NUTR-309: Statistical Methods for Nutrition Research II
Spring 2019

Class Meetings: Tuesdays 5:00-6:00pm, Jaharis 118

Lab Meetings: Thursdays 5:00-6:00pm, Sackler 514
Exceptions: 1/17 and 4/25: Jaharis 118

Instructor: Tania M. Alarcon Falconi, PhD
Email: tania.alarcon_falconi@tufts.edu
Office hours: To be determined on the first day of class.

Teaching Asst.: Heesun Eom
Email: heesun.eom@tufts.edu
Office hours: To be determined on the first day of class.

Semester Hour Units: 3.0

Prerequisites: Statistical Methods for Nutrition Research I (NUTR-206) or Statistical Methods for Nutrition Science and Policy (NUTR-207) or equivalent, and graduate standing or instructor consent. Ability to conduct exploratory data analysis using R. NOTE: Students cannot receive semester hour units for both NUTR 309 and NUTR 307: Regression Analysis for Nutrition Policy.

Course Description: This course is part two of a one-year, two-semester course on statistical methods for nutrition research. The focus of this course is on simple and multiple regression methods for continuous, binary, and survival data. Emphasis is on developing a conceptual understanding of the application of these techniques to solving problems rather than on numerical details. In the computer lab sessions, students will use concepts learned during lecture to analyze data using statistical software R and RStudio, an integrated development environment for R.

Course Objectives: Students will learn to analyze quantitative nutrition data using simple and multiple regression models. Upon successful completion of the course, students should be able to:
- Analyze continuous outcomes using multiple linear regression models
- Investigate and test associations with binary outcomes using logistic regression
- Apply survival modeling methods for censored and time-to-event analyses
- Apply repeated measures methods for longitudinal data
- Run simple and multiple regressions in RStudio
- Document data analysis using data management plans and R scripts that are liberally commented


Class Materials: All class materials, including lecture notes and assignments, will be posted on Canvas.

Statistical Software: 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.
**Communication Policy:**
For questions on the course, we will use Piazza – a link is available on the course Canvas site. Students should try to seek out information for themselves before contacting the instructor or the TA. If you cannot find your answer, ask a question on Piazza. Students are strongly encouraged to answer others’ questions or take part in any discussion. We will discuss guidelines and expectations for Piazza use during the first day of class.

**Classroom Conduct:**
Students are expected to attend all classes and all 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.

Students who use computers in class should limit the use solely for purposes related to the course. Cell phone use is not allowed in class.

**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 and Tufts University policies (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.

**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.

**Snow emergency:**
If the campus closes down due to snow, announcements will be sent via e-mail. You can also check the Tufts Emergency Preparedness website for updates: https://emergency.tufts.edu/weather/closing/
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>5%</td>
</tr>
<tr>
<td>Lab participation</td>
<td>15%</td>
</tr>
<tr>
<td>Homework (5 x 6%)</td>
<td>30%</td>
</tr>
<tr>
<td>Exams (2 x 15%)</td>
<td>30%</td>
</tr>
<tr>
<td>Final project</td>
<td>20%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Final course grades will be based on the following (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 (5%)**

Students are expected to attend all classes and read all assigned materials before class. Students will be asked to post a question, comment, or response to another student’s post in Piazza for the required reading material. Posts will be due by 11:59 pm the day before class. Late submissions will not receive any credit.

**Lab participation (lab exercises 10%, quizzes 5%)**

Students are expected to attend all labs and complete all lab exercises. Throughout the semester, there will be short quizzes during lab to assess students’ continued understanding of the course material. Missing a lab without prior agreement with the instructor will result in receiving a 0% for that lab exercise and if applicable, quiz.

**Homework (30%)**

Five problems sets, each worth 6% 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 at 5:00 PM on the specified date. 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. 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.

**Exams (30%)**

There will be two in-class exams, each worth 15% of the final grade. Additional information on the format, grading and content of the exams will be distributed prior to each exam. Any scheduling conflicts should be discussed with the instructor as soon as possible.

