MA482 Course Design

Thursday, December 3, 2020

Course-Level Objectives

As in any statistics course, we emphasize statistical literacy (interpretation and clear communication of statistical methods, results, and concepts) and statistical reasoning (modeling variability in a process, defining the need for data to address questions, choosing appropriate methodology, and critiquing an analysis). Specifically, after taking this course, students will be able to accomplish the following tasks:

(A) **Describe** situations for which multi-predictor regression models are needed to address the research question of interest. Specifically, **describe** the role multi-predictor regression models play in isolating the effect of a variable and investigating the interplay between multiple variables.

(B) **Compare** and **contrast** the four primary regression modeling techniques (linear, non-linear, survival, or repeated measures); and, given an analysis situation, **state** the appropriate regression modeling technique and **justify** your choice.

(C) **Formulate** research questions as measurable statements about parameters in a regression model.

(D) Given a research question from the biological sciences, use appropriate software to **conduct** inference on the corresponding parameters and **interpret** the resulting output in context of the research question.

(E) Clearly **communicate** an analysis and its implications using a variety of media: written paper, scientific poster, scientific abstract, and oral presentation.

(F) **Collaborate** with others to formulate a statistical analysis plan for addressing a research question.

(G) **Appreciate** the value and limitations of regression modeling for addressing research questions in the biological sciences.

(H) **Express** a desire for researchers in the biological sciences to be trained in statistical thinking and literacy.

(I) **Assess** the strength of evidence presented by a scientific publication in addressing a research question and **provide** constructive feedback for improving a study.

These are based on Fink's Significant Learning Outcomes (http://www.buffalo.edu/ubcei/enhance/designing/learning-outcomes/finks-significant-learning-outcomes.html):

- Foundational Knowledge - (A) and (B)
- Application - (C) and (D) and (E)
- Integration - (B)
- Human Dimension - (F)
- Caring - (G) and (H)
- Learning to Learn (I)

Course Alignment

Miscellaneous Course Ideas

Structure:

Course Alignment

Timing for Primary Modules:

1. In-Class Activity: 2-5 hours (balanced with HW, 7 total)
2. Reading/Videos/Guided Notes: 6 hours
3. Article Review: 2 hours (1 in-class, 1 outside)
4. Homework Assignments: 2-5 hours (balanced with in-class activities, 7 total)
5. Module Quiz: 2 hours
6. Concept Check: 0.5 hours
7. Analysis Task: 2 hours
8. Time towards capstone project: 2.5 hours

Total Time per Module: 22 hours
Total Time per Capstone Project: 10 hours

Course Objective F will be supported through the capstone project.

Course Objectives G and H will be assessed through optional writing assignments which earn tokens for revisions in the course.

Course Objectives E and I will be supported through Article Reviews.

Module 0: Basic Statistics and Probability
An introductory statistics course is designed to increase your statistical literacy (understanding and communicating statistical results) and statistical reasoning ("thinking with data"). The class is not designed to make you an expert analyst, but the concepts discussed there permeate all of statistical methodology. We take a moment to review some of the most important concepts covered in that course. We also introduce basic probability for those without prior exposure; probability is useful for modeling distributions, which we study extensively in statistical modeling.

Objectives	***Activities***	***Assessments***
Describe the four key distributions involved in statistical inference: *distribution of the population, distribution of the sample, the sampling distribution of a statistic, the null distribution of a statistic*.	Concept, Example	Automated Moodle quiz for identification.
Given a dataset, **compute** numerical summaries of a variable using the course software.	Concept, Example, Coding	Automated Moodle quiz with computation.
Given a dataset, **construct** a graphical summary of the distribution of a variable using the course software.	Concept, Example, Coding	Automated Moodle quiz for identification.
Describe the role of a *density function* for modeling a probability distribution.	Concept, Example	
Given a common model for a probability distribution, **compute** a probability about a random variable using the course software.	Concept, Example, Coding	Automated Moodle quiz with computation.
Contrast a *discrete random variable* and a *continuous random variable*.		

Completion of assessments in this section will earn ability to revise and resubmit further assignments.

