NUTR 0309: Statistical Methods for Nutrition Research II  
Spring 2022

Class and Lab Meetings: Wednesdays, 9:00am to 12:00pm
Instructors: Mei Chung, PhD, MPH
Email: mei_chung.chung@tufts.edu
Office hours: after class on Wednesdays and by appointment

Teaching Asst.: Bingjie Zhou
Email: bingjie.zhou@tufts.edu
Office hours: after class on Wednesdays and by appointment

Semester Hour Units: 3.0
Prerequisites: Biostatistics I (NUTR 0206) or Statistical Methods for Nutrition Science and Policy (NUTR 0207) or equivalent, and graduate standing or instructor consent. **NOTE: Students cannot receive semester hour units for both NUTR 309 and NUTR 307: Regression Analysis for Nutrition Policy. Students who have not taken Principles of Epidemiology (NUTR 0204) or equivalent course are strongly encouraged to take Principles of Epidemiology (NUTR 0204) concurrently with NUTR 0309.**

Course Description: This course is part two of a one-year, two-semester course on statistical methods for nutrition research (any study designs). The focus of this course is on multiple regression models for continuous, binary, and count data. Emphasis is on developing a conceptual understanding of the application of these techniques to solving problems, and on exposure to statistical modeling concepts and theory (non-mathematical) within the context of nutrition research. In the computer lab sessions, students will use concepts learned during lecture to analyze data using statistical software R and RStudio, STATA, or SAS.

Note: While students are not required to use a particular software program for this course, examples are provided primarily using Stata and R. Students who are using SAS may not receive sufficient support from the instructor. In addition, the statistical examples in the textbook were written in Stata.

Course Goals: Students will learn to analyze quantitative data using multiple regression models. Upon successful completion of the course, students should be able to:

- Gain working knowledge of statistical modeling procedures common in nutrition research.
- Analyze continuous outcomes using multiple linear regression models
- Analyze binary outcomes using logistic regression models
- Analyze count outcomes using poisson regression models
- Analyze complex survey data
- Analyze time-to-event outcomes using survival analysis
- Build and run proper multiple regression models in R and RStudio, STATA, or SAS.
- Document and report data analysis using data management plans and either R scripts, STATA do-files, or SAS code files that are liberally commented.
- Practice communicate statistics professionally and ethically (written and oral).


Note: Statistical examples in Vittinghoff textbook were written in Stata. Other required readings (journal articles) are specified in the syllabus.

Class Materials: All class materials, including lecture notes and assignments, will be posted on Canvas (https://canvas.tufts.edu).

Statistical Software: Students can choose which statistical software they would like to work with: R, STATA, or SAS. R statistical software and RStudio, an integrated development environment for R, will be used for this course. The Tufts Data Lab has instructions for downloading R and RStudio. R is available
on the computers in the Jaharis student room, in the Boston Data Lab (Sackler 510 and 514), and on laptops at the HHSL library. R is also available on the computers at the Eaton computing lab and the Data Lab on the Medford campus.

STATA: Tufts provides an enterprise site license for Stata SE enabling students, faculty, and staff to install Stata SE on their Tufts-owned or personal computer for educational use. Follow the self-service installation instructions at the site provided below or contact Tufts Technology Services (617-627-3376). You can install a new version or upgrade your existing version to Stata SE 16.1 on your computer by following the steps outlined at [Tufts STATA access](https://access.tufts.edu/stata).

SAS: SAS is a PC-based statistical software suite provided by Tufts for free for students, faculty, and staff. Tufts Technology Services can install SAS on your personal device (in person at either the Boston or Medford campuses, or via remote install) through a request to [Tufts SAS access](https://access.tufts.edu/sas). If you have a Mac, you can run SAS through the [TTS Virtual Lab](https://access.tufts.edu/virtuallab). Contact Tufts Technology Services (617-627-3376) for help with either of these.

**Communication Policy:**
Students should try to seek out information for themselves before contacting the instructor or the TA. If you cannot find your answer, we have included a 'Clarity Thread' Discussion on Canvas. Please go there first to post your question and check if any student has already asked the same question. We will aim to check the Clarity Thread daily for any inquiries, but if other students know the answer they should go ahead and post it. If you need to ask a more personal question, please get in touch with either the TA or instructors by email.