**Final project (20%)**

Students will complete an individual data analysis project using a dataset provided by the instructor. Additional details on the project and format will be provided during Week 11. The final written project will be due on Thursday, 5/9/2019 at 5:00 pm. Late submissions will not be accepted unless an extension is approved in advance.
Course Topics and Assignment Schedule at a Glance*:

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments (due date)</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thu, 1/17</td>
<td>Course overview and review of Biostatistics I (Jaharis 118)</td>
<td></td>
<td>Vittinghoff 3.1-3.4 (optional)</td>
</tr>
<tr>
<td>2</td>
<td>Tue, 1/22</td>
<td>Linear regression and controlling for confounding</td>
<td>Piazza post (1/21)</td>
<td>Vittinghoff 4.1-4.4</td>
</tr>
<tr>
<td></td>
<td>Thu, 1/24</td>
<td>Computer Lab 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tue, 1/29</td>
<td>Interaction and collinearity in multiple linear regression</td>
<td>Piazza post (1/28)</td>
<td>Vittinghoff 4.6</td>
</tr>
<tr>
<td></td>
<td>Thu, 1/31</td>
<td>Computer Lab 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tue, 2/5</td>
<td>Regression model building techniques</td>
<td>Piazza post (2/4)</td>
<td>Vittinghoff 10.1-10.5</td>
</tr>
<tr>
<td></td>
<td>Thu, 2/7</td>
<td>Computer Lab 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Tue, 2/12</td>
<td>Sample size and power</td>
<td>Piazza post (2/11)</td>
<td>Vittinghoff 4.8, 5.7</td>
</tr>
<tr>
<td></td>
<td>Thu, 2/14</td>
<td>Computer Lab 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tue, 2/19</td>
<td><strong>EXAM 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thu, 2/21</td>
<td>No lab. Substitute Monday’s class schedule on Thursday.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Tue, 2/26</td>
<td>Introduction to logistic regression</td>
<td>Piazza post (2/25)</td>
<td>Vittinghoff 5.1-5.2</td>
</tr>
<tr>
<td></td>
<td>Thu, 2/28</td>
<td>Computer Lab 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tue, 3/5</td>
<td>Conditional logistic regression and model diagnostics</td>
<td>Piazza post (3/4)</td>
<td>Vittinghoff 5.3-5.4</td>
</tr>
<tr>
<td></td>
<td>Thu, 3/7</td>
<td>Computer Lab 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Tue, 3/12</td>
<td>Introduction to survival analysis</td>
<td>Piazza post (3/11)</td>
<td>Vittinghoff 3.5</td>
</tr>
<tr>
<td></td>
<td>Thu, 3/14</td>
<td>Computer Lab 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SPRING BREAK - Saturday, March 16, 2019 to Sunday, March 24, 2019**

| 10      | Tue, 3/26  | Regression techniques for survival analysis                         | Piazza post (3/25)              | Vittinghoff 6.1-6.2          |
|         | Thu, 3/28  | Computer Lab 8                                                       | Final project assigned          |                              |
| 11      | Tue, 4/2   | Additional considerations for survival analysis                     | Piazza post (4/1)               | Vittinghoff 6.3-6.4          |
|         | Thu, 4/4   | Computer Lab 9                                                       |                                 |                              |
| 12      | Tue, 4/9   | Introduction to longitudinal analysis                                | Piazza post (4/8)               | Vittinghoff 7.1-7.2          |
|         | Thu, 4/11  | Computer Lab 10                                                      | Homework 5 assigned             |                              |
| 13      | Tue, 4/16  | Introduction to longitudinal data analysis                          | Piazza post (4/15)              | Vittinghoff 7.3-7.4          |
|         | Thu, 4/18  | Computer Lab 11                                                      | Homework 5 (4/20)               |                              |
| 14      | Tue, 4/23  | Statistical methods in practice                                     |                                 |                              |
|         | Thu, 4/25  | **EXAM 2 (Jaharis 118)**                                            |                                 |                              |
| 15      | Tue, 5/7   | Computer lab working session                                         |                                 |                              |
|         | Thu, 5/9   | **FINAL PROJECT DUE**                                               |                                 |                              |

*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 review of Biostatistics I

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, exams, and final project, as well as on the expectations for using R and RStudio, Canvas, and Piazza. During the second part of the class, we will revisit the basics of Biostatistics I with examples in RStudio.