Distributional Quartet:
Cover an example of making a graphic and running through the Distributional Quartet. This should be in

Miscellaneous Course Ideas

Structure:
Each module will be broken into roughly three sections. Each section will have lectures, an in-class activity, and a homework assignment. The module will culminate with the Article Review.

Lectures:
Lecture should divide up the conceptual, the applied, and the implementation.

Class Activities/Discussions:
Each section of the module will have a class activity/discussion. These will focus on the most challenging concept of the section, generally with an exercise that helps make the concept less abstract and an example to work through together. Time permitting, students will be able to begin their homework assignment in class.

Forums:
Use forums to handle the elements of the capstone project; this will primarily be done asynchronously. On days when we have Article Review discussions, we will also discuss the project a bit to continue progress.

Assessments:
Homework Assignments will be low-stakes assessments, primarily for giving feedback to students practicing the course concepts. Module Quizzes will assess primarily concepts, though some computation, related to the course material. Concept Checks will primarily assess interpretation of results. Analysis Tasks will assess the complete analysis of a problem from framing the question, graphically summarizing the data, performing an analysis, and reporting the conclusions; these will be individual assessments.

Project:
If possible, have several groups working on the same project but with their own data. This allows us to share ideas but implement unique variations. Possible ideas include:

- Comparing fruit savers (salt water, sprays, lemon juice, etc.). These are meant to prevent sliced apples from "browning." This would have repeated measures and be longitudinal. We would need time-lapse photography and a way of measuring "browning" (pixel color?)
- Comparing music or talking to plants. While this has been done, we could look at some variation (like the topic of conversation) with regard to how quickly plants grow. This would be repeated measures with nonlinear growth.
- Impact of types of water (tap, rain, river, bottled) on plant growth. Likely nonlinear growth with repeated measures.
- The impact of weather on the number of walk-in appointments at the health clinic on campus (may need approval for this?). Likely linear models.
- Water conservation if we urinated in the shower. This will require an IRB waiver (just a survey, so should be exempt). This is likely linear models.
- Impact of homemade fertilizer on plant growth. Likely repeated measures with nonlinear growth.
- Impact of soda (different time, different sugar content, diet vs. regular, etc.) on cavities in eggs.

completion of assessments in this section will earn ability to revise and resubmit further assignments.

Distributional Quartet:
Cover an example of making a graphic and running through the Distributional Quartet. This should be in the context of one-sample inference or simple linear regression.

Probability Modeling:
Cover an example of modeling these distributions with probability and how that is helpful. This should include the idea of modeling the sampling distribution through the Central Limit Theorem.

Project:
Project groups (potentially) are formed. The project topic is given to students with an initial piece of literature to read. Students will then gather a few additional sources of literature with a key point that should be considered as a result.

Module 1: General Linear Model
The general linear model, also referred to as multiple linear regression, provides a framework appropriate for modeling a continuous outcome (response) as a function of several predictors (covariates). This module serves as a unifying framework to the topics discussed in an introductory statistics course. It also provides a platform for introducing several flexible modeling strategies and the foundation of our modeling approach in the class.

Objectives	***Reading***	***Activities***	***Assessments***
Given an analysis situation, **decide** if linear regression would appropriately address the research objective. (Supports Course Objectives A and B)	Ch 4		
Translate research questions appropriate for a regression model into specific questions about the coefficients of the model. (C)	Ch 4	Example	Homework, Module Quiz, Concept Check, Analysis Task
Using appropriate software, **test hypotheses** about relationships between variables in a linear regression model, including confounding and interaction. (C, D)	Ch 4.3, 4.4, 4.6	Example, Coding	Homework, Module Quiz, Analysis Task
Identify the key conditions of the "classical" linear regression model and **describe** their implications. (A, B, D)	Ch 4.7	Concept	Homework, Module Quiz
Define the term *fitted value* and *residual*. (D)	Ch 4.7	Concept, Coding	
Given data, **assess** the key conditions of the "classical" linear regression model. (D)	Ch 4.7	Example, Coding	Homework, Analysis Task
Interpret regression output (coefficient estimates, p-values, confidence intervals, and statistics summarizing the quality of the model fit) in the context of the research objective. (D, E)	Ch 4	Concept, Example	Homework, Module Quiz, Concept Check, Analysis Task
Describe three approaches for modeling the sampling distribution of an estimator (or the null distribution of a standardized statistic): exact, large-sample theory, bootstrapping. (B, D)		Concept	
Define the term *indicator variable* and **describe** its use in regression modeling. (C)		Concept, Example, Coding	Homework
Given a hypotheses, **formulate** it within the *General Linear Hypothesis Testing Framework*, if appropriate. (C)		Concept, Example, Coding	Homework, Module Quiz, Analysis Task
Describe and **implement** (given data) an approach for relaxing		Concept, Example, Coding	Homework, Module Quiz, Analysis