**Classroom Conduct:**
Students are expected to attend all classes and lab sessions. Absences should be explained in writing at least 24 hours before class. Missing more than one or two classes/labs per semester will usually result in substantial underperformance. Students are also expected to read all assigned materials before class and come prepared to participate in class discussions and group activities.

**Academic Conduct:**
Each student is responsible for upholding the highest standards of academic integrity, as specified in the [Friedman School’s Policies and Procedures Handbook](https://access.tufts.edu/policies/procedures) and Tufts University policies ([http://students.tufts.edu/student-affairs/student-life-policies/academic-integrity-policy](http://students.tufts.edu/student-affairs/student-life-policies/academic-integrity-policy)). It is the responsibility of each student to understand and comply with these 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:**
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](https://access.tufts.edu/zoom)) or WebEx ([https://it.tufts.edu/webex](https://it.tufts.edu/webex)) on days when the Boston campus is closed due to pandemic, weather or a temporary cancellation issue. Students should expect to be notified by email 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 or WebEx meeting video and audio will be recorded and posted on Canvas when completed. If an on-campus Examination/Presentation 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.
Diversity Statement:
We believe that the diversity of student experiences and perspectives is essential to the deepening of knowledge in this course. We consider it part of our responsibility as instructors to address the learning needs of all of the students in this course. We will present materials that are respectful of diversity: race, color, ethnicity, gender, age, disability, religious beliefs, political preference, sexual orientation, gender identity, socioeconomic status, citizenship, language, or national origin among other personal characteristics.

Assessment and Grading:
Grading for the course will be based on the following distribution:

<table>
<thead>
<tr>
<th>Components</th>
<th>Proportion of final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation</td>
<td>10%</td>
</tr>
<tr>
<td>Lab participation</td>
<td>10%</td>
</tr>
<tr>
<td>Homework</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm project</td>
<td>20%</td>
</tr>
<tr>
<td>Final project</td>
<td>30%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Final score</th>
<th>Letter grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 97 %</td>
<td>A+</td>
</tr>
<tr>
<td>94 to &lt; 97 %</td>
<td>A</td>
</tr>
<tr>
<td>90 to &lt; 94 %</td>
<td>A-</td>
</tr>
<tr>
<td>87 to &lt; 90 %</td>
<td>B+</td>
</tr>
<tr>
<td>84 to &lt; 87 %</td>
<td>B</td>
</tr>
<tr>
<td>80 to &lt; 84 %</td>
<td>B-</td>
</tr>
<tr>
<td>77 to &lt; 80 %</td>
<td>C+</td>
</tr>
<tr>
<td>74 to &lt; 77 %</td>
<td>C</td>
</tr>
<tr>
<td>70 to &lt; 74 %</td>
<td>C-</td>
</tr>
<tr>
<td>&lt;70</td>
<td>F</td>
</tr>
</tbody>
</table>
Class participation (10%)
Students are expected to attend all classes and read all assigned materials before class. Students will be expected to be active class participants by participating in class discussions and activities.

Lab participation (10%)
Students are expected to attend all labs and complete all lab exercises. Missing a lab without prior agreement with the instructor will result in receiving a 0% for that lab exercise.

Homework (30%)
Four problem sets, each worth 7.5% of the final grade, will be assigned throughout the semester. Guidelines for homework will be discussed during the first class. Homework assignments must be submitted through Canvas prior to the lab session. Students who are unable to complete a problem set on time for any reason should contact the instructor by email at least 48 hours prior to the deadline. Without prior notification, assignments received after their deadline will be penalized by a 20% grade reduction for the first 24 hours and 50% for the next 24 hours. Late submissions will not be accepted if late more than 48 hours.

Midterm project (exploratory model write-up) and peer review (20%)
Scenario: You have been asked, as a statistical expert, to co-author an article. While you may not be a subject matter expert, your co-authors are certainly not experts in statistical modeling. So, you split up the work on the manuscript. Your co-authors will write the article Introduction and Discussion sections, and you will not only carry out the statistical modeling, but you will write the majority of the Statistical Methods section and the entirety of the Results section. For this assignment, you will write the enhanced Statistical Methods and Results sections for a journal article. Thus, you should follow the author guidelines for a major journal. Additional information on the format, grading, peer review instructions, and datasets of the project will be provided in session 5.

