for a homework assignment). For information or questions not suitable for sharing on the discussion board, email us directly or visit our office hours. We are delighted to hear from you.

Office Hours:

Instructor: *Wed 3-4pm (open office hours)*
Tues 4-5pm (NUTR 307 dedicated appointments link: https://calendly.com/parke-wilde)
And by appointment at other times (https://calendly.com/parke-wilde)

Teaching Asst.: ***Yu-Hsiang Chiu****: Thurs 9-10:30am location Jaharis Room 254 (and Zoom on Canvas) or by appt.*
Zoom link - https://tufts.zoom.us/j/96993943198?pwd=N2d5aHJiOFVhdlJWZFlJbHdhdWd0UT09
Ananyaa Mohan: *Thurs 7-8:30pm (Zoom) or by appt.*
Zoom link- https://tufts.zoom.us/j/97002641570?pwd=TW5YRG52aFVVNWFvME9MT3FhSEdrdz09

Do you have a question about the homework, a thought about concept in class, or an idea about a career path you want to explore? All of these topics and more can be discussed in office hours. Our doors are open – just stop by or make a Calendly appointment.

Course Summary:

Part two of a one-year, two-semester course sequence in statistics. This course is intended for students whose main focus is non-experimental or survey-based research. The course covers research design, simple linear regression, multiple regression, analysis of variance, non-linear functional forms, heteroskedasticity, complex survey designs, and real-world statistical applications in nutrition science and policy. Students will make extensive use of Stata.

Course Goals:

Goal:
Master the basic statistical methods most useful for ***non-experimental*** *and* ***survey-based*** *research in nutrition science and policy.*

Topics:
Review

- ***Review.*** *Review the basic building blocks of statistics: discrete and continuous variables, expectation, conditional expectation, variances, covariances, and the normal distribution.*
- ***Simple linear regression****. Review the theory and practice of simple linear regression, with one dependent variable and one independent variable.*

Models

- ***Multiple linear regression****. Learn the practice of multiple linear regression. Intuition and understanding will be developed without matrix algebra, mainly by analogy with simple linear regression.*

- ***Analysis of Variance (ANOVA).*** *Study problems where the dependent variable is continuous while key explanatory variables are binary or categorical. For example, a binary variable may identify whether a subject is in a study group or control group.*
- ***Brief introduction to Logistic Regression (Logit) and Linear Probability Models.*** *Learn statistical analyses that are appropriate when the dependent variable is binary.*
- ***Panel data.*** *Introduce statistical methods for longitudinal or panel data.*
- ***Introduction to complex surveys.*** *Correctly interpret surveys that have complex sampling designs, using sampling weights and design-corrected standard errors.*

Solving problems

- ***Confounding variables and problems with functional form****. Learn solutions to the most serious specification problems, which cause bias in all parameter estimates.*
- ***Heteroskedasticity.*** *Recognize that heteroskedasticity biases standard errors but does not bias regression parameter estimates. Learn to use robust standard errors in large samples.*
- ***Non-normality****. Recognize that non-normality causes problems for hypothesis tests of regression parameter estimates in small samples, but not in large samples.*
- ***Instrumental variables and program evaluation.***

Texts or Materials:

Primary text. *Lecture notes are provided on Canvas.*

Secondary text*. In addition to instructor lecture notes and videos on Canvas, students may refer to Introductory Econometrics: A Modern Approach (fifth edition through seventh edition), by Jeffrey M Wooldridge (Mason, OH: Thomson South-Western). For electronic rental options, see "Mindtap Course List on Amazon" and also RedShelf, which had a temporary free or low-cost option as of Spring 2020 (https://www.about.redshelf.com/redshelfresponds). Purchase is optional. Many students complete this course with the instructor lecture notes and videos alone.*

Applied readings*. Several empirical research articles and reports will illustrate how the methods learned in this class are applied to real-world problems in (1) agriculture, food, and the environment, and (2) food and nutrition policy and programs.*

How to be Successful in this Course:

This syllabus section, new this year, gives an opportunity to make explicit the "hidden curriculum." Here below are my tips now at the start of the semester. Let us have a conversation about this "how to" section at the end of the semester. I will revise it with your input for the benefit of future students.

