EDUC 645 Syllabus & Schedule

# Student Learning Outcomes

By the end of this term, it is expected that students will be able to:

1. Conduct and interpret regression analyses with dichotomous and count outcomes and predictors in response to independently developed research questions.
2. Determine when it is appropriate to conduct analyses with generalized linear models and multilevel models over linear least squares regression models.
3. Describe the assumptions of logistic and Poisson regression analyses and test analytic models for the extent to which they satisfy the models’ assumptions.
4. Compare different approaches to fitting models with predictors that are categorical, interacting, and that have non-linear relationships with the outcome variable.
5. Explain how to apply regression models to analyze clustered data for linear and non-linear outcomes.
6. Analyze discrete and continuous outcomes using multilevel modeling.
7. Conduct diagnostics to assess the fit and use of multilevel modeling.
8. Describe the purposes for assessing inter-rater reliability (e.g., ICC) and item discrimination.
9. Evaluate measurement data using classical-test-theory reliability, generalizability theory, and item analysis.
10. Explain the applications and limitations of different methods for treating missing data.
11. Implement modern methods for examining and treating missingness of data in the GLM framework.
12. Perform all analyses using the R programming language.

# Textbooks and Reading Materials

## Textbooks

We will primarily be referring to chapters in [Beyond Multiple Linear Regression](https://bookdown.org/roback/bookdown-BeyondMLR/) (**BMLR**) by Paul Roback and Julie Legler. We will also rely on a couple of chapters in [Introduction to Educational and Psychological Measurement Using R](https://www.thetaminusb.com/intro-measurement-r/) (**IEPM**) by Tony Albano. Both textbooks are available for free online. You may choose to purchase a paper copy of **BMLR** if you wish, but it is not required.

## Supplemental reading materials

Further readings will be provided in the form of peer-reviewed journal articles. Links to these papers can be found in the schedule.

## R and RStudio

Students must have the latest version of R, which can be downloaded [here](https://ftp.osuosl.org/pub/cran/). It is strongly recommended that students also download the RStudio GUI, available [here](https://www.rstudio.com/products/rstudio/download/#download). Both softwares are free. We will provide tutorials on R/RStudio installation and they are also accessible [here](.slides/tutorial.html).

### Resources for R and RStudio

While we will teach you how to effectively use R and RStudio to conduct analyses, one of the key skills required to use R is the ability to find answers on your own. Many common questions or problems are either available on blogs or have been asked and answered in discussion forums already. Finding and deciphering those answers is an important skill you should seek to hone. *You will never remember all of the programming commands!*

Here are some sites where you can find the answers to many R questions and learn new tricks:

* [*YaRrr, the Pirate’s Guide to R*](https://bookdown.org/ndphillips/YaRrr/) by Nathaniel Phillips
* [*R for Data Science*](https://r4ds.had.co.nz/) by Hadley Wickham
* [*R Cookbook*](http://www.cookbook-r.com/) by Winston Chang
* [*An Introduction to Statistical Learning*](http://www-bcf.usc.edu/~gareth/ISL) by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
* [Cheat Sheets](https://www.rstudio.com/resources/cheatsheets)
* [Quick-R](https://www.statmethods.net)
* [StackOverflow](https://stackoverflow.com/questions/tagged/r)
* [Stack Exchange](https://stats.stackexchange.com)

# Schedule

For more details, see [here](./schedule.html).

