SYLLABUS

SURV 615 (UMD)/SURVMETH 685 (UMICH)

Statistical Methods I

Fall 2011

(Last updated:September 20, 2011)

Instructor:	Sunghee Lee
Meeting Time:	Thursday 1-3:30 PM
Meeting Places:	1208 LeFrak (Maryland) and 368 ISR (Michigan)
Office Hours:	Thursday 3:30 - 4:30 PM or by appointment 4032 ISR or Skype (Skype name: sunghee0930)
Office Phone:	734-615-5264
Email:	sungheel@umich.edu
TA:	Julia Lee (julialee@umich.edu) Available Wednesday 1-2 PM via Ctools chat. Yiting Dai (ydai@survey.umd.edu) Available Tuesday 2-3 PM in-person. Send questions a day ahead to save time, if possible.
Class Website:	https://ctools.umich.edu/portal All the course-related material will be posted on the website: - Class notes by 9AM on Thursdays. - Homeworks and associated data files by 4PM on Thursdays. JPSM students need to create U Michigan friend's accounts to access the ctools website. If you have not received email about this, please contact Jill Esau (jesau@isr.umich.edu) ASAP.
Video site:	http://www.jpsmclasses.umd.edu Use "Username: umich" and "Password: psm" to access videos. Search for SURV 615. All lectures will be recorded and available on the mediasite 1 or 2 days after the class. More information on http://www.jpsm.umd.edu/support/ViewRecordings.pdf Instructions for remote access to JPSM network for running SAS at home: http://www.jpsm.umd.edu/support.htm. Non-JPSM students need to contact Duane Gilbert (dgilbert@survey.umd.edu) for an access.

Textbook(required):	***Applied Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS*** (Heiberger, R. & Holland, B., Noted as "HH" throughout the class.) (NOTE: Make sure to download data and programs as instructed in the preface. Look for "Online Files from Authors" on the website.) **Old class notes** (complied by Miller, S., Valliant, R. and Lee, S. Noted as "OC.") available on the class website.
Others (useful references):	***Statistical Methods***, 8th ed. (Snedecor, G.W. & Cochran, W. Noted as "SG.") ***The Little SAS Book***, 4th ed. (Delwiche, L.D. & Slaughter, S.J.)
Useful statistics review:	JPSM/PSM Orientation meterial
Useful computing help:	http://www.ats.ucla.edu/stat/
Useful SAS help:	http://www.jpsm.umd.edu/SAS/SASOnlineTutor/sot12/en/60476/saselmenu/ JPSM/PSM Orientation material
Useful Calculators:	http://danielsoper.com/statcalc3/

OVERVIEW

The purpose of this class is to learn basic statistical methods through the use of linear model theory and regression. The emphasis is to understand and apply the methods presented and develop a feel for how problems in data analysis can be viewed from several different ways. In all cases, the focus will be on understanding the techniques, rather than deriving their theoretical properties. The student are expected to increase their understandings about the techniques through homework assignments, apply the techniques in a final project and professionally deliver their understandings of applying such techniques to a real world problem through a project presentation. This class is not designed to offer programming instructions. The students are expected to have basic SAS skills and make an improvement throughout the course.

OUTLINE

(Note: All readings associated with a given class should be completed before the class meets, except for the first class. Readings in SC are supplements.)

CLASS 1 (Sep 8) Guest lecture by James Wagner

Mathematics vs. Statistics; Data and Matrix algebra; Review of basic statistical concepts; Basic graphical displays; Review of Statistics Classes in JPSM/PSM

Reading: HH Ch 1,2,3,4, Appendix A,C,F; OC 1&2; SC Ch 1,2,3; JPSM/PSM Orientation material

Homework #1

CLASS 2 (Sep 15)

Inferences with univariate and bivariate normal variables from single, two and paired samples. Test of normaility. Multiple comparisons

Reading: HH Ch 5,7; OC 1&2; SC Ch 4,5

Homework #2

CLASS 3 (Sep 22)

Simple linear regression. Basic assumptions and theory. Introductory analysis of residuals. Tests of normality, and tests for skewness and kurtosis.

Reading: HH Ch 8; OC 3; SC Ch 9

Homework #3

CLASS 4 (Sep 29) Instructor at MD

Multivariate normal distribution. Multiple linear regression. Basic assumptions. Matrix notation. Properties of the Least Squares estimates. Interpreting the coefficients. Partial correlation coefficients. ANOVA table.