in eggs.

Planned Videos

Module 0:
- C: Distributional Quartet
- R: Introduction to software (similar to 223, but shorter)
- C: Density function
- E: Computing a probability, density to CDF
- R: Computing a probability
- E: Computing a quantile
- R: Computing a quantile

Module 1:
- C: Alternate Characterization (focus on idea of capturing parameters)
- E: Interpretation of Parameters
- R: Interpretation of Parameters
- E: Inference on Parameter
- R: Inference on Parameter
- E: Assessing Mean 0
- R: Assessing Mean 0
- E: Assessing Constant Variance
- E: Assessing Independence
- R: Assessing Independence
- E: Assessing Normality
- R: Assessing Normality
- E: Assessing Linearity
- R: Assessing Linearity
- C: Confounding
- C: Categorical Predictors
- E: Categorical Predictors
- R: Categorical Predictors
- C: Interaction Terms
- E: Interaction Terms
- R: Interaction Terms
- C: General Linear Hypothesis Tests (relationships, ands)
- E: General Linear Hypothesis Tests
- R: General Linear Hypothesis Tests
- C: Bootstrapping (residual and case?)
- E: Large Sample Theory
- R: Large Sample Theory
- C: Splines
- E: Splines
- R: Splines

Module 2:
- E: Distinguishing fixed/random effects 1
- E: Distinguishing fixed/random effects 2
- C: Sources of Variability
- E: Specifying an individual-level model
- E: Specifying a population-level model
- E: Specifying a mixed-effects model
- R: Fitting mixed-effects model
- C: Correlation Structures
- C: Robust Variance-Covariance Matrix
- E: Specify a population-averaged model

Given a hypotheses, **formulate** it within the *General Linear Hypothesis Testing Framework*, if appropriate. (C)		Concept, Example, Coding	Homework, Module Quiz, Analysis Task
Describe and **implement** (given data) an approach for relaxing the condition of "normality" and "linearity" within the linear regression framework. (D)		Concept, Example, Coding	Homework, Module Quiz, Analysis Task

Model Foundations:
Focus on the interpretation and allowing for causal inference in the presence of confounding. Also include indicator variables, which may not be familiar to everyone. Divide the class up based on stereotypes and then talk about pairing which happens when we "hold all other things constant."

Effect Modifications and General Linear Hypothesis Testing:
Focus on interpretation of interaction terms. Using the idea of testing an interaction with a multi-level categorical variable, give example of general linear hypothesis test. Consider the relationship between the height and weight among 3 athletic sports?

Relaxing Conditions:
Focus on relaxing linearity. Discuss relationships which we believe are not linear, but the exact relationship is unknown (try to avoid exponential growth, for example). Then, talk about how we can use splines to model the relationship that is unknown. Circle back to using the general linear hypothesis test to assess whether a spline term is necessary.

Article Review:
Nevitt (2001), "The Effect of Estrogen Plus Progestin on Knee Symptoms and Related Disability in Postmenopausal Women."

Project Discussion:
Project discussion will focus on developing the data collection procedure.

Module 2: Repeated Measures
When the response is measured at multiple times on the same subject, we refer to this as repeated measures. This induces a relationship among the responses that violates the assumption of independence often made during an analysis. This relationship must be addressed in the modelling stage if the standard errors produced are to be relied upon. Further, careful consideration of the study design and use of such analyses can improve the power of a study in many situations.