1. *Attend class. Turn off electronic distractions.*
2. *Study the lecture notes. Read them once briefly before a new topic lecture (usually Monday), but on this first reading do not feel nervous if some parts seem confusing. That is normal. Read the lecture notes again after the second topic lecture (usually Wednesday), and this time make a plan to resolve any parts that still seem confusing.*
3. *Usually begin the homework after the Wednesday class, so that you have enough time to meet with student colleagues or attend TA office hours on Thursday or Friday to resolve any difficulties. If you are nervous about grades, getting high scores on the homework can help balance out quizzes. The homework tasks have easier timing and permit friendly collaboration.*
4. *Exercises in statistics are like practice in learning a musical instrument. Some people enjoy them, and some do not, but they are utterly unavoidable. Sometimes, watching a colleague solve a puzzle generates a good "aha!" feeling in your own brain, but this feeling is misleading. Watching is not the same as doing. If you have help on a particular homework problem, you should rehearse variations of the same type of problem before quiz time.*
5. *For additional practice, we have built up a growing repertoire of helpful problems. Still, I want to teach you also to construct variations on existing problems as part of your rehearsal and study process. For example, if a homework*

question asks you to interpret one regression coefficient in a table, you can try your hand at another coefficient from the same table and then check your answer with the instructor or TA.

6. *Consider front-loading your effort in this course toward the first half of the semester. The grade weight on the final in May is just 10 percentage points bigger than the quizzes throughout the course. This avoids excessive stress during finals week, but it also makes early effort important. Mastery of early material will help you with the content of later sections.*
7. *Before quizzes, study on several days, not just the night before. Reserve study time on your calendar in advance. The quizzes are now all open-book on Canvas, but it is good to prepare in advance as you would have done in the past for a closed-book test.*
8. *Split your study effort between new topics and old topics. Sometimes, students may feel pressure to keep up with the new topics and feel bad if they are still solidifying old topics, but be reassured: it is natural for your mastery of the basics to get stronger and stronger as the course proceeds.*
9. *Students with all different levels of prior preparation have done well in this course. If you are worried, it is important to visit TA and instructor office hours frequently starting early in the semester, when our support and assistance and your own good work have plenty of time to pay off. We value your communication!*

Assignments and Grading:

Homework and Labs (30%)
Quizzes (45% for 3 quizzes)
Final examination (25%)

Grading Range:

*A **passing grade in the course is B- or better**. Course grades will be based on the below (subject to revision during the course):*

A	*94%+*
A-	*90 - <94%*
B+	*87 - <90%*
B	*84 - <87%*
B-	*80 - <84%*
C+	*77 - <80%*
C	*74 - <77%*

Instructions for Submission of Assignments and Exams:

Homework and labs:

Weekly problem sets are submitted through the course site on the dates indicated below, usually due Sunday evening (any time before that date is fine). There is a 10% score penalty for late submission.

Quizzes and examination:

There will be three 60-minute quizzes and a 90-minute final examination, open book, on Canvas, but with no communication with people, on the dates indicated below. Please plan ahead and avoid schedule conflicts on these dates.

Academic Conduct:

You are responsible for upholding the highest standards of academic integrity, as specified in the Friedman School's Policies and Procedures Handbook located at this web page: https://nutrition.tufts.edu/about/policies-and-procedures, as well as Tufts University's policies (https://students.tufts.edu/community-standards/support-resources/academic-integrity-resources). This includes understanding and avoiding plagiarism, which is defined as the unacknowledged use of someone else's published or unpublished work. It is the responsibility of each student to understand and comply with academic integrity standards, as violations will be sanctioned by penalties ranging from failure on an assignment and the course to dismissal from the school.