Final project (30%)
Students will be creating a scientific poster using BRFSS or NHANES dataset (use of other datasets requires additional approval process with the instructor). The number of points you can earn on the final project is 50, with 39 points based on the scientific poster rubric (see Appendix), 10 points based on your syntax (included with the final poster), and 1 point for your DAG. Detailed instructions are described in the Appendix.
<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wednesday, 1/19</td>
<td>Course overview and expectations</td>
<td>Reply to Welcome Message and Pre-class Questions</td>
<td>Vittinghoff 3.2-3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Review bivariate analyses (correlations, covariance, and simple linear regression)</td>
<td>on the course Canvas site before class</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wednesday, 1/26</td>
<td>Review multiple linear regression models and model diagnostics</td>
<td>Homework 1 assigned</td>
<td>Vittinghoff 4.0-4.3.4; 4.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: correlations and linear regression models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Wednesday, 2/2</td>
<td>Principles of model building – Part 1 (Types of models; research questions and study designs; causal inference; confounding and effect modification)</td>
<td>Homework 1 due; Homework 2 assigned</td>
<td>Vittinghoff 4.4; 4.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: Dealing with variables with non-linear relationships with the outcome variables (very common for nutrient-disease relationships) in regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Wednesday, 2/9</td>
<td>Principles of model building – Part 2: Regression model building techniques</td>
<td>Homework 2 due</td>
<td>Vittinghoff 10.0-10.1.3.2, 10.1.4, 10.2-10.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: regression model building techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Wednesday, 2/16</td>
<td>Logistic regression model building</td>
<td></td>
<td>Vittinghoff 5.1-5.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: Introduce the mid-term project dataset and provide detailed instructions on the mid-term project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Wednesday, 2/23</td>
<td>Logistic regression model: diagnostics</td>
<td></td>
<td>Vittinghoff 5.3-5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: logistic regression building and diagnostics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Wednesday, 3/2</td>
<td>Reporting Guidelines for Statistical Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Students’ presentation of the draft methods of the mid-term project for feedback (ungraded)</td>
<td>Draft methods of the mid-term project due</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Wednesday, 3/9</td>
<td><strong>Mid-term project presentations and peer reviews</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Wednesday, 3/16</td>
<td>Analysis of Complex Surveys</td>
<td>Homework 3 assigned</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: analysis of complex surveys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Wednesday, 3/30</td>
<td>Poisson regression model: building</td>
<td>Homework 4 assigned</td>
<td>Campbell Ch 6, pg 84-97 (scan on Canvas)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: Poisson regression model building</td>
<td>Homework 3 due</td>
<td></td>
</tr>
</tbody>
</table>

**Spring break: 3/20-3/28**
<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
<th>Computer lab: Sample size and power</th>
<th>Homework</th>
<th>Vittinghoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Wednesday, 4/6</td>
<td>Sample size and power</td>
<td></td>
<td></td>
<td>4.8, 5.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: Sample size and power</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wednesday, 4/13</td>
<td>Data Visualizations</td>
<td></td>
<td>Preparation of your final e-poster due</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: Data Visualizations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Wednesday, 4/20</td>
<td>Survival analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer lab: run survival analysis and check proportional hazard assumptions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wednesday, 4/27</td>
<td>Final e-poster Presentations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This schedule is subject to modification at the instructor’s discretion*

**Detailed Description of Course Topics, Assignment Schedule, and the Learning Objectives:**

**Session 1: Course overview and expectations. Review bivariate analyses (correlations, covariance, and simple regression)**

**Learning Objectives:**

During the first part of the class, we will meet and greet and go over the syllabus in detail. Emphasis will be put on the assignments, lab exercises, midterm, and final project, as well as on the expectations for using R/RStudio, STATA, or SAS, and Canvas. During the second part of the class, we will review correlations, covariance, and simple linear regression.

