STAT 102 Spring 2014
Section 003
Course Syllabus

The Wharton School of the University of Pennsylvania

Administrative Details

Class Meets: JMHH F85 - Tues. and Thurs. 3:00 – 4:30PM
Pre-Requisites: STAT 101 (or AP Stat)

Course Instructor: Justin Bleich
- Office: JMHH 414.2
- Contact: `jbleich@wharton.upenn.edu`
- Office Hours: *Tentative*: Thursday, 4:30 - 5:30 PM (or by appointment)

Teaching Assistant: Dan McCarthy
- Office: JMHH 454
- Contact: `danielmc@wharton.upenn.edu`
- Office Hours: TBA

Teaching Assistant: Xin Lu Tan
- Office: JMHH 476
- Contact: `xtan@wharton.upenn.edu`
- Office Hours: TBA

Note: Unlike in previous semesters of STAT 102, this section of the course is standalone from sections 001 and 002.

Course Overview

As a second semester of business statistics, the focus of this course is on regression analysis. Relying on the building blocks from STAT 101, the course will cover the fundamentals of building regression models, analyzing and interpreting associated output, and understanding the roles of statistical inference and prediction in regression settings.

As the business world rapidly progresses towards a paradigm of data-driven decision making, the *primary goal* of this course is on understanding both the power and limitations of regression analysis so that students can comfortably apply regression models in *real-world settings*. Equipped with the course tools, the hope is that each student can be a successful data analyst in any career path that he or she may choose. While some computation and theory is necessary

to achieve our goal, it is not meant to be the emphasis of the course.

A tentative list of topics that may be covered includes:

- Review of Frameworks for Statistical Inference

- Two-Sample Hypothesis Testing and Confidence Intervals

- Simple Linear Regression: Properties, Inference, and Model Diagnostics

- Multiple Linear Regression and Associated Statistical Inference

- Model Building and Diagonstics for Multiple Regression

- Categorical and Interaction Terms in Regression Models

- Collinearity in Multiple Regression

- Out-of-sample Validation and Models for Pure Prediction

- Time Series Analysis

- Logistic Regression Models

- Additional Topics from Modern Regression (time permitting)

As a core Wharton course, the focus of the examples will be on how to use models to assist in decision making across fields such as finance, management, marketing, logistics, etc., but data sets and analysis from other interesting disciplines may appear as well.

Course Materials

Lecture Slides: The lecture slides will be the primary learning guide for the course and should be fairly complete (with some details to be filled in during lecture). The slides will be posted by the morning of lecture (sometime the night before) and you are *highly* encouraged to bring them to class as an aid.

Textbook: *Applied Regression Analysis: A Second Course in Business and Economic Statistics*, Dielman, 4th Edition. This book is required for the course. Some lectures will line up with the text and others will not. The correspondence will be noted in each lecture. Homework problems may be taken from the text as well.

Additional readings and material may be provided on Canvas.

Computer Software: **JMP 11** (or 10), available at `http://upenn.onthehub.com` (3-year license for $59.95). Make sure to select your platform (Windows/Mac). JMP is also available on all Wharton computers under the Statistics tab in the Departmental Apps folder. As an official course policy, we do *not* recommend not buying a copy and relying on public Wharton computers. This will likely be inconvenient.

I will make use of `R` occasionally throughout the semester to demonstrate certain concepts via simulation. I will post the scripts for these simulations, which you can run yourself if you wish. You can obtain a free copy of `R` from `http://cran.us.r-project.org/` by clicking on the appropriate download link, but **you are not required to download** `R`.

Course Requirements

Grading Policy

Your course grade will be calculated as follows:

1. Midterm Examination - 30%

2. Final Examination - 45%

3. Quizzes - 12.5%

4. Homework- 12.5%

More information about final course grades will be provided in class.

Homework

Homework is an extremely important aspect of the course. Success is quantitative courses *heavily* rests on practice and experience. To quote a colleague of mine, Adam Kapelner, "homework is kind of like weightlifting; you have to lift weights to build muscle."