Accommodation of Disabilities:

We will do our best to ensure each of you has the resources you need to succeed. Tufts University is committed to providing equal access and support to all students through the provision of reasonable accommodations so that each student may access their curricula and achieve their personal and academic potential. If you have a disability that requires reasonable accommodations, please contact the Friedman School Assistant Dean of Student Affairs at 617-636-6719 to make arrangements for determination of appropriate accommodations. Please be aware that accommodations cannot be enacted retroactively, making timeliness a critical aspect for their provision.

Tufts Zoom:

The Friedman School's on-campus courses may be offered by Tufts Zoom (https://access.tufts.edu/zoom) on days when the Boston campus is closed due to pandemic, weather or a temporary cancellation issue. Students should expect to be notified through a Canvas announcement in the event that class is cancelled and will be provided with the Zoom link for students to attend any remote class sessions during the normally scheduled class period. The Zoom meeting video and audio will be recorded and posted on the course's Canvas site (https://login.canvas.tufts.edu/) when completed. If an on-campus Examination, Presentation, etc. was scheduled on a day when the Boston campus is closed due to weather or a temporary cancellation issue and cannot be conducted by Zoom, the exam/presentation will be rescheduled for an alternate on-campus class session date.

On-Campus and Remote Participation:

- This course will be delivered in the classroom. Only students who have received approval from the Friedman School's Assistant Dean for Student Affairs will be permitted to participate remotely by Zoom on a regular basis.
- If you are ill, please do **NOT** come to campus. Contact the instructor or TA to let them know you would like to participate by Zoom.
- If you will need to participate remotely for a particular class session, please contact the instructor or TA in advance. You don't need to wait for their approval; after sending notice, you can go ahead and use the Zoom link.
- In the event of inclement weather leading to campus closure the instructor may choose to conduct the class remotely by Zoom.
- Some class sessions may be recorded. All students in the course will have access to these recordings. Massachusetts law states that students have the right to not have their voices recorded in the classroom; if this is your wish you may refrain from participating verbally in class when the session is being recorded.

Course Overview:

Course Schedule (subject to future modification)**:**

Topic #	*Topic*	*Secondary Readings*	*HW/Quiz*
1 *Jan-18, 23, 25*	***Multiple regression: estimation***	*Wooldridge, Ch 3*	*HW1 (Jan 29)*
2 *Jan-30, Feb-1*	***Multiple regression: inference***	*Wooldridge, Ch 4*	*Lab 1 3:15pm (Jan 30, Feb 1, due Feb 5)*
3 *Feb-6, 8*	***Large samples, functional form, goodness-of-fit***	*Wooldridge, Ch 5, 6*	*HW2 (Feb 12)*

4 *Feb-13, 15*	***Putting it together: reading regression analysis (I)***	*Article on Canvas*	*Quiz 1 (Feb 15)*
5 *Feb-22, 23 (Thurs!)*	***Analysis of variance (categorical vars)***	*Wooldridge, Ch 7*	*HW3 (Feb 27)*
6 *Feb-27, Mar-1*	***Heteroskedasticity***	*Wooldridge, Ch 8*	*Lab 2 3:15pm (Feb 27, Mar 1, due Mar 5)*
7 *Mar-6, 8*	***Specification problems***	*Wooldridge, Ch 9*	*HW4 (Mar 12)*
8 *Mar-13, 15*	***Putting it together (II)***	*Article on Canvas*	*Quiz 2 (Mar 15)*
Recess	*-*	*-*	*-*
9 *Mar-27, 29*	***Longitudinal or panel data***	*Wooldridge, Ch 13*	*HW5 (Apr 2)*
10 *Apr-3, 5*	***Complex surveys***	*Documentation on Canvas*	*HW6 (Apr 9)*
11 *Apr-10, 12*	***Putting it together (III)***	*Article on Canvas*	*Quiz 3 (Apr 12)*
12 *Apr-19, 21 (Fri!)*	***Logit***	*Wooldridge, Ch 17*	*HW7 (Apr 24)*
13 *Apr-24, 26*	***Instrumental variables***		
May-1	***Review and conclusion***		