Objectives	***Reading***	***Activities***	***Assessments***
Recognize a hierarchical data situation and **explain** the consequences of ignoring it. (Supports Objective B)	Ch 7.1, 7.2	Example	Concept Check
Compare and **contrast** *generalized estimating equations* and *mixed effects models* for addressing the *correlation structure* induced in hierarchical data. (A, B)	Ch 7.2-7.5		
Given a description of a data collection procedure and some discipline knowledge, **determine** whether a variable is a *fixed* or *random* effect. (C, D)	Ch 7.5	Example	Homework, Module Quiz
For generalized estimating equations, **discuss** the role of the *working correlation structure* and the *robust variance-covariance estimator.* (D)	Ch 7.4	Concept, Example, Coding	
Given a description of a data collection procedure, **select** a working correlation structure and **justify** the selection. (D)	Ch 7.4	Example	Homework, Module Quiz, Analysis Task (?)
Interpret regression output (coefficient estimates, p-values,	Ch 7	Example, Coding	Homework, Module Quiz, Analysis

- R: Fitting mixed-effects model
- C: Correlation Structures
- C: Robust Variance-Covariance Matrix
- E: Specify a population-averaged model
- R: Fitting a population-averaged model

Module 3:
- C: Need for nonlinear models vs. flexible linear framework
- R: Fitting a nonlinear model with built-in functions
- R: Fitting a nonlinear model from scratch
- E: Inference for nonlinear models
- R: Inference for nonlinear models
- C: Allowing relationships to vary
- E: Allowing the relationship to vary
- R: Allowing the relationship to vary
- R: Plotting the results of a nonlinear fit
- C: Types of bootstrapping
- R: Wild bootstrap
- C: Why logistic regression
- C: Likelihood
- E: Fitting a logistic model
- R: Fitting a logistic model
- C: Interpreting the coefficients of a logistic model
- E: Computing the odds ratio
- R: Computing the odds ratio
- E: Comparing models
- R: Comparing models

Module 4:
- C: Ways of modeling survival (graph illustrating survival vs. hazard)
- C: Staggered entry and right-censoring (impacts on estimation)
- E: Life-table estimate/inference
- R: Life-table estimate/inference
- E: Kaplan-Meier estimate
- R: Kaplan-Meier estimate
- E: Log-rank test
- R: Log-rank test
- C: Proportional Hazards assumption
- E: Cox PH model with inference
- R: Cox PH model with inference

Zoey:
2. I've Got the Music in Me - probability is behind what we do.
3. I'm Gonna Be (500 Miles) - linear regression will go the distance.
4. I Wanna Dance with Somebody - interactions…together stronger
5. Stronger - letting go of conditions
6. Just Give Me A Reason - knowing why the variability is there can help us adjust…it's not broken, just bent
7. The Boy is Mine - we need to make the decision about whether each term should be allowed to vary or not; they fought over him.
8. All I do is Win - robust variance-covariance matrix
9. Numb - little less like linear models
10. Tightrope - when letting a model vary across groups, we let parameters depend on another variable (interactions), but we should consider the question carefully, don't go overboard.
11. Happier - in order to handle binary stuff, we need to move beyond the basics and embrace a future that is different (maximum likelihood)

Given a description of a data collection procedure, **select** a working correlation structure and **justify** the selection. (D)	Ch 7.4	Example	Homework, Module Quiz, Analysis Task (?)
Interpret regression output (coefficient estimates, p-values, confidence intervals, and statistics summarizing the quality of the model fit) in the context of the research objective. (D, E)	Ch 7	Example, Coding	Homework, Module Quiz, Analysis Task
Discuss the benefits of considering repeated measures when designing a study. (F)		Concept	
Using appropriate software, **test hypotheses** about relationships between variables in a repeated measures model using either generalized estimating equations or a mixed effects model, including confounding and interaction. (C, D)		Example, Coding	Homework, Module Quiz, Analysis Task

Sources of Variability:
Need to really spend time discussing how we partition various sources of variability. This should primarily be in a video, but this day should be running through an example and trying to determine the various sources of variability and classifying them. Maybe speed at which someone can send a text message as a function of the length of the text message.