Upon completion of this week, students should be able to:

1. Reference the syllabus for information on expectations and requirements of the course.
2. Use the syllabus to find information on the structure of the course, including important dates and leaning objectives for each class.
3. Get to know their fellow classmates
4. Review correlations and simple regression analysis
   a. Differentiate between different types of correlations
   b. Understand the relationship between correlation and simple regression analysis

**Required Reading before class:**

Vittinghoff 3.2-3.3

**Assignments Due before class:**

1. Do the SOLSTICE survey
2. Reply to Welcome Message and Pre-class Questions on the course Canvas site

**Optional:**

If you need a quick refresher of Biostats I, read Vittinghoff 3.1
If you need a quick refresher on the use of R and RStudio, the DataLab has tutorials including: R and RStudio basics.
If you need a quick refresher on the use of Stata, watch STATA Video Tutorials below:

L01: Intro to Stata
L02: Results and Do-files
L03: Creating Variables and Labeling

**Session 2: Review multiple linear regression and model diagnostics**

**Learning Objectives:**

During class, we will revisit the uses of multiple regression, basic principles for reporting the results of multiple regression models, including regression diagnostics and discuss the remedies to avoid violations of linear regression
assumptions. During the computer lab, we will run bivariate analyses, a multiple linear regression, and model diagnostics.

Upon completion of this week, students should be able to:

1. Describe the assumptions and steps required to carry out linear regressions.
2. Run multiple linear regression models and interpret regression coefficients.
3. Understand how to evaluate regression models for outliers, influence and collinearity, and for violations of normality and homoscedasticity.
4. Understand how to interpret log-linear transformed models

**Required Reading/Assignments:**
1. Vittinghoff 4.0-4.3.4; 4.7
2. Homework 1 assigned

**Optional readings:**
- Stata users: Regression With Stata, chapter 2, Regression Diagnostics: [https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-stata-chapter2-regression-diagnostics/](https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-stata-chapter2-regression-diagnostics/)

**Session 3: Principles of model building – Part 1: Types of models; research questions and study designs; causal inference; confounding and effect modification**

**Learning Objectives:**
During class, we will think about WHY we are creating models, types of models (explanatory, descriptive, and predictive models) and their associated research questions, review study designs (intervention and observational study designs) and causal inference, and finally discuss techniques to deal with confounding and evaluate effect modifications in the regression models. During the computer lab, we will practice variable transformation and other modeling techniques to investigate non-linear relationships with the outcome variable (very common for nutrient-disease relationships) in multiple regression.

Upon completion of this week, students should be able to:

1. Gain a basic understanding of model types and their associated research questions
2. Understand the basic principles of confounding and effect modification
3. Understand how to investigate and interpret non-linear relationships with the outcome variable in multiple regression.

**Required Reading/Assignments:**
1. Vittinghoff 4.4; 4.6
2. Homework 2 assigned

**Assignments Due:**
1. Homework 1 is due before the lab session

**Session 4: Principles of model building – Part 2: Regression model building techniques**

**Learning Objectives:**
During class, we will describe and contrast different multiple regression model building techniques, including methods for selecting predictors (including using causal reasoning [Directed Acyclic Graph]) and running and refining models to avoid overfitting. During the computer lab, we will build multiple regression models using the variable selection techniques covered in class.
Upon completion of this week, students should be able to:
1. Recognize the differences in model building strategies for explanatory, descriptive, and predictive models.
2. Understand the methods for selecting predictors
3. Gain basic understanding of Directed Acyclic Graph and its use for model building
4. Generate and interpret statistics for comparing different regression models, including $R^2$, adjusted $R^2$, extra sum of square F-test, AIC, and BIC.
5. Discuss the drawbacks of predictor selection procedures.
6. Build models with multiple predictors

Required Reading/Assignments:
1. Vittinghoff 10.0-10.1.3.2, 10.1.4, 10.2-10.6

Optional Reading:

Assignments Due:
1. Homework 2 is due before the lab session

---

**Session 5: Introduction to logistic regression**

**Learning Objectives:**
During class, we will discuss logistic regression to model dichotomous outcomes and how to interpret the results of those models. We will also discuss the relationship of logistic regression models with odds ratios. During lab, we will fit simple and multiple logistic regression models.

Upon completion of this week, students should be able to:
1. Understand the basic principles of creating a multivariable model of dichotomous outcomes.
2. Understand how to interpret the results of logistic regression models in the context of odds ratio.
3. Use logistic regression to analyze binary outcomes using R, STATA, or SAS.

Required Reading/Assignments:
1. Vittinghoff 5.1-5.2

Special Note:
1. This week we will also introduce the mid-term project dataset and provide detailed instructions on the mid-term project.

**Session 6: Logistic regression model diagnostics**

**Learning Objectives:**
During class, we will discuss different diagnostic methods to assess logistic regression model fit, and conditional logistic regression for case-control studies (including nested case-control studies). We will also discuss the relationship of logistic regression models with 2x2 tables. During the computer lab, we will build and assess model fit of multiple logistic regression models.