| Units | Uniti Topics | Weeks | Lecture Topics | Lab Objectives | Required Readings | Recommended Readings | Assignments | Quizzes |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Logistic Regression | 1 | a. Review OLS | a. R basics and R packages | BMLR Ch. 2 (pp. 39-68), 6.1-6.4 (pp. 151-155) | BMLR 3.3 ‘Discrete Random Variables’ (pp. 72-79) |  |  |
|  |  |  | b. Dichotomous outcomes | b. Visualize/describe dichotomous data when explanatory variables are continuous, categorical, and interactions |  |  |  |  |
|  |  | 2 | a. Interpreting logistic regression coefficients | a. Logistic regression programming in R | BMLR Ch. 6.5-6.7 (pp. 159-180) |  |  |  |
|  |  |  | b. Significan testing and model fit |  |  |  |  |  |
|  |  | 3 | a. Diagnostics | a. Logistic regression programming in R (cont’d) |  |  |  |  |
|  |  |  | b. Classification and accuracy | b. Preview Unit 1 assignment/quiz |  |  | Assignment 1 (Due: TBD) | Quiz 1 (Due: TBD) |
| 2 | Poisson Regression | 4 | a. Count data and probability | a. Review Unit 1 assignment/quiz | BMLR Ch. 4.1-4.4 (pp. 93-112) |  |  |  |
|  |  |  | b. Poisson regression analysis | b. Visualize/understand Poisson data |  |  |  |  |
|  |  | 5 | a. Coefficient estimation and interpretation | a. Poisson regression programming in R | BMLR Ch. 4.5-4.10 (pp. 113-132), Ch. 5 (pp. 145-148) |  |  |  |
|  |  |  | b. Significance tests | b. Preview Unit 2 assignment/quiz |  |  | Assignment 2 (Due: TBD) | Quiz 2 (Due: TBD) |
| 3 | Nested Data | 6 | a. Clustering within datasets | a. Review Unit 2 assignment/quiz | BMLR Ch. 8.1-8.5 (pp. 211-231) |  |  |  |
|  |  |  | b. OLS | b. Visualize/understand nested data |  |  |  |  |
|  |  | 7 |  | a. Nested data and OLS estimator | BMLR Ch. 8.6-8.11 (pp. 234-251) |  |  |  |
|  |  |  |  | b. Preview Unit 3 assignment/quiz |  |  |  |  |
|  |  | 8 |  | a. Nested data and Logistic regression | BMLR Ch. 9.1-9.7 (pp. 263-306), Ch. 11 (pp. 373-398) |  |  |  |
|  |  |  |  | b. Preview and prepare for final project |  |  | Assignment 3 (Due: TBD) | Quiz 3 (Due: TBD) |
| 4 | Measurement and Assessment | 9 |  | a. Review Unit 3 assignment/quiz | BMLR 7.1-7.8 (pp. 193-206) |  |  |  |
|  |  |  |  | b. Measurement and assessment in R |  | IEPM Ch. 5 |  |  |
|  |  |  |  | c. Preview Unit 4 assignment/quiz |  |  | Assignment 4 (Due: TBD) | Quiz 4 (Due: TBD) |
| 5 | Missing Data | 10 | Data cleaning and handling missing data | a. Review Unit 4 assignment/quiz |  |  |  |  |
|  |  |  |  | b. Dealing with missing data in R |  |  | No assignment | Quiz 5 (Due: TBD) |
|  | Final | 11 |  |  |  |  | Final Project |  |

# Grading Components and Criteria

Final grades will be based on the following components:

* Quizzes: 20% (5 quizzes 4% each)
* Assignments: 60% (4 assignments, 15% each)
* Final: 20%

## Quizzes

Five classes will begin with a very short multiple-choice quiz that is designed to test your knowledge of the theoretical principles underlying the statistics we are studying for the week. While some may feel that this is overly paternalistic, research evidence shows that frequent quizzing increases learning (see a summary of [one study](https://www.nytimes.com/2013/11/21/education/frequent-tests-can-enhance-college-learning-study-finds.html) from the University of Texas). Quizzes will be open book and notes (but not laptop/computer).

## Assignments

The goal of the assignments is to practice the concepts and vocabulary we have been modeling in class and implement some of the techniques we have learned. Each assignment has an associated data source, short codebook and detailed instructions for the required data and analytic tasks. You may work on your own or collaborate with one (1) partner. Please make sure that you engage in a a full, fair and mutually-agreeable collaboration if you do choose to collaborate. If you do collaborate, you should plan, execute and write-up your analyses together, not simply divide the work. Please make sure to indicate clearly when your work is joint and any other individual or resource (outside of class material) you consulted in your response. Further assignment details are available [here](./assignments.html).

# 

# Student Engagement Inventory

Graduate students are expected to perform work of high quality and quantity, typically with forty hours of student engagement for each student credit hour. For this course, the following table shows the number of hours a typical student would expect to spend in each of the following activities:

| Educational activity | Hours | Explanatory comments |
| --- | --- | --- |
| Class attendance | 30 | Three hours per week over 10 weeks |
| Class reading and prep | 30 | Includes reading and review of slides |
| Homework Assignments | 20 | Five assignments taking 5 hours each (on avg.) |
| Quizzes | 5 | One hour per quiz (40 min preparation, 20 min completing quiz) |
| Lab attendance | 10 | One hour-long lab per week over 10 weeks |
| Final project | 25 | Includes familiarization with data, data analysis, preparation of displays and writing |
| **Total hours** | **120** | **These are approximations.** Reading and *especially* analytic time will vary per individual |