Final Examination. *May 8 (Mon), 1:30-3pm, online. Please plan ahead for this date.*

Topics, Assignments, and Learning Objectives for Each Class Session:

Preparation for NUTR 307: *Principles of Statistics*

Objectives:

- Charts
  - Be able to interpret charts showing values of a single variable. Recognize which axis shows the variable of interest (such as the vertical axis in a vertical bar chart).
  - Be able to interpret histogram charts for a single variable. Recognize that the horizontal axis typically shows the variable of interest, and the vertical axis shows the frequency with which values of this variable occurs.
  - Be able to interpret charts showing values for two variables (such as scatter plots).
- Mathematics review
  - Know how to convert a proportion into a percentage, and vice versa.
  - For any continuous random variable, given a value at time 0 and time 1, be able to explain the percentage change.
  - Given a percentage at time 0 and time 1, understand the distinction between a percentage change and a percentage-point change.
  - Know the meaning of a logarithm (base 10) and a natural logarithm (with a base of *e*).
- Random variables
  - For a continuous random variable, understand the meaning of "normally distributed."

- Samples and populations
    - Know and understand the equation for a population mean (which also happens to be the equation for a sample mean, useful for estimating the population mean). Given sample data, be able to estimate a population mean.
    - Know the equation for a population variance. Know the equation for a sample variance, when it is used to estimate a population variance. Be able to compute a sample variance. Understand that a variance is an average of squared deviations from means. Given some sample data, be able to estimate a population variance.
    - Know the equation for a sample standard deviation, when used to estimate a population standard deviation. The standard deviation is simply the square root of a variance. Given some sample data, be able to estimate a population standard deviation.
    - Know the equation for a sample covariance, when used to estimate a population covariance. Understand that the covariance of x and y is an average of the products of the deviations from means for x and y. Given some sample data on x and y, be able to estimate a population covariance.
    - Know the equation for a correlation. Given some sample data on x and y, be able to estimate a population correlation.
    - Know the meaning of a standard error of a mean. Be able to explain the difference between a standard error of a mean and a standard deviation of a random variable.
- Hypothesis testing
    - For a simple test of the difference in two means, know how to state a null hypothesis. Know how to state a one-tailed or two-tailed alternative hypothesis.
    - Know the equation for a t-statistic. Given a t-statistic for a hypothesis test, decide whether to reject the null hypothesis.
    - Know the meaning of a p-value. Given a p-value for a hypothesis test, decide whether to reject the null hypothesis.
    - Given a statistic and a confidence interval, conduct a hypothesis test.
- Simple linear regression
    - Be able to interpret the slope coefficient of a simple linear regression model (using the correct units in an interpretation sentence).

Secondary Readings: *Statistics Review Materials on Course Site*
Activities for this week: Self-assessment using "Preparation for NUTR 307."

Topic 1: *Multiple regression: estimation*

Objectives:

- Recognize the wide diversity of useful applications for regression tools in non-experimental studies in food policy, nutrition policy, applied nutrition, agriculture, food, and environmental studies.
- Given a regression coefficient, be able to provide a correct interpretation sentence, including the correct units.
- Unbiasedness is the "number one" good quality we wish to find in a regression model. Know the meaning of unbiasedness and the assumptions required for unbiased regression estimates.
- Use the "3-variable game tool" to explore the consequences of failed assumptions, focusing specifically on circumstances that cause bias in a coefficient for a simple regression model.
- Understand how R-square is computed and what it means.
- Be able to interpret regression results as they appear in either Stata output or published tables from journal articles.