Upon completion of this week, students should be able to:
1. Assess model fit and carry out model diagnostics for logistic regression models.
2. State the relationship between 2x2 tables and logistic regression models.
3. Fit conditional regression models for case-control studies.

Required Reading/Assignments:
1. Vittinghoff 5.3-5.4
2. Draft methods of the mid-term paper

**Session 7: Reporting Guidelines for Statistical Analysis**

**Learning objectives:**
During class, we will discuss reporting guidelines for different study designs. Students will present the draft methods of their mid-term paper to get feedback from the instructor and their peers.

Upon completion of this week, students should be able to:
- understand the basic principles of good reporting of scientific articles

**Required Reading/Assignments:**
3. Complete and turn in midterm assignment

**Session 8: Midterm Presentation and Peer Review**

**Learning Objectives:**
Students will submit their midterm project paper and conduct peer review.

**Required Reading/Assignments:**
1. Homework 3 assigned – analytical plan for final project

**Assignments Due:**
1. midterm project paper
2. peer review form

**Session 9: Analysis of Complex Surveys**

**Learning Objectives:**
During class, we will become familiar with different types of sampling design, complex survey weights, and how to incorporate sample weights in analysis to be able to produce adjusted sample statistics that are generalizable to the population of interest.

Upon completion of this week, students should be able to:
1. Be aware of different sampling designs and implications for the data
2. Use complex survey weights
3. Generate and interpret adjusted sample statistics

**Optional Readings:**
- Stata users: Analyzing Complex Survey Data: Some key issues to be aware of Richard Williams, University of Notre Dame, [https://www3.nd.edu/~rwilliam/stats2/SvyCautions.pdf](https://www3.nd.edu/~rwilliam/stats2/SvyCautions.pdf)
- R users: Complex sampling and R of Thomas Lumley, [https://r-survey.r-forge.r-project.org/survey/survey-wnar.pdf](https://r-survey.r-forge.r-project.org/survey/survey-wnar.pdf)

**Session 10: Poisson regression model building and diagnostics**

**Learning Objectives:**
During class, we will discuss Poisson regression to model count data and how to interpret the results of those models and discuss diagnostic methods to assess model fit. During lab, we will fit and diagnose simple and multiple Poisson regression models.

Upon completion of this week, students should be able to:
1. Describe situations in which Poisson regression analysis is needed.
2. Interpret the results of Poisson regression models.
3. Assess model fit and carry out model diagnostics for Poisson regression models.
4. Use Poisson regression to analyze count data using R, STATA, or SAS.

Required Reading/Assignments:
1. Campbell Ch 6, pg. 84-97 (scan on Canvas)

Assignments Due:
1. Homework 3 is due before the lab session

Session 11: Sample size and power
Learning Objectives:
During class, we will discuss power and sample size calculations for binary and measured outcomes. We will also introduce software, in addition to R, that can be used for power calculations. During lab, we will fit simple and multiple logistic regression models.

Upon completion of this week, students should be able to:
1. Recognize that power calculations are an essential part of both study design and analysis.
2. Identify the factors which influence power and sample size.
3. Compute sample sizes for binary and measured outcomes.
4. Discuss strategies to calculate sample sizes when not all information is available, and to address sample sizes that are not feasible.

Required Reading/Assignments:
1. Vittinghoff 4.8 & 5.7
2. Homework 4 assigned

Session 12: Visualizations
Learning Objectives:
The class will learn about tips for improving visualizations of statistical data.

Upon completion of this week, students should be able to:
1. Understand key concepts in data visualization
2. Identify common problems in data visualization
3. Identify best practices in data visualization
4. Critically evaluate graphical displays
5. Visualize data in R, SAS, or STATA

Required Reading/Assignments:
1. Preparation of your final e-poster due

Assignments Due:
1. Homework 4 is due before the lab session
Session 13: Survival Analysis

Learning objectives:
• understand the basic principles of modeling time-to-event outcomes
• be able to carry out and interpret survival models
• understand how to evaluate the validity of the assumptions of survival models

Required Reading/Assignments:
• Vittinghoff, chapters 6.0-6.2.6.2, 6.2.7-6.2.13

Assignments Due:
1. none

Session 14: Final e-poster presentation

Learning Objectives:
Students will present their final e-poster.