Reading: HH Ch 6,9; OC 4; SC Ch 17
Homework #4

CLASS 5 (Oct 6)
Prediction (Dummy variable technique). General F-testing. ANOVA table. Linear transformations of predictors. Testing linear restrictions: Plug-in method, Testing coefficients method.
Project discussion 1.
Reading: HH Ch 9,10; OC 5
No Homework

CLASS 6 (Oct 13)
Model building with predictor variables. Polynomial models, and dummy variables. Analysis of covariance. Lack of fit tests with repeated observations.
Reading: HH Ch 10; OC 6; SC Ch 18
Homework #5

CLASS 7 (Oct 20) Instructor at MD
Regression diagnostics and model assessment. Studentized residuals. Measures of influence. Properties of Ordinary Least Squares when the assumptions are violated. Weighted Least Squares, and Generalized Least Squares. Lack of fit tests (variance known).
Reading: HH Ch 11; OC 7
Homework #6

CLASS 8 (Oct 27)
Transformations of variables in regression. Box-Cox transformations. Variance stabilizing transformations. Transformations to linearity. Testing for heteroscedasticity of error variances, and estimating variance functions. Estimated Weighted Least Squares.
Project discussion 2.
Reading: HH Ch 11; OC 8
No Homework

CLASS 9 (Nov 3)
Collinearity, and variable selection bias. Variable selection methods. Stepwise regression.
Reading: HH Ch 9; OC 9
Homework #7

CLASS 10 (Nov 10)
Instrumental variables. Regression with errors-in-variables. Regression estimation with samples with unequal probability of selection. Testing if the weights can be ignored.
Reading: OC 11; SC Ch 12

Homework #8

CLASS 11 (Nov 17) Instructor at MD
Fixed effects linear models. One-way analysis of variance. Balanced and unbalanced designs.
Project discussion 3.
Reading: HH Ch 12,13; OC 12; SC Ch 14,16
No Homework.

Nov 24-No class, HAPPY THANKSGIVING!

CLASS 12 (Dec 1)
Two-way analysis of variance. Balanced and unbalanced designs. Cell means model versus overparameterized model. Missing treatment combinations.
Reading: HH Ch 12,13; OC 12; SC Ch 14,16
No Homework

CLASS 13 (Dec 8)
Random effects linear model. One-way model. Two-way model, and nested model. Estimation of variance components.
Project discussion 4.
Reading: HH Ch 12,13; OC 13; SC Ch 13
No Homework.

CLASS 14 (Dec 15) Instructor at MD
Final project in-class presentation

GROUND RULES

Software:

a. SAS will be used exclusively in the class. Refer to HH Appendix C, ***The Little SAS Book***, and Orientation material.
b. Students are encouraged to learn R independently. Refer to HH Appendix B.
c. R can be used for the assignments and project, but the instructions will not be given.

Assignments:

a. Prepare in MS-Word format. Use MathType of other math symbol typing software for mathematical expressions.
b. Submit via the ctools website.
c. Homework assignments are due BEFORE the following class. Late homework assignments will be accepted until one week after the due date but will receive 50% of the points.
d. Late project abstracts and reports will not be accepted unless permitted by the instructor 3 weeks prior to the due date.
e. Graded assignments will be returned via the ctools website a week after the due date along with solutions.
f. Follow naming conventions for assignment files: `ASSIGNMENT_YOURNAME.xxx`
e.g., `hw1_sunghee_lee.doc; abstract_sunghee_lee.doc`
g. Students are encouraged to work together for assignments but in a productive way.
    - Limit what is shared.
    - Copying other's work will NOT benefit your education.
h. Students are expected to spend a number of hours on assignments.

Class etiquettes:

a. If you are at the remote site, stay within the camera angle.
b. No ringing phones.
c. Refrain from using personal computing/communication devices during lectures.

    - This includes laptops, netbooks, iPod, iPads, smart phones and cell-phones.
    - Of course, feel free to use them during breaks.

PROJECT

Projects are designed to allow each student some individual creativity in developing and expressing their own approach to data analysis, within the framework of the stated goals of SURV 615/SURVMETH 685.

Important Dates

Oct. 13, 2011:	Project abstract due
Nov. 3, 2011:	Project progress report due
Dec. 1, 2011:	Final project reports due
Dec. 15, 2011:	Final project presentation

Data

You can use datasets from your work or another source. The main requirement is that the statistical techniques covered in class can be applied to analyze the data.

The web site for the Inter-University Consortium for Political and Social Research, http://www.icpsr.umich.edu/, has many downloadable datasets that can be used for analysis. Some of the datasets are collected using sample surveys that include strata, clusters, and weights. Students in SURV 615/SURVMETH 685 are NOT expected to use the specialized techniques, taught in other courses, that have been developed for analyzing survey data. You can treat each dataset as if it were a simple random sample. However, keep in mind that if a dataset from a complex sample were analyzed appropriately, the relationships found could be different from those that you uncover.