Secondary Readings: *Wooldridge, chapter 3*

Topic 2: *Multiple regression: inference*

Objectives:

- Precision is the "number two" good quality we wish to find in a regression model. Understand how to assess the precision of a coefficient estimate.
- Understand each part of the equation for the variance of a regression coefficient estimate var($\hat{\beta}_1$) -- though you need not memorize this equation. Know the factors that make regression coefficients more or less precise.
- Know the meaning of the standard error se($\hat{\beta}_1$), and recognize that it can be computed as the square root of var($\hat{\beta}_1$). Know what var($\hat{\beta}_1$) and se ($\hat{\beta}_1$) tell us about the precision of estimates.
- Be able to carry out a hypothesis test using five steps: (1) state assumptions, (2) state null and alternative hypotheses, (3) compute a test statistic whose distribution is known under the assumption that the null hypothesis is true, and (4) compute the corresponding p-value, (5) state a conclusion regarding the truth of the null hypothesis in the population.
- Find the information needed for hypothesis tests from published results tables or default Stata output when possible (t statistics and F statistics).
- Be able to request additional hypothesis tests from Stata when necessary (F statistics).
- Be able to compute confidence intervals from regression results, and know the interpretation of confidence intervals.
- Know the 5 assumptions needed for computing the "usual" regression standard errors, and the 6 assumptions needed for conducting the "usual" hypothesis tests in small samples.

Secondary Readings: *Wooldridge, chapter 4*

Topic 3: *Large samples, functional form, goodness-of-fit*

Objectives:

- Know the central limit theorem and be able to explain how the central limit theorem makes it easier to satisfy the assumptions of the regression model in large samples.
- Know the meaning of consistency and the assumptions required for consistent regression estimates.
- Be able to identify which difficulties in regression analysis are or are not remedied by large sample sizes.
- Recognize that the actual conduct of hypothesis tests is similar in small and large samples.
- Be able to explain quantitatively how regression coefficients change if the scale of a dependent or explanatory variable changes.
- Interpret regression coefficients when the dependent or explanatory variable is in logarithmic form.
- Use quadratic terms to estimate non-linear relationships between an explanatory variable and an outcome variable.
- Use interaction terms to estimate the effects of two variables that have synergistic effects.
- Use regression results to predict the value of an outcome variable, when given particular values of the explanatory variables.
- Be able to describe the precision of such predictions for (a) a particular observation or (b) the estimated expected value of the outcome variable conditional on the explanatory variables.

Secondary Readings: *Wooldridge, chapters 5 & 6*

Topic 4: *Putting it together: reading regression analysis (I)*

Objectives:

- In applied examples of published research, recognize whether the organization of quantitative material is typical or atypical.

- For descriptive statistics and regression results, be able to interpret tabular estimates (including locating additional information if needed for correct interpretations).
- Recognize common variations in the display of quantitative information.
- Read article conclusions and judge for yourself whether they are justified on the basis of statistical evidence presented.

Secondary Readings: *Article on course website.*

Topic 5: *Analysis of variance (categorical variables)*

Objectives:

- Understand the relationships between two categorical variables, including the meaning of joint distributions and conditional distributions.
- Interpret regression coefficients when an explanatory variable is binary or categorical.
- Using an example of one continuous explanatory variable and one binary explanatory variable, explore once again the circumstances in which regression coefficients may be biased.
- Estimate regression models with categorical explanatory variables using both ANOVA tools and regression tools in Stata, and confirm that the two approaches are equivalent.
- Use a linear probability model to estimate regression models when the outcome is a binary variable.

Secondary Readings: *Wooldridge, chapter 7*

Topic 6: *Heteroskedasticity*

Objectives:

- Know the consequences of heteroskedasticity, without either overstating or understating these consequences.
- Test for the presence of heteroskedasticity in empirical examples.
- When sample size is large, use robust standard errors to compensate for the problem of heteroskedasticity.
- Understand the motivation for weighted least squares models in the presence of heteroskedasticity (but it is beyond the scope of this course to know weighted least squares models in detail).