Required Reading/Assignments:
1. none

Assignments Due:
1. final individual e-poster presentation

* This schedule is subject to modification at the instructor’s discretion.
Appendix. Instructions and Grading Rubric for Final e-Poster

Parameters
- Use the BRFSS or NHANES data set
- Generate an original hypothesis to test using an adjusted multivariable model; you choose the outcome
- Build the appropriate adjusted model according to the model building principles you have learned in this course
- You should have at least three scientific references that either justify the importance of the hypothesis or justify the predictors selected for the model (no fishing!)
- Create a directed acyclic graph to illustrate the logic of your statistical model (include this model on your poster)
- Model assumptions should be demonstrated as tenable and any remediation steps clearly described

a. **Exploratory outputs**: Tests or visuals that were used to mainly assess assumptions on which statistical inference were based, support for the selection of analytical techniques, and investigate potential outliers or implausible entries.

b. **Descriptive outputs**: Tests or visuals that are included on the poster (e.g., the descriptive tables and results section); for example, the mean and SD of the continuous outcome, sample size, percent of missing, etc.

c. **Analytical outputs**: Tests or visuals that are included in the results section of the poster as the main results of analysis. This part should directly answer the research question.

d. **Diagnostic outputs**: Tests or visuals for checking the analytical results; e.g., residual plots, leverage plots, Cook’s distance plot, etc.

Part of your grade (22%, or 11/50 points) will be turning in well-documented syntax that includes the following clearly denoted sections:

**Preparation (Due Session 12)**
Familiarize yourself and begin to prepare the BRFSS or NHANES data set for your final project. This week, you will submit
- two possible hypotheses using the data in the BRFSS or NHANES data set (reference at least one article to justify each hypothesis),
- output of all cleaned and reformatted variables that they would need to test their hypotheses (this is the one time it’s ok just to paste the output tables, as I’m wanting to make sure there are no mistakes in the variables you’ll need to create your final models), and
- the syntax used to create the variables you used to create your models.

**Creation and Creativity (Weeks 13 and 14)**
There is nothing to turn in during these two weeks. I expect you to be working on the models and poster.

**e-Poster (due session 14)**
Please listen to these YouTube videos for ideas on how to best create your scientific poster:
- [https://www.youtube.com/watch?v=1RwJbhkCAS8](https://www.youtube.com/watch?v=1RwJbhkCAS8) and [https://www.youtube.com/watch?v=SYk29tnxASs](https://www.youtube.com/watch?v=SYk29tnxASs) (for BIG IDEA design)
- [https://www.youtube.com/watch?v=AwMFhyH7_5g](https://www.youtube.com/watch?v=AwMFhyH7_5g) (for a more traditional design)

You are free to choose your own look. Choose one of the templates from:
- [https://osf.io/6ua4k/](https://osf.io/6ua4k/) (for the BIG IDEA format)
• https://www.makesigns.com/SciPosters_Templates.aspx for more traditional poster design (but please do NOT choose the billboard poster template). Your poster may be horizontal or vertical.

The grading rubric for the e-poster is below.
Below are the criteria on which your final project scientific poster will be judged. Within each criteria category are a set of more specific criteria. For each criterion, you can get a maximum of 3 points (for Excellent) and a minimum of 0 points (for unacceptable). With 14 specific criteria, the maximum number of points is 39. As you can see, some sections (e.g., Technical) are more heavily weighted than others.