Types of Analysis

The analysis you conduct should be ones that we have covered in the class. These include linear regression modeling, model diagnostics, model assessment, checks for collinearity, use of transformations, and any other techniques we will have studied. Although fitting of statistical models is expected, do not overlook the need for simple descriptive analysis to better understand the variables in your data. Your project goals should be to:

(a) Demonstrate that you know how to apply the methods covered in class in a systematic and thorough way to investigate relationships in a data set or to explore a particular problem, and

(b) Present your results coherently in a written report.

Reports should include tables and figures that illustrate the points you want to make and the analyses you have done.

Format of the Abstract

- Loosely follow Structured Abstract Format
  (Refer to http://research.mlanet.org/structured_abstract.html)
- 300 words limit

- Include the following three sections:
  - Objectives: State the essential issue the project attempts to address.
  - Data: Introduce the data source.
  - Analysis: Provide an analysis plan.
  - A sample abstract will be posted on the website.

Format of the Progress Report

- 2 page limit
- Include the following three components:
  - Objectives: State the essential issue the project attempts to address.
  - Data: Introduce the data source.
  - Variables: Clearly list dependent and independent variables.
  - Basic descriptive statistics: Provide means (or frequencies) and standard errors of the selected variables.
  - Models: Specify models of interest.

Format of the Report

The project reports take the form of a Lab Report (as discussed in Stat Labs: Mathematical Statistics Through Applications by Deborah Nolan and Terry Speed). Appendix A of Stat Labs (by permission of the authors) will be posted on the class website, as a guide to writing project reports. Reports should be 15-20 pages, counting tables and figures but excluding the abstract and appendix. Use the following format:

- MS-Word
- Space and one-half (do not use single spacing)
- 1 inch margins
- Include page numbers
- Parts of report
  - Title page with title of project, name, class, date
  - Abstract– 200 words limit
  - Introduction – state the problem you are addressing, the goals of your project, and your findings
  - Methodology – describe your data, how they were collected, the variables to analyze, and the models to consider
  - Results – present your findings along with a description of the techniques used for analysis. You can refer to particular sections of the class notes or text as part of your explanations. Graphs of different kinds are often a good way to display results (e.g., histograms, qqplots, residual plots, scatterplot matrices).
  - Discussion – what did you learn? Describe any limitations to your findings and follow-up work that might be appropriate.
  - Technical Appendix – this part should include SAS code and the parts of the output that are *relevant* to the discussion in the main body of the report. The appendix is not counted in the suggested length of 15-20 pages but should not itself exceed 10 pages. If, in the main body of the report, you refer to results

in the Appendix, you should refer to specific pages, figure numbers, or tables of the Appendix for clarity.

- A sample paper will be posted on the website.

Format of the In-Class Presentation

Each student will be given 10 minutes*. Follow conference or professional meeting presentation styles (refer to http://www.gwu.edu/~capstone/symposium/presentation_styles.htm "General Consideration" section). Present the issue, data, analysis, results and implications in a clear, concise and engaging manner. A sample powerpoint prestnation will be posted on the website.

* Presentation length subject to change.

Grading

The project will account for 50% of the grade in the class. Grades will be based on complexity of the project, thoroughness of analysis, correctness of results, quality of the written report and quality of the in-class presentation.

Extremely simple projects will receive poorer marks than more complex ones. Points will be deducted if reports are poorly written or poorly organized.

Presentation will be scored based on the usage of time, the clarity of the delivery, and the level of audience engagement.

Note

On purpose, the sample project whose abstract, report and presentation slides are provided in this class focuses on a categorical, not continuous, variable.

GRADING

Homework (50%):	Each homework assignment will receive a numerical score. The sum of all homework assignment scores will be re-scaled to 100.
Project (50%):	
Abstract (5%):	A numerical score scaled to 100.
Progress Report (5%):	A numerical score scaled to 100.
Final Report (30%):	A numerical score scaled to 100.
Presentation (10%):	A numerical score scaled to 100.

The final course grade is determined by assigning the letter grade corresponding to the result of the weighted average of the numerical scores for each of the assignments as follows.

A+	[98,100]
A	[93,98)
A-	[90,93)
B+	[87,90)
B	[83,87)
B-	[80,83)
C+	[77,80)
C	[73,77)
C-	[70,73)
D+	[67,70)
D	[63,67)
D-	[60,63)
F	¡ 60