Statistics 2118
Regression Analysis
Fall 2014

11:10-12:25 MW  MPA  B07

Instructor:  Dr. Efstatia Bura
Office:  801 22nd Street NW, Rm 554 (Rome Hall)
Office Phone:  (202) 994-6358
Office Hours:  M: 2:30 p.m. to 3:30 p.m., W: 3:30-4:30, or by appointment.
Email:  ebur@gwu.edu  (Preferred means of communication)

TAs:  Jesse Jeter and Josh Touyz

Email:  jessejeter@email.gwu.edu
        joshuahtouyz@email.gwu.edu

TA Office Hrs:  Jesse Jeter:  1-2 and 3:30-4:30 on Tuesdays
               Josh Touyz:  4:50 on Tuesdays, 10:00-11:00 on Fridays
               in Old Main room 301A (1922 F Street)

A one semester introductory course in regression analysis. It covers the basic ideas of
Applied Regression Analysis including: Simple Linear Regression; Diagnostics for
Simple Linear Regression; Multiple Regression; Parameter Estimation and Testing;
Indicator Variables; Diagnostics; ANOVA; Transformations and Weighted Least Squares;
Model Selection; Logistic regression. Any changes will be announced in class.

PREREQUISITES
Knowledge of the basic concepts of probability and statistics (Stat 51, 53, or 111 and 157, 158).

TEXTS
Required:  Regression Analysis by Example, Chaterjee, Hadi & Price.

A good freely downloadable reference book is Practical Regression and Anova in R by
Faraway.

BLACKBOARD REGISTRATION
All students are required to register for the course in Blackboard, the GWU web-based
instructional resource. Course information and materials, including notes, grades, and
details about course assignments and quizzes will be posted there periodically. It is the
student’s responsibility to check the Stat 2118 Blackboard website frequently for up-to-
date information about assignments. Once enrolled in the course, you should automatically be registered on Blackboard. Log into the course website at: https://blackboard.gwu.edu/webapps/portal/frameset.jsp

COURSE STRUCTURE:
The class consists of two lectures and one lab per week. Your grade is based on homework, projects and two exams.

LEARNING OUTCOMES:
As a result of completing this course, students should be able to:
1. Enter tabular data using R. Plot data using R, do some exploratory data analysis.
2. Know basic methods required for data analysis and interpretation.
3. Formulate regression models for the data, while understanding some of the limitations and assumptions implicit in using these models.
4. Fit models and interpret the output.
5. Use diagnostic plots and tests to assess the adequacy of a particular model.
6. Find confidence intervals for the effects of different explanatory variables in the model.
7. Use some basic model selection procedures, as found in R, to find a best model in a class of models.
8. Fit simple ANOVA models, treating them as special cases of multiple regression models.
9. Fit simple logistic regression models.
10. Draw conclusions and communicate analysis results.

GRADING
- homework 40%
- data analysis project 20%
- two exams 20% each

- **HOMEWORK:** 5 assignments which will be collected and graded. All students must work independently on the homework sets. The HW will be posted on Blackboard or handed out in class, and must be submitted AND received through Blackboard. In the rare instance, Blackboard is unavailable and you are reaching the deadline, you must e-mail the assignment to the TA outside of Blackboard. Your TA will go over problems similar to the HW problems and software examples during lab sessions.

- **DATA ANALYSIS PROJECT:** There will be one data analysis project involving the use of statistical software to analyze data. A report along with the code and output will be turned in.

- **EXAMS:** Two closed book examinations will be given. The second exam will serve as the final exam for the course and will be given according to GW calendar. If an exam is missed, you will receive zero credit for that part of the grade. In order to qualify to take a make-up examination, you
must contact me prior to the exam and have a University approved excuse (e.g. well-documented medical problems) with documentation to verify that excuse. If the absence is not excused according to University guidelines, the examination grade will be recorded as a zero.

A grade of INCOMPLETE will ONLY be given to a student who is passing the course and cannot complete the course due to illness or other (well documented) circumstances beyond their control.

COURSE OUTLINE
1. Review: One and two sample problems, The Student’s $t$ distribution, Confidence intervals and Tests of Hypotheses
2. Intro to R: Functions, source, concatenation, sequences, plotting, saving plots, loops, built-in help, distributions in R.
3. Simple linear regression: Confidence and prediction intervals from linear models; Indicator variables;
4. Diagnostics for simple linear regression model
   a. Goodness of fit, $R^2$, sums of squares
   b. Outliers and non-constant variance
5. Multiple linear regression model
   a. Partial regression coefficients
   b. Matrix representation of coefficients
   c. Confidence intervals by hand
   d. A general $F$ test
6. Diagnostics and influence
   a. Standard diagnostic plots
   b. DFBETAS
   c. Cook’s distance
   d. DFFITS
   e. Bonferroni outlier test
   f. All influence measures
   g. Added variable plots
   h. Component + residual plots
   i. Hat values
7. Interactions & ANOVA
   a. Testing for interactions
   b. Visualizing an interaction
   c. Comparing models
   d. One-way ANOVA
   e. Two-way ANOVA
8. Transformations & Weighted Least Squares
9. Correlated Errors and Whitening
10. Model selection
    a. Election example: best subset
    b. Election example: stepwise
c. Cross-validation

11. Logistic regression
a. Logistic transform
b. Comparing models
c. IRLS (Iteratively Reweighted Least Squares)

12. Penalized regression
a. Bias-variance tradeoff
b. Cross-validation
c. Ridge regression
d. LASSO
e. Elastic Net

Computing environment: You will need to use a statistical package in order to perform most of the statistical analyses covered in this course. We will use mostly R, but I will be giving you instructions how to run regression analyses also in SAS. A brief introduction to the SAS package can be found at http://www.umass.edu/statdata/software/handouts/sas_online/


A list of books using R is given at http://www.r-project.org/doc/bib/R-publications.html

The R Manuals can be found by clicking here http://www.r-project.org/ and selecting Manuals under Documentation on the left column.

CLASS ATTENDANCE
Class discussions and programming demos will extend the material covered in the text. Therefore, it is important that you attend classes. You are responsible for any material covered or any announcements made in class, even if you are not present.

SUCCESS
Your success in this course depends upon you. Success in this course requires a great deal of time and effort. You should expect to spend at least 8 hours a week on this course.

While everyone would like to get an “A,” set a realistic goal for yourself, acknowledging that you may be unwilling (or unable) to devote the time and effort necessary to earn an “A.” If you are unable to make such a time commitment this semester, you might consider taking this course another semester.

ACADEMIC INTEGRITY
I personally support the GW Code of Academic Integrity. It states: “Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information.” For the remainder of the code, see: http://www.gwu.edu/~ntegrity/code.html
SUPPORT FOR STUDENTS OUTSIDE THE CLASSROOM

DISABILITY SUPPORT SERVICES (DSS)
Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 to establish eligibility and to coordinate reasonable accommodations. For additional information please refer to: http://gwired.gwu.edu/dss/

UNIVERSITY COUNSELING CENTER (UCC) 202-994-5300
The University Counseling Center (UCC) offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include:
- crisis and emergency mental health consultations
- confidential assessment, counseling services (individual and small group), and referrals
  http://gwired.gwu.edu/counsel/CounselingServices/AcademicSupportServices

SECURITY
In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.