<table>
<thead>
<tr>
<th>Criteria Category</th>
<th>Excellent = 3 points</th>
<th>Good = 2 points</th>
<th>Adequate = 1 point</th>
<th>Unacceptable = no points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation of Research</td>
<td>• Prominently positions title/authors of paper</td>
<td>• Contains title/authors of paper</td>
<td>• Contains title/authors of paper</td>
<td>• Title/authors absent</td>
</tr>
<tr>
<td></td>
<td>• Thoroughly but concisely presents main points of introduction, hypotheses, research methods, results, and conclusions in a well-organized manner</td>
<td>• Adequately presents main points of introduction, hypotheses, research methods, results, and conclusions in a fairly well-organized manner</td>
<td>• Presents main points of introduction, hypotheses, research methods, results, and conclusions, but not as sufficiently and not as well-organized</td>
<td>• Does not sufficiently present main points of introduction, hypotheses, research methods, results, and conclusions, and is not well-organized</td>
</tr>
<tr>
<td>Visual Presentation</td>
<td>• Overall visually appealing, not cluttered, colors and patterns enhance readability</td>
<td>• Overall visually appealing, not cluttered, colors and patterns support readability</td>
<td>• Visual appeal is inadequate, somewhat cluttered, colors and patterns detract from readability</td>
<td>• Not visually appealing, cluttered, colors and patterns hinder readability</td>
</tr>
<tr>
<td></td>
<td>• Uses font sizes/variations which facilitate the organization, presentation, and readability of the research</td>
<td>• Adequate use of font sizes/variations to facilitate the organization, presentation, and readability of the research</td>
<td>• Use of font sizes/variations to facilitate the organization, presentation, and readability of the research is somewhat inconsistent or distracting</td>
<td>• Use of font sizes/variations to facilitate the organization, presentation, and readability of the research is inconsistent or distracting</td>
</tr>
<tr>
<td></td>
<td>• Graphics (e.g., tables, figures, etc.) are engaging and enhance the text</td>
<td>• Graphics (e.g., tables, figures, etc.) enhance</td>
<td>• Graphics (e.g., tables, figures, etc.) do not</td>
<td>• Graphics (e.g., tables, figures, etc.) do not</td>
</tr>
<tr>
<td>Documentation of Sources, Quality of Sources</td>
<td>understand order without narration</td>
<td>the text is arranged so that the viewer can understand order without narration</td>
<td>figures, etc.) adequately enhance text</td>
<td>enhance the text adequately and does not adequately assist the viewer in understanding order</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cites all data obtained from other sources; APA orAMA citation style is accurate</td>
<td>Cites most data obtained from other sources; APA or AMA citation style is accurate</td>
<td>Cites some data obtained from other sources; citation style is either inconsistent or incorrect</td>
<td>Does not cite sources</td>
<td></td>
</tr>
<tr>
<td>Spelling and Grammar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No spelling and grammar mistakes</td>
<td>Minimal spelling and grammar mistakes</td>
<td>Noticeable spelling and grammar mistakes</td>
<td>Excessive spelling and/or grammar mistakes</td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical model is appropriate for the hypotheses and data</td>
<td>Statistical model is appropriate for the hypotheses and data</td>
<td>Questionable decision about the choice of statistical model given the hypotheses or the data</td>
<td>Statistical model is not appropriate for the hypotheses and data</td>
<td></td>
</tr>
<tr>
<td>Model logic is represented with a directed acyclic graph</td>
<td>Model logic is not presented graphically</td>
<td>Model logic is not present</td>
<td>No logic model</td>
<td></td>
</tr>
<tr>
<td>Table(s) of outcomes are well organized and font size is appropriate</td>
<td>Minimal formatting issues with table(s)</td>
<td>Some formatting issues with table(s) or not formatted according to journal standards</td>
<td>Serious formatting issues with table(s)</td>
<td></td>
</tr>
<tr>
<td>Model results are correct with no flaws</td>
<td>Model results are correct with minimal flaws</td>
<td>Model results are mostly correct with minimal flaws</td>
<td>Model results are mostly correct with minimal flaws</td>
<td></td>
</tr>
<tr>
<td>Appropriate model assumptions are justified as tenable</td>
<td>Most model assumptions are violated or not justified as tenable</td>
<td>Some model assumptions are violated or not justified as tenable</td>
<td>Model assumptions are not justified</td>
<td></td>
</tr>
<tr>
<td>Graphical representation of results (if used) are clear</td>
<td>Graphical representation of results (if used) are difficult to read or interpret (e.g., not adequately labeled).</td>
<td>Graphical representation of results (if used) are uninterpretable</td>
<td>Graphical representation of results (if used) are uninterpretable</td>
<td></td>
</tr>
<tr>
<td>Extra Credit</td>
<td>+3 points: Graphical representation of results are clear, well-formatted, and clearly support the findings</td>
<td>+2 points: Graphical representation of results are clear, adequately formatted, and support the findings</td>
<td>+1 points: Graphical representation of results have some formatting problems and/or do not support the findings</td>
<td>+0 points: No graphs or figures</td>
</tr>
<tr>
<td>Comments</td>
<td>do not clearly support the findings</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>