Secondary Readings: *Wooldridge, chapter 8*

Topic 7: *Putting it together: reading regression analysis (II)*

Secondary Readings: *Article on course website.*

Spring Recess

Topic 8: *Specification problems, functional form, and regression discontinuity*

Objectives:

- Test for particular non-linearities or synergies by adding additional quadratic terms or interaction terms.
- More generally test for incorrect functional form using Ramsey's Regression Specification Error Test (RESET).

- Explain the consequences of measurement error in dependent and explanatory variables, recognizing how these consequences depend on assumptions about the nature of the measurement error.
- Understand the effects of non-response and identify situations in which non-response does and does not lead to biased coefficient estimates.
- Quantify non-response problems using response rates.

Secondary Readings: *Wooldridge, chapter 9*

Topic 9: *Longitudinal or panel data*

Objectives:

- Understand how longitudinal or panel data are arranged and indexed with subscripts i and t.
- Know the consequences in Ordinary Least Squares (OLS) of ignoring unobserved individual-level time-constant confounding variables.
- Use fixed effects models to remedy the problem of unobserved individual-level time-constant confounding variables.
- Analyze panel data using the panel data tools in Stata.
- Understand the purpose of random effects estimators for panel data, including how this purpose differs from that of the fixed effects estimator.

Secondary Readings: *Wooldridge, chapter 13*

Topic 10: *Complex surveys*

Objectives:

- Recognize the structural similarity between complex survey data discussed this week and the panel data discussed in the previous week.
- Understand both stratification and clustering in complex surveys.
- Be able to define design effects, effective sample size, and intra-cluster correlation coefficients.
- Know the consequences of estimating OLS with the usual standard errors when the data come from complex surveys.
- Use the survey procedures in Stata to compute correct standard errors for descriptive statistics and regression estimates.
- Understand how complex surveys design choices are related to the economics of cost-per-cluster and cost-per-observation.

Secondary Readings: *Introduction, Stata Survey Reference Manual (Release 15)*.

Topic 11: *Putting it together: reading regression analysis (III)*

Secondary Readings: *Article on course website.*

Topic 12: *Logit, probit, and a second look at the linear probability model*

Objectives:

- Know the meaning of probability, odds, and log odds, and be able to translate from one to another in quantitative examples.
- Interpret odds ratios as they appear in logit model tabular output.
- Understand the equation for a logit model.
- For logit model coefficients, give interpretation sentences in terms of both log odds and odds ratios.

- For probit coefficients, give interpretation sentences in terms of the z-score for the probability that an outcome happens.
- Compare results from logit models and linear probability models.

Secondary Readings: *Wooldridge, chapter 17*

Assessment of Competencies for the Master of Public Health (MPH) Degree

For students in the Epi/ Biostatistics concentration in the Tufts MPH program, this is a required course. This course delivers concentration competencies required for the MPH degree. To pass this course, you must successfully complete each competency-based assignment. If you do not get a passing grade on the competency-based assignment, please arrange to meet with me to discuss next steps so we can ensure you attain the competency.

Competency	Sessions Where Competency is Taught	Competency Assessment
Demonstrate the reproducibility of work when engaged in any phase of the life cycle of data from conceptualization to dissemination.	Lab 1 and Lab 2 use Stata .do files (rather than merely pull-down menus) for reproducibility. Three class sessions emphasize how journal article methods sections facilitate reproducibility.	Lab 1 and Lab 2 are accompanied by homework submissions. The understanding of journal articles is assessed in Quizzes 1, 2, 3, and the final examination.
Apply appropriate epidemiologic design and statistical analysis techniques to minimize bias and confounding in public health studies.	The control for observable factors to minimize bias is emphasized in Topic 1-4, and control for unobservable confounding factors is discussed when covering the fixed effects model (Topic 9) and instrumental variables model (Topic 13).	The understanding of techniques to minimize bias attributable to confounding variables is assessed in Quizzes 1, 2, 3, and the final examination.