Mixed Effects Models:
Focus is on constructing the subject-level and population-level model. Continue with the text messaging example, specifying the various models and fitting it in class.

Generalized Estimating Equations:
Focus on selecting a working correlation structure and then creating a marginal model. Continue with the text messaging example, specifying the correlation structure and model and fitting it in class.

Article Review:
Mystery article with a discussion about the refereeing process.

Project Discussion:
Project discussion will focus on updating the group on the status of the data collection and thinking about key summaries to create; drafting the beginning portion of the statistical analysis plan by considering demographic tables, for example.

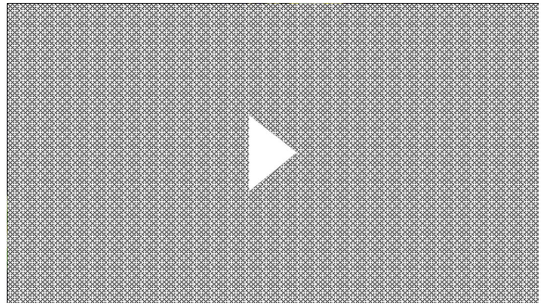
Module 3: Nonlinear Models
Models which are nonlinear in the parameters have applications to cellular biology, ecology, chemical engineering, and more broadly when modeling categorical data (such as when the response is binary). Embedding nonlinear models into a statistical framework allows us to make inference on the underlying parameters. In this unit, we examine such models and discuss logistic regression in particular. We also discuss extensions to repeated measures data and touch on model selection for nonlinear regression models.

Objectives	***Reading***	***Activities***	***Assessments***
Identify a nonlinear model and **describe** situations in which nonlinear models are necessary, including logistic regression. (Supports Objective B)		Concept	Module Quiz, Concept Check
Translate research questions to statements about model parameters in a nonlinear model. (C)		Example	Homework, Module Quiz, Analysis Task
Interpret regression output (coefficient estimates, p-values, confidence intervals, and statistics summarizing the quality of the model fit) in the context of the research objective. (D, E)		Example, Coding	Homework, Module Quiz, Analysis Task

overboard.
11. Happier - in order to handle binary stuff, we need to move beyond the basics and embrace a future that is different (maximum likelihood)
12. One Call Away - handling censored data just requires making the right call, and Life Table methods are the building block to that.
13. Carry On - we can extend Life Table methods to handle individual censoring.
14. Juice - this model is hugely popular because it is extremely flexible; the recognition is well-deserved.
15. American Pie or I want you to want me or Sucker - invokes nostalgia and a sense of losing the good stuff…I hope you look back fondly on this class in the future. I want you to want statistics; I want you to be a sucker for modeling.

Using appropriate software, **test hypotheses** about relationships between variables in a nonlinear model, including whether the model parameters should vary across specific groups. (C, D)		Example, Coding	Homework, Module Quiz, Analysis Task
State the relationship between odds ratios and the parameters in a logistic regression model. (C, D, E)	Ch 5.1	Concept, Example, Coding	Homework, Module Quiz
Describe the semiparametric moment-model approach taken in modeling nonlinear models compared to a fully parametric approach. (B, D)		Concept	
Describe an approach, and **implement** it, for relaxing the condition of constant variance in a nonlinear model. (B, D)		Concept, Example, Coding	Homework, Analysis Task (?)
Compare and **contrast** logistic regression with other nonlinear models. (B)	Ch 5.1	Concept	
Describe the tension at play in information criteria used for model selection. (D, E, I)		Concept	
Using model selection summaries, **determine** which of two models better fits the data. (D, E, I)		Example, Coding	Homework, Module Quiz

Nonlinear Modeling Framework:
Revisit the idea of specifying the mean and variance (constant for now) and using least squares for estimation. Emphasize that inference is based on large-sample theory. Maybe we could examine the growth of a sunflower over time (Sunflower growing time lapse 42 days of growing - 4k)? I can't think of another great nonlinear thing we could gather in class directly. Could also do my body temperature over the course of a day (and Jamie's for comparison?)