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 learning objectives for each class.
3. Locate and interpret important statistics (e.g., p-values, t-statistics,) from printouts of statistical tests in RStudio.

Required Reading/Assignments:
No required readings this week.

Assignments Due:
No assignments due this week.

Optional:
If you need a quick refresher of Biostats I, read Vittinghoff 3.1-3.4
If you need a quick refresher on the use of R and RStudio, the DataLab has tutorials including: R and RStudio basics.

Session 2: Linear regression and controlling for confounding

Learning Objectives:
During class, we will revisit the basics of linear regression, including assumptions, and discuss the concept of confounding and how to use different approaches to minimize its impact. During the computer lab, we will provide a quick refresher on the use of RStudio; and will fit regression models, check model assumptions, and test for confounding.

Upon completion of this week, students should be able to:
1. Describe the assumptions and steps required to carry out linear regressions.
2. Run simple and multiple linear regression models and interpret regression coefficients and model diagnostic plots.
3. Examine and test confounding in regression models.
4. Suggest study designs and analytical remedies to address confounding.

Required Reading/Assignments:
1. Vittinghoff 4.1-4.5
2. Homework 1 assigned

Assignments Due:
1. **Monday, 1/21**: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.
Session 3: Interaction and collinearity in multiple linear regression

Learning Objectives:
During class, we will discuss the consequences of interaction and how to model them. We will also discuss collinearity among independent variables. During the computer lab, we will fit regression models with interactions and test for collinearity.

Upon completion of this week, students should be able to:
1. Recognize the impact of unadjusted interaction in a regression model.
2. Specify and run a hypothesis test for interaction.
3. Interpret regression models with statistically significant interaction terms.
4. Produce and interpret VIF and Tolerance statistics.

Required Reading/Assignments:
1. Vittinghoff 4.6

Assignments Due:
1. **Monday, 1/28**: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm
2. **Saturday, 2/2**: Homework 1 due by 5:00 pm.

Session 4: Regression model building techniques

Learning Objectives:
During class, we will describe and contrast different multiple regression model building techniques, including automatic model selections and use of model-based statistics. 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. Generate and interpret statistics for comparing different regression models, including $R^2$, adjusted $R^2$, extra sum of square F-test, AIC, and BIC.
3. Discuss the drawbacks of predictor selection procedures.
4. Build models with multiple predictors in RStudio.

Required Reading/Assignments:
1. Vittinghoff 10.1-10.5
2. Homework 2 assigned

Assignments Due:
1. **Monday, 2/4**: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.
Session 5: 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

**Assignments Due:**
1. **Monday, 2/11:** Post a question, comment, or response to another student's post in Piazza for the required reading material by 11:59 pm
2. **Saturday, 2/16:** Homework 2 due by 5:00 pm.

---

Session 6: Exam 1 (no class on 2/21)

**Learning Objectives:**
Exam 1 will evaluate students' understanding of the material covered up to session 5.

**Required Reading/Assignments:**
No required readings this week.

**Assignments Due:**
No assignments due this week.
Session 7: 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 with and without interaction terms.

Upon completion of this week, students should be able to:
1. Describe situations in which logistic regression analysis is needed.
2. Interpret the results of logistic regression models in the context of odds ratio.
3. Include and interpret interaction terms in logistic regression models.
4. Use logistic regression to analyze binary outcomes with and without interaction terms using RStudio.

Required Reading/Assignments:
1. Vittinghoff 5.1-5.2
2. Homework 3 assigned

Assignments Due:
1. Monday, 2/25: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm

Session 8: Conditional logistic regression and 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. 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. Assess model prediction accuracy with sensitivity, specificity, and ROC curves.
3. State the relationship between 2x2 tables and logistic regression models.
4. Fit conditional regression models for case-control studies.

Required Reading/Assignments:
1. Vittinghoff 5.3-5.4

Assignments Due:
1. Monday, 3/4: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.
2. Saturday, 3/9: Homework 3 due by 5:00 pm.
Session 9: Introduction to survival analysis

Learning Objectives:
During class, we will introduce survival data, which take into account the time until an event occurs, and will discuss tests for group comparison. During lab, we will create Kaplan-Meier survival curves and compare two survival curves.

