Overview and Objectives:
This course is designed for researchers in the clinical and translational sciences who want to understand and use linear regression techniques. Topics covered include multiple linear regression, regression diagnostics, analysis of variance, analysis of covariance, confounding, mediation, moderation, interactions, and model selection. At the completion of the course, trainees should be able to demonstrate an understanding of the basic principles of ANOVA and linear regressions, to assess their adequacy and assumptions, to analyze simple data sets taken from the fields of medicine and public health, to summarize results from regression models via written communication, and to recognize situations that require more advanced methodologies.

Responsibilities:

- There will be reading assignments in the textbook. All readings are expected to be completed before class. You should anticipate quizzes and class discussions about the material you are expected to read.
- Students will be assigned three homework assignments that will be graded. These assignments will be used to reinforce material reviewed in class. All homework assignments will be assigned a due date. You are encouraged to discuss and work together on problems, but you must write up your results individually, i.e. very similar papers will not be accepted. Homework assignments are to be turned in at the beginning of class on the due date. No assignment will be accepted via email. Late homework assignments will be penalized 10% per day past the due date (unless prior arrangements have been made with the instructor) until the following lecture, at which no late homeworks will be accepted.
- There will be a quiz on the 2nd to last day of class consisting of short answer and multiple choice questions. The focus of this exam will be on understanding 1) what to do and when to do it, 2) how to interpret what you did, and 3) more focused questions on the important issues. We will discuss the results on the last day of class.
- There will be a final project due two weeks from the last day of class. This project will be a compilation of the homework assignments but with synthesized text that pulls the analyses together in a coherent, scientific manner. This project must be turned in on time to receive full credit, and will be subject to the same late penalties as the homework (outlined above). Students must work on the final project independently. Discussion, consulting, or working with other people is not permitted.

Course Requirements:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation and attendance</td>
<td>0%</td>
</tr>
<tr>
<td>Homework assignments</td>
<td>55%</td>
</tr>
<tr>
<td>Final project</td>
<td>30%</td>
</tr>
<tr>
<td>Quiz</td>
<td>15%</td>
</tr>
</tbody>
</table>
**Attendance Policy:**

- Students are expected to sign-in to each class (computer provided in suite lobby). If a problem is encountered with the sign-in system, please contact the course instructor(s) as well as Lauren Talotta (talottals@upmc.edu) immediately.

**Course Grading Scale:**

For the computation of the final course grade as well as for the course assignments, the following grading scale will be used:

- 90 - 100 = A
- 80 - 85 = B
- 70 - 75 = C
- 60 - 65 = D
- 86 - 89 = B+
- 76 - 79 = C+
- 66 - 69 = D+
- < 60 = F

**Required Textbook(s):**


**Supplemental Textbook(s):**


Books are available at the ICRE library in Parkvale 306 (Holliman office).

**Website resources:**

All homework assignments, course information, and communication will be available at http://courseweb.pitt.edu.

**Academic Integrity:**

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity (http://www.provost.pitt.edu/info/ai1.html). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.
**Course Schedule**

### Session 1: Correlation and Simple Linear Regression

**At the conclusion of this lecture, the student will be able to:**

Describe correlation and the linear regression model and explain its appropriate use answering biomedical research questions.

**Topics:**

1. Course description
2. Review of correlation
3. Review of simple linear regression (model, assumptions, estimation, inference, prediction)
4. Prediction and confidence bands
5. Introduction to multiple linear regression

**Required Reading(s):**

Reading: Vittinghoff Ch 1 (Introduction), Ch 2 (Explor & Descr) and Ch 3 pgs 33-42 (Correlation & SLR)
Supplemental reading: Kleinbaum Ch 4 -7 (Intro Regression, SLR, Correlation, ANOVA table)

**Homework assignment(s):**

Homework 1: Correlation, simple linear regression, and multiple linear regression

**Competencies**

1. Research Design: Problem Formation
2. Data Analysis: Applied Analytical Techniques
3. Professional Skills: Written Communication

### Session 2: Multiple Linear Regression

**At the conclusion of this lecture, the student will be able to:**

Confidently use the linear regression model with multiple explanatory variables to discover important relationships between variables.

**Topics:**

1. Multiple (multivariable) regression model
2. Estimation
3. Evaluating model and predictor significance
4. Dummy (indicator) variables
5. Introduction to confounding, mediation, and interaction

**Required Reading(s):**
Session 3: Confounding, Mediation and Interaction

At the conclusion of this lecture, the student will be able to:

Identify and appropriately address confounding, mediation, moderation, and interaction.