Allowing Parameters to Vary Across Groups:
Altering the parameters across groups is really a modeling exercise, but the implementation is often challenging. We want to start with the example we had from the previous framework and allow it to vary across several groups and implement it.

Logistic Regression:
Really focus on the need for something different with binary response. Also highlight that the distribution is known. Therefore, we can do better than least squares. We can estimate the probability of a heads for pennies when spun vs. flipped? We could ignore the repeated measures of people here. While this could be done without logistic regression, the comparisons of the various probabilities with easy computations could be useful. We could then extend the idea to a continuous predictor.

Article Review:
Pharmacokinetics of Ibuprofen in children. This has a good illustration of **not** using nonlinear methods and we can talk about the impacts here. Could compare this to the poster a student did a few years ago in which nonlinear methods were used.

Project Discussion:
Project discussion will focus on sketching out the analysis plan as well as dealing with any data-cleaning issues. We will have a good idea of where the analysis should be taken, expecting a draft soon.

Module 4: Survival Analysis
Many studies involve studying the time until an event occurs. Unfortunately, in biological settings, the event is often not observed for all subjects, a phenomena referred to as censoring. In this module, we examine methods for addressing censored data. In particular, we look at nonparametric approaches leading up to the Cox Proportional Hazards model, an extension of regression which accounts for the censoring.

Objectives	***Reading***	***Activities***	***Assessments***
Define key characterizations in time-to event models: *survival function, mortality rate, hazard function*. (Supports Objectives C, D)	Ch 3.5	Concept, Example	
Identify various types of censoring: *left, right, interval*. (A, B)	Ch 3.5	Concept, Example	Homework, Module Quiz
Describe situations in which survival analysis techniques are necessary, including *life tables, Kaplan-Meier curves and log-rank tests,* and *Cox Proportional Hazards models*. (B)	Ch 3.5, 6.1		
Describe the impact of censoring on traditional analyses. (A, B)	Ch 6.1	Concept	Concept Check
Using appropriate software, **construct** and **interpret** a life-table estimate of survival. (D)	Ch 3.5	Concept, Example, Coding	Homework, Module Quiz
Using appropriate software, **construct** and **interpret** Kaplan-Meier estimates of a survival curve. (D)	Ch 3.5	Concept, Example, Coding	Homework, Module Quiz, Analysis Task
Using appropriate software, **construct** and **interpret** a log-rank test for comparing multiple survival curves. (D)	Ch 3.5	Concept, Example, Coding	Homework, Module Quiz
Describe the assumption of *proportional hazards*. (B, C, D, E)	Ch 6.1	Concept, Example	
Describe the primary benefits of the semiparametric Cox PH model. (B)	Ch 6.2	Concept	
State the relationship between hazard ratios and the parameters in a Cox PH model. (C, D, E)	Ch 6.2	Concept, Example, Coding	Homework, Module Quiz, Analysis Task
Interpret regression output (coefficient estimates, p-values, confidence intervals, and statistics summarizing the quality of the model fit) in the context of the research objective. (D, E)	Ch 6.2	Example, Coding	Homework, Module Quiz, Analysis Task
Using appropriate software, **test hypotheses** about relationships between variables in a Cox PH model, including interactions and after adjusting for confounders. (C, D)	Ch 6.2	Example, Coding	Homework, Module Quiz, Analysis Task
State the common pitfalls and conditions for the Cox PH model. (E, G)		Concept	Module Quiz

Life-Table Methods:
Discuss big impacts of censoring, but also really focus on optimistic and pessimistic estimation methods. This leads to the life-table estimate. We could consider the "length of time a penny spins," as our primary method, but consider data censored if it falls off the table or hits something. We can talk about binning things in every 10-second intervals. While we wouldn't want to bin data necessarily, we can see how only looking every 10 seconds would lead to this particular data structure.