Upon completion of this week, students should be able to:
1. Describe time to event data.
2. Interpret Kaplan-Meier survival curves.
3. Calculate median survival from an estimated survival curve.
4. Compare groups with a time to event outcome.

Required Reading/Assignments:
1. Vittinghoff 3.5
2. Homework 4 assigned

Assignments Due:
2. **Monday, 3/11**: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm

SPRING BREAK - Saturday, March 16, 2019 to Sunday, March 24, 2019
Session 10: Regression techniques for survival analysis

Learning Objectives:
During class, we will discuss regression techniques for analysis of survival data, taking into account confounding. We will emphasize description and use of the Cox proportional hazards model. During the computer lab, we will model survival data in RStudio.

Upon completion of this week, students should be able to:
1. Describe the relationship between the survival function and the hazard function.
2. Recognize the benefits of using the semi-parametric Cox proportional hazards model instead of parametric models.
3. Control for confounding using a Cox proportional hazards model.
4. Interpret the coefficients from a Cox proportional hazards model.

Required Reading/Assignments:
1. Vittinghoff 6.1-6.2

Assignments Due:
1. Monday, 3/25: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.

Session 11: Additional considerations for survival analysis

Learning Objectives:
During class, we will introduce additional considerations for modeling survival data using a Cox proportional hazards model, including checking for model fit and incorporating time-dependent covariates. During lab, we will continue to fit Cox proportional hazards models and will assess model fit.

Upon completion of this week, students should be able to:
1. Fit and interpret a Cox proportional hazards model with interaction terms.
2. Incorporate time varying covariates in a Cox proportional hazards model.
3. Assess the proportional hazards assumption.
4. Fit a stratified Cox proportional hazards model.

Required Reading/Assignments:
1. Vittinghoff 6.3-6.4

Assignments Due:
1. Monday, 4/1: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm
2. Saturday, 4/6: Homework 4 due by 5:00 pm.
Session 12: Introduction to longitudinal data

Learning Objectives:
During class, we will discuss important features of longitudinal that complicate their analysis and will present various examples of different types of longitudinal studies. During the computer lab, we will discuss the basics of longitudinal data, including their structure, and how to create different plots in RStudio.

Upon completion of this week, students should be able to:
1. Describe the main objectives of longitudinal studies.
2. Recognize the features of longitudinal data that complicate their analysis.
3. Explain the consequences of ignoring those specific features.
4. Plot longitudinal data in RStudio.

Required Reading/Assignments:
1. Vittinghoff 7.1-7.2
2. Homework 5 assigned

Assignments Due:
1. Monday, 4/8: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.

Session 13: Introduction to longitudinal data analysis

Learning Objectives:
During class, we will introduce analysis techniques for longitudinal data, and discuss some benefits and limitations of those methods. During the computer lab, we will analyze some longitudinal datasets.

Upon completion of this week, students should be able to:
1. List common methods that can be used to analyze longitudinal data.
2. Describe the benefits and limitations of the “Analysis of Response Profiles” technique.
3. Use “Analysis of Response Profiles” to analyze longitudinal data in RStudio.

Required Reading/Assignments:
1. Vittinghoff 7.3-7.4

Assignments Due:
1. Monday, 4/15: Post a question, comment, or response to another student’s post in Piazza for the required reading material by 11:59 pm.
2. Saturday, 4/20: Homework 5 due by 5:00 pm.
Session 14: Statistical methods in practice and Exam 2 on 4/25

Learning Objectives:
On Tuesday, 4/23, we will discuss research papers that show how all the statistical methods explored in this course are used in practice. We will also use mistake finding exercises as a self-assessment tool to help students prepare for Exam 2. There will be no lab on Thursday, 4/25, instead, we will have Exam 2, which will evaluate students’ understanding of the material covered up to session 13.

Upon completion of this week, students should be able to:
1. Assess the quality of statistical analysis in published media.
2. Propose alternative analyses when a specific statistical analysis is not appropriate.

Required Reading/Assignments:
No required readings this week.

Assignments Due:
No assignments due this week.

Session 15: Computer lab working session and Final Project due on 5/9

Learning Objectives:
During this week, we will have a computer lab working session to help students with the final project.

Required Reading/Assignments:
No required readings this week.

Assignments Due:
1. Thursday, 5/9: Final project due by 5:00 pm.

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