Topics:

1. Covariates
2. Confounding
3. Interaction and testing for equal slopes
4. Mediation vs moderation
5. Introduction to diagnostics

Required Reading(s):

Reading: Vittinghoff Chapter 4 89-108 (Confounding, Mediation, and Interaction)

Supplemental reading: Kleinbaum Chapters 11 and 12 (Confounding and Interaction)

Barrera M, et al Social support and social-ecological resources as mediators of lifestyle intervention effects for Type 2 Diabetes. *Journal of Health Psychology* 2006;11;483-495.

Competencies

1. Research Design: Sampling
2. Data Analysis: Applied Analytical Techniques
3. Professional Skills: Written Communication

Homework assignment(s):

Homework 2: Confounding, interaction, and regression diagnostics

Due Today:

Homework 1: Correlation, simple linear regression, and multiple linear regression

Session 4: Regression diagnostics

At the conclusion of this lecture, the student will be able to:

Use regression diagnostic tools to assess validity of assumptions for a linear regression model.

Topics:

1. Simple approaches to diagnosing problems
2. Residual analysis and types of residuals
3. Detecting outliers (leverages, jackknife residuals, Cook’s d)
4. Assessing model assumptions
5. Alternate strategies (if assumptions are not met)
6. Transformations
7. Summary of how to validate assumptions
8. Collinearity

**Required Reading(s):**

Vittinghoff Chapter 4 109-130 (Model Assumptions & Fit) and Ch 10 421-423 (Collinearity & number of predictors)

**Competencies**

1. Data Analysis: Applied Analytical Techniques
2. Professional Skills: Ethics and Professional Norms

**Session 5: Model Selection**

**At the conclusion of this lecture, the student will be able to:**

Correctly select the best linear regression model to fit exploratory data.

**Topics:**

1. Model selection
2. Best prediction of Y
3. Steps in selecting the best regression equation
4. Criteria for selecting a model
5. Backward, forward, and stepwise procedures
6. Evaluating a primary predictor and multiple important predictors
7. Introduction to hierarchical regression

**Required Reading(s):**

Reading: Vittinghoff Chapter 10 (Predictor Selection) and Ch 8 257 & 267 (Hierarchical Data)
Supplemental reading: Kleinbaum Ch 14 and 16 (Diagnostics and Selecting the Best Regression Equation)

**Homework assignment(s):**
Homework 3: Regression diagnostics, model selection

**Due Today:**
Homework 2: Confounding, interaction, and regression diagnostics

**Competencies**

1. Research Design: Problem Formation
2. Data Analysis: Applied Analytical Techniques
3. Professional Skills: Written Communication

**Session 6 : One- & Two-Way ANOVA, Analysis of Covariance & Review**
At the conclusion of this lecture, the student will be able to:

Describe the ANOVA model and explain when it should be used compared to the linear regression model.

Topics:

1. ANOVA vs. Dummy Regression
2. Reference cell coding
3. Multiple comparisons
4. Two-way ANOVA
5. ANCOVA and linear regression
6. Adjusted means example
7. Review of class topics

Required Reading(s):
Reading: Vittinghoff Chapter 3 pgs 28-32 (t-test and ANOVA)
Supplemental reading: Kleinbaum Ch 13 and Ch 17 (One-way ANOVA and ANCOVA)

Competencies

1. Research Design: Problem Formation
2. Data Analysis: Applied Analytical Techniques

Homework assignment(s):
Homework 4: ANOVA

Due Today: Homework 3: Regression diagnostics and model selection

Session 7: Quiz, Sample Size, Correlated Data

At the conclusion of this lecture, the student will be able to:

Determine the meaningful sample size and power for a given biomedical research question.
Recognize studies that have or require correlated data.

Topics:

1. Two-way ANOVA, ANCOVA (if needed from previous lecture)
2. Sample size and power calculations
3. Correlation data and structures

Required Reading(s):
Vittinghoff Chapter 7 Repeated Measures Analysis pgs 261-271

Supplemental reading: Kleinbaum Ch 25-26 (Analysis of Correlated Data) and Sections 27.1-27.3
Note about reading: some of the reading in Ch’s 25-26 is statistically dense; the examples will be most helpful but feel free to use the lecture as your primary source for information

Competencies

1. Research Design: Sampling
2. Research Design: Measurement
3. Data Analysis: Data Management and Biomedical Informatics
Due Today: Homework 4: ANOVA

Session 8: Review of quiz, Assignment of project & Missing data

At the conclusion of this lecture, the student will be able to:

Explain how and when to use the linear regression and ANOVA models. Explain different types of missing data mechanisms.

Topics:

1. Review of regression and ANOVA
2. Missing data

Suggested Reading(s):

Competencies

1. Research Design: Problem Formation
2. Research Design: Sampling
3. Research Design: Measurement
4. Data Analysis: Data Management and Biomedical Informatics
5. Data Analysis: Applied Analytical Techniques
6. Professional Skills: Written Communication