KM Estimates and Log-Rank tests:

Using the penny-spin data, continue by creating KM estimates. Maybe we can do this for two groups of students (athletes and non-athletes?). We can compare the survival curves using the log-rank test. We really need to emphasize the hypothesis in the log-rank test.

Cox PH Model:
The key idea of proportional hazards should be emphasized and the interpretation. We can take the same penny-spin data and model it. With such a simple model, we should be able to see if the assumption of proportional hazards is met and discuss the interpretation. The common pitfalls should be emphasized here.

Article Review:
Hannan (2005). This paper is a good discussion on adjusted vs. unadjusted comparisons and the impacts that can have.

Project Discussion:
Project discussion will focus on discussing the results and writing up the paper completely.

Module 5: Projects
Each module has been self-contained. In this final module, we simply bring the ideas together while focusing on the blending of analysis plans and study design. This also provides a chance to finish up the capstone project for the course and outstanding assessments.

This module has no assessments; instead, this focuses on class discussions and data collection activities.

Objectives
Given a research question, **describe** how a multi-predictor regression model covered in class could be used to address the question if data were available. (A, B, C)
Given a research objective, **design** an appropriate study to address the question. (F)
In addition to the primary variables under study, **describe** the importance of collecting additional information related to the research objective. (F, G)

Each of the following idea is a potential in-class lab. Which ones work well depend on the number of students in the class (if we want to actually do the analysis, stick with linear and repeated measures). The class should discuss the entire study design, variable collection, etc. If possible, students should collect the data in class and discuss the analysis techniques. The instructor can email the results to students after the fact.

Memory Association:
Perform a study to assess how well we remember words from a list (though that is not explicitly described as the objective) when we focus on the utility of the word verses some fact about the word (like the number of vowels). Theory suggests that focusing on utility should improve memory. This can be achieved by asking students to focus on utility or structure of several words (without true purpose being revealed). Then, students are asked to identify as many words as possible from the list. This introduces when patients should not be aware of treatment or goal of study and the use of "distractors" to clear short term memory. Primarily linear regression.

Maze and Eye Coordination:
Generally, an image is formed in our brain when our eyes both take in information together. Occasionally, patients develop a "lazy eye" in which the two signals from the eyes are disjoint; so, the brain only records one image at a time. Consider reading red text while wearing red/green glasses; if both eyes work correctly, the text is visible. If the dominant eye is looking through the red lens, the

color negates the text and the text becomes "invisible." We can run an activity in which students complete a maze in red ink while wearing red/green glasses. The maze should be harder (take longer) if one eye is weaker than the other. This can be used to assess paired design since both eyes should be tested. Primarily repeated measures. Or, could be survival analysis if people do not finish and no pairing is used.

Dress that Broke the Internet:
There is a famous image of a dress which looks blue or gold for different people. It is unclear why people see different colors. This could be due to the screen as well as ability to see different colors. Students could complete a survey to find predictors of which color we see. Logistic regression (nonlinear models).

Music on Speed:
How does music impact the speed with which you complete academic tasks? Is music distracting or helpful? Students could be randomized to a particular genre of music and then asked to complete a Sudoku puzzle (or calculus questions) and timed to see how quickly the complete the puzzle. Not everyone will finish. To introduce more random censoring, you could have a warm-up puzzle which is done without music; so, the amount of time spent on the puzzle of interest is different for everyone because they start that one at a unique time. Primarily survival analysis.

Alternate Module 5: Paradoxes
Each module has been self-contained. In this final module, we discuss some of the common paradoxes that are likely to be encountered in practice. This also provides a chance to finish up the capstone project for the course and outstanding assessments.

This module has no assessments; instead, this focuses on class discussions and data collection activities.

Objectives
Given a research question, **describe** how the data could lead to non-intuitive conclusions and how multi-predictor regression model covered in class could be used to address the paradox if data were available. (A, B, C)
In addition to the primary variables under study, **describe** the importance of collecting additional information related to the research objective in order to overcome a paradox. (F, G)

Each of the following idea is a potential paradox that could be discussed.