# Course Schedule

## Unit 1: Dichotomous Outcomes

**Learning Objectives:**

* Describe the linear probability model and discuss its limitations
* Differentiate between the concepts of probability, odds, and logit
* Estimate and interpret model parameters logistic regression model with continuous and categorical predictors
* Discuss predictive power, model fit, and diagnostics for logistic regression models

**Readings**:

* BMLR Chapter
* Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67-82. 861-874.[https://doi-org.libproxy.uoregon.edu/10.1093/esr/jcp006](https://www.jstor.org/stable/40602478)

**Lecture**:

* Lecture slides: [html](./slides/EDUC645_log_regression.html)
* Dataset: [principal.csv](./data/principal.csv)

**Assignment**:

* Assignment description: [assignment1.html](./assignments/assignment1.html)
* Dataset: [nscs.csv](./data/nscs.csv)

## Unit 2: Count Outcomes

**Learning Objectives:**

* Apply regression models to analyze count and rates as outcome variables
* Describe situations which merit the use of Poisson regression
* State the assumptions about rates in Poisson regression
* Use Poisson regression model with continuous and categorical predictors, and interpret the model parameters and inference for model parameters
* Evaluate dispersion of the outcome variable, model fit, and diagnostics for a Poisson regression model

**Readings**:

* BMLR Chapter
* Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121-136. [https://doi.org/10.1080/00223890802634175](https://doi-org.libproxy.uoregon.edu/10.1080/00223890802634175)

**Lecture**:

* Lecture slides: [html](./slides/EDUC645_poisson_regression.html)
* Dataset: [ceps\_sch.csv](./data/ceps_sch.csv)

**Assignment**:

* Assignment description: [assignment2.html](./assignments/assignment2.html)
* Dataset: [ah\_smoke.csv](./data/ah_smoke.csv)

## Unit 3: Nested Data Structures

**Learning Objectives:**

* Describe nested data structures (two-level, e.g., repeated measures within individuals, students within schools)
* Demonstrate how intercepts, slopes, and regression coefficients may vary across clusters
* Compare and contrast using fixed-effects and random-effects in regression models
* Explain the covariance matrix of random-effects and describe implications of ignoring random-effects in terms of the inference for fixed-effects and statistical power
* Measure within-cluster dependency using intra-class correlation, and describe how to evaluate whether multilevel modeling is appropriate
* Create fixed and random intercept-only models and interpret the model parameters using the R lmer package
* *Time permitting*: Discuss alternatives to multilevel modeling of nested data

**Readings**:

* BMLR Chapter

**Lecture**:

* Lecture slides: [html](./slides/EDUC645_nested_data.html)
* Dataset: [seda.csv](./data/seda.csv) and [portugal.csv](./data/portugal.csv)

**Assignment**:

* Assignment description: [assignment3.html](./assignments/assignment3.html)
* Dataset: [gpa.csv](./data/gpa.csv) / [gpa\_above.csv](./data/gpa_above.csv)

## Unit 4: Measurement and Assessment

**Learning Objectives:**

* Differentiate and define different aspects of measurement and understand when they apply
* Demonstrate how to compute and interpret Cronbach’s alpha
* Define the assumptions and limitations for methods to assess reliability

**Readings**:

* BMLR Chapter
* IEPM Chapter

**Lecture**:

* Lecture slides: [html](./slides/EDUC645_measure_assess.html)
* Dataset:

**Assignment**:

* Assignment description: [assignment4.html](./assignments/assignment4.html)
* Dataset:

## Unit 5: Missing Data

**Learning Objectives:**

* Discuss the concept of model-based imputation in the context of linear regression, logistic regression, and Poisson regression models
* Identify and describe missing data mechanisms
* Differentiate listwise and pairwise deletion methods from single-imputation methods
* Describe the “missing-value-dummy” method (Cohen) and some issues with this approach
* Compare and contrast the conceptual steps for the multiple-imputation with using full-information maximum likelihood estimators to account for missing data
* Using the R mice package to practice applying multiple-imputation to fit linear, logistic, and Poisson regression models for a data set with missing values

**Readings**:

* Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5-37. [https://doi.org/10.1016/j.jsp.2009.10.001](https://pubmed.ncbi.nlm.nih.gov/20006986/)

**Lecture**:

* Lecture slides: [html](./slides/EDUC645_cleaning_missing.html)
* Dataset: [galo.csv](./data/galo.csv)

**Assignment**:

No assignment for this unit.

**Final Project**:

* Final project description: [final.html](./assignments/final.html)
* Dataset: