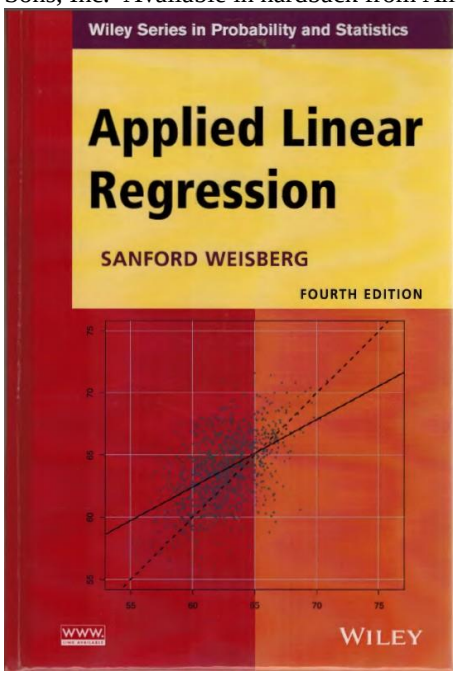
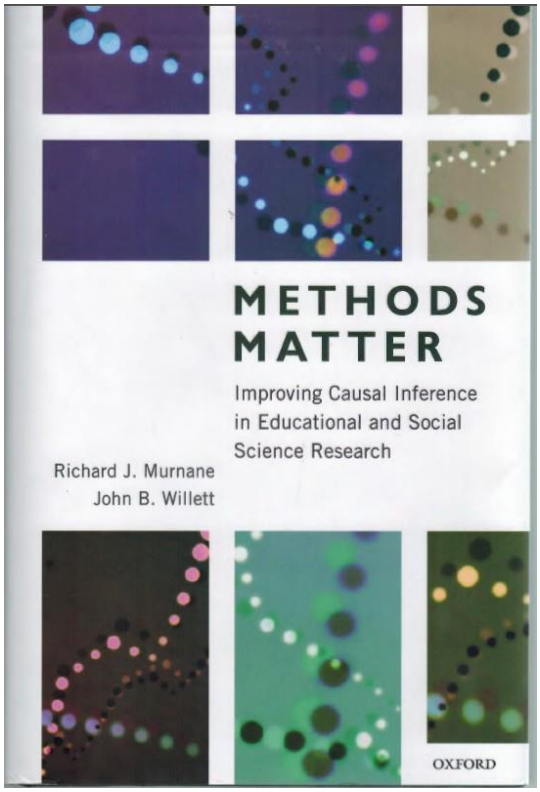
Psychology 313 – Correlation and Regression (Graduate)

Instructor:	James H. Steiger, Professor Email: *james.h.steiger@vanderbilt.edu* Department of Psychology and Human Development Office: Hobbs 215A Phone: 615-322-7060 (On campus 2-7060) Office Hours: TTh 11:10-12:00; W (By Appointment)
Course Meeting Times:	T R 0935–1050 Hobbs 107
Instructor Website:	http://www.statpower.net All course materials, including a detailed schedule of readings and class assignment by day, are available for downloading at this website.
Course Objectives	The objective of Psychology 313 is to provide students with a firm background in modern correlation and regression techniques. More specifically, the student will attain full competence in • Linear regression • Polynomial regression • Interaction effects • Multiple Regression • Univariate and Multivariate Outlier Detection • Data Transformation Algorithms • Handling of Missing Data • Non-linear Regression • Matrix Algebra for Statistical Applications • Logistic Regression • Poisson Regression • Zero-inflated Models • Regression Diagnostics and Graphics • Variable Selection Procedures • Instrumental Variables Analysis • Propensity Score Analysis • Regression Discontinuity Designs • Power and Sample Size Estimation • Regression Component Analysis

Required Textbooks:	Weisberg, Sanford (2014). *Applied linear regression (4th Ed.)* Hoboken, NJ: John Wiley & Sons, Inc. Available in hardback from Amazon.  Murnane, Richard J., & Willett, John B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York, NY: Oxford University Press.  It is essential that you read the assigned reading prior to class, as discussion will center on the readings. Understanding will be greatly enhanced by adhering to this policy (as will your grades!). Occasionally the text will be supplemented by other readings provided by the instructor.

Software:	The freeware statistical program R will be used throughtout the course, in the classroom, in laboratory exercises, on homework assignments. All work is open-book, open note, and students are allowed to use R, and any statistical functions they have written for R, during exams.
Prerequisites:	There are no official prerequisites for the course, but an introductory course in applied statistics at the graduate level is strongly recommended.

<table>
<tr><td>Grading:</td><td>Homework 85% There are homework assignments throughout the course. These are deep and time-consuming, and go beyond the kind of superficial coverage we can achieve in an examination format. These exercises develop extensive competence in R programming to generate both statistical and graphical analysis.

Class Presentation 15%

Grading Standards

<table>
<tr><td>90–100</td><td>A</td><td>excellent</td></tr>
<tr><td>86–89</td><td>A–</td><td>superior</td></tr>
<tr><td>83–85</td><td>B+</td><td>strongly competent</td></tr>
<tr><td>80–82</td><td>B</td><td>competent</td></tr>
<tr><td>76–80</td><td>B–</td><td>competent</td></tr>
<tr><td>70–75</td><td>C+</td><td>adequate</td></tr>
<tr><td>66–70</td><td>C</td><td>adequate</td></tr>
<tr><td>61–65</td><td>C–</td><td>adequate</td></tr>
<tr><td>57–60</td><td>D+</td><td>marginally adequate</td></tr>
<tr><td>54–56</td><td>D</td><td>inadequate</td></tr>
<tr><td>50–53</td><td>D–</td><td>barely passing</td></tr>
<tr><td>00–49</td><td>F</td><td>failing</td></tr>
</table>
</td></tr>
<tr><td>Analysis Project and Class Presentation:</td><td>Time permitting, several class meetings (depending on enrollment) during the final weeks will be reserved for original conference-style presentations by each student.</td></tr>
<tr><td>Honor Code:</td><td>Your presence here presupposes a commitment to principles of academic honesty, integrity, and responsible citizenship. Consult the University Student Handbook regarding academic misconduct. You are encouraged to work together on computer code, but interpretations and write-ups must be your own. Continued enrollment in this course assumes tacit agreement with this policy.

The Honor Code is defined in Vanderbilt University Student Handbook:

http://studentorgs.vanderbilt.edu/HonorCouncil/

Please abide by the Honor Code. Academic misconduct will be dealt with through official channels. I encourage you to study and discuss topics with your fellow students, but all submitted work should reflect only your own knowledge and understanding of the subject matter, not your fellow students, unless a team project is explicitly authorized.
If there is any doubt whatsoever about what actions constitute infractions of the honor code, please discuss the issue with the course instructor.</td></tr>
<tr><td>Disabilities:</td><td>If you are (or become) learning, sensory, or physically disabled, and feel that you need special assistance in lecture, reading, testing, or any other work in this course, please contact me to discuss your specific needs as soon as possible.</td></tr>
</table>

How to Succeed in This Course	
Attend Lecture Regularly:	Attendance at all lectures is strongly encouraged. This course relies heavily on cumulative information, and your success will depend heavily on your ability to keep up.
Use Office Hours:	I cannot recommend strongly enough that you use office hours. This is a way for you to get one-on-one clarification of any questions or problems that you might have.
Work with Others:	Study groups can be very useful. I encourage you to work together to master difficult concepts and use nonparametric statistics software.
Ask Questions:	If you have a question about something, I guarantee that at least one other person has the same question. Ask questions in class. Ask questions in office hours. Ask me over e-mail. Use the resources that are available to you.
Be Inclusive and Supportive:	This introductory course welcomes students with a rich variety of backgrounds. Over the years, we have found that students who arrive with a superior skill set profit immensely by sharing their knowledge. The act of sharing and teaching others helps reinforce your own knowledge, while alerting you to new ways of thinking about the subject matter. Students who start the course with less than the typical background generally find that their experience is enhanced by networking with other students. Let's celebrate our differences, and support each other with an atmosphere of mutual respect and inclusiveness.

Outline of Topics by Week[1]	
Week 1	Administrative Issues Review of the Basic Algebra of Linear Combinations Review of Bivariate Correlation and Regression Sample Applications of Linear Regression and Correlation
Week 2	Tools for Analyzing Scatter Plots – Resizing, Smoothers, Transformations Scatterplot Matrices Linear Regression as a Model The Conditional Mean Function The Conditional Variance Function
Week 3	A Minimal Introduction to Matrix Algebra for Correlation and Regression Matrix Algebra in R
Week 4	Linear Regression as a Predictive System Fitted Values Residuals Estimating σ^2 The Analysis of Variance and *F* Test The Coefficient of Determination R^2
Week 5	Review of the Algebra of Expected Values, Variances, Covariances, Linear Combinations Matrix Expected Value Algebra Multiple Regression Adding a Predictor to a Simple Linear Regression Added Variable Plots Polynomials, Interactions, and Factors The Multiple Regression Surface Matrix Algebra of Multiple Regression The Geometry of Multiple Regression The Analysis of Variance in Multiple Regression
Week 6	Drawing Conclusions from a Regression Analysis Understanding Parameter Estimates Rate of Change, Signs of Estimates Interpretational Dependencies between Coefficients Rank Deficient and Over-Parameterized Mean Functions Experimentation Versus Observation Sampling from a Normal Population Missing Data Missing at Random Computationally Intensive Methods Regression Inference without Normality Nonlinear Functions of Parameters Predictors Measured with Error
Week 7	Weighted Least Squares Regression Testing for Lack of Fit, Variance Known Testing for Lack of Fit, Variance Unknown General F Testing Non-Null Distributions and Power Joint Confidence Regions

[1] The pace of a course is difficult to predict. This schedule is subject to change.

Outline of Topics by Week (ctd.)	
Week 8	Polynomial Regression Factors, Adding a Predictor: Comparing Regression Lines Many Factors Partial One-Dimensional Mean Functions, Random Coefficient Models,
Week 9	Transformations and Scatterplots Power Transformations Transforming Only the Predictor Variable Transforming the Response Only The Box and Cox Method Transformations and Scatterplot Matrices Automatic Choice of Transformation of Predictors Transformations of Nonpositive Variables
Week 10	Regression Diagnostics Residuals Testing for Curvature Testing for Nonconstant Graphs for Model Assessment Checking the Putative Mean and Variance Functions
Week 11	Outliers and Influence Cook's Distance Other Measures
Week 12	Variable Selection Forward Selection Backward Selection Stepwise Selection Information Criteria
Week 13	Nonlinear Regression Estimation Techniques Bootstrap Inference
Week 14	Thanksgiving Break
Week 15	Logistic Regression The Binomial Regression Model Deviance Generalized Linear Models Constrained Linear Regression Linear Regression via Structural Equation Modeling
Week 16	Poisson and Zero-Inflated Regression Special Models for Count Data Modeling Overdispersion: the quasi-Poisson and negative binomial approaches
Week 17	Student Presentations