There will be two types of homework assigned regularly throughout the semester: problem sets and data analysis mini-projects.

Both types of assignments are to be turned in to specified boxes in the Statistics department in JMHH by the due date specified on the assigment. Graded work will also be available for pick-up in these boxes.

Late Homework Policy: Late assigments will be penalized at 10% of the maximum grade per day for up to three days and are ineligible to be handed in for credit after this time. Late assignments should placed directly in my mailbox in the Statistics department.

Regrade Policy: Regrade requests for homework will be entertained for up to two weeks after the homework is officially returned and solutions are posted (starting from whichever occurs later). After two weeks, the homework grades are finalized. It is your responsibility to remain current with the course.

Requests require a written explanation on a seperate cover page as to why the assignment is incorrectly graded, referencing specifically the questions for regrade. Failure to abide by this policy will result in a loss of regrade privileges.

Problem Sets

Problem sets will consist of textbook problems and other questions that are meant to (usually) have a single solution. The goal of problem sets is to obtain a basic mastery of concepts. Problem sets may be completed **individually or in pairs of two students**. While collaboration beyond each pair is encouraged at a high level, there will be zero tolerance for write-ups that are not completed independently by each student or pair of students.

Additionally, the lowest problem set score (not mini-project) will be dropped.

Data Analysis Mini-Projects

Unlike many mathematics courses, there is generally no right answer in a statistical analysis of a data set. Often, a variety of issues appear and the data analyst is required to make certain tradeoffs. The aim of these mini-projects is to tackle analyzing a data set as if it were a task assigned in the real-world. Although general guidelines towards best practices may be provided, these projects will largely be open-ended with no specific right answer. An emphasis will be placed on proper interpretation of findings and students may be asked to write an executive summary that is suitably understandable for managers or clients lacking a quantitative background.

There will be up to three mini-projects and the size of groups for these assignments will be determined at a later date.

Forecasting Competitions (Optional)

Kaggle, a startup based in San Francisco, offers a solution to clients requiring predictive models for their businesses by holding forecasting competitions. Any data scientist is allowed to build and submit forecasting models and the winning model receives a monetary prize (often $10,000) in addition to being implemented by the sponsoring company.

For at least one of the data analysis mini-projects, there will be the option to participate in our own STAT 102 forecasting competition using the project data set. Students will submit predictions and the team with the "best" predictions (to be defined more rigorously) will receive extra credit towards either the midterm or final examination. More details will be provided later in the semester.

Examinations

- The midterm examination will be held from 6:00 – 8:00 PM on Tuesday, March 4th.

- The final examination will be held from 6:00 – 8:00 PM on Monday, May 12th.

More details about the examinations, such as format, rules for cheat sheets, and room assignments will be announced closer to the exam dates.

Regrade Policy: As with homeworks, requests will be entertained for up to two weeks after the solutions are posted and exams returned, and a cover sheet of questions to be regraded must

be provided. Additionally, the course staff reserves the right to regrade the entire exam if it is submitted.

Quizzes

There will be periodic quizzes consisting of 5 multiple choice questions that will take place in class (for about 10-15 minutes). The aim of these quizzes is to encourage students to remain up-to-date with course material. These quizzes will be announced one week in advance.

Use of Canvas

All necessary materials, grades, and information will be posted on Canvas.

You are encouraged to use the discussion board to post questions and comments about the course. I encourage you all to reply to your fellow classmates inquiries, as one of the best ways to learn is through teaching another. The course staff will do its best to provide answers as well. I may also post pertinent questions I receive via email in addition to my reply, leaving the original student's indentity anonymous.

If you are not a Wharton student and need access to Canvas, you can create an account by visiting `accounts.wharton.upenn.edu` once the semester begins.

Datasets

If you have an interesting data set that you'd like to contribute to the course, please let me know! I'm always happy to incorporate novel and interesting data sets into some aspect of the course when possible, and it can help improve the overall course experience for everyone.