70208-Regression Analysis

Summer 2017

Instructor

Kole Reddig
Office: 315C GSIA (Tepper School of Business)
E-mail: kreddig@andrew.cmu.edu
Office hours: Monday and Thursday 10:30am-12pm or by appointment

Course Website

Located on Canvas

Dates and Lectures

Classes are held Monday through Friday, 9:00-10:20am, Baker Hall, 235B.
First day of class is Monday, July 3, 2017.
The last day of class is Thursday, August 10, 2017.
The Final Exam will be held Friday, August 11, 2017.

Class Overview

This class is a part of the core statistics sequence for undergraduate business majors, following 70-207. The previous course discussed univariate probability concepts to understand the behavior of a single random variable. In this class, we build on these concepts to study the statistical relationship between two or more random variables with a method called *linear regression*.

Throughout the course, we will consider both the theory and applications of linear regressions. The goal of this course is that students 1) learn how to interpret regression results presented by someone else, and understand the strengths and shortcomings of a particular analysis; 2) learn how to use regressions to do their own analysis—including gathering information about relationships between variables, assessing the accuracy of this information, making predictions according to this information, and assessing the accuracy of these predictions.

Recommended Textbook

Applied Regression Analysis, 4th edition, by Terry Dielman (Duxbury Press). ISBN 053446548. This book is recommended, but *not* required.

Class topics and readings

Lectures will not strictly follow the book, so you should come to class and take notes. The book is a complement, not substitute, for attendance and taking notes during lecture.

Here is a list of the topics covered in the course. I also listed the readings in the textbook for each section. Not all topics are covered in the book. For those topics, you will have to rely on lecture notes.

1. Inference about a population mean
   - Population vs sample (2.1)
   - Sampling distributions (2.5)
   - Law of Large Numbers
   - Central Limit Theorem (2.6)
   - Confidence interval for the sample mean (2.6)
   - Hypothesis testing for the sample mean (2.7)
2. Multivariate populations
   - Bivariate probability distributions
   - Marginal distributions
   - Conditional expectation and variance
   - Covariance and coefficient of correlation
   - Statistical independence
   - Conditional expectation function
3. Univariate Regression
   - Least squares (LS) line (3.1 and 3.2)
   - Least squares algebraic properties
   - Correlation vs causality (3.7.1)
4. Classical (univariate) regression model
   - Assumptions (3.3.1)
   - Estimation by LS
5. Inference in the classical regression model
   - Standard error of the regression (3.3.2)
   - Sampling distribution of the LS estimators (3.3.2)
   - Hypothesis testing and confidence intervals in the classical regression model (3.3.2)
6. Prediction and fit in the classical regression model
   - Prediction and prediction error (3.5.1 and 3.5.2)
   - Standard error of the regression
   - R-squared (3.4.1 and 3.4.2)
7. Multiple regression
   - LS for multiple independent variables (4.1)
   - "Short" vs "long" regression
   - Omitted variable rule

8. Classical regression model: multiple regression
   - Assumptions (4.2.1)
   - Estimation by LS
   - Inference (4.2.2)
   - Fit of the multiple regression (4.3.1)
   - F-tests (4.3.2)
   - Multicollinearity (4.6.1 and 4.6.2, until "Large F, small t")
9. Functional form issues in the classical regression model
   - Polynomials (5.1 and 5.2.1)
   - Transformation of the dependent and/or the independent variables (5.2.2, 5.2.3, 5.2.4, 5.2.5)
10. Dummy variables
    - Fixed effects (7.1)
    - Interaction effects (7.2)
    - Application to seasonal effects (7.3)

Lecture materials

Lecture notes will be posted before each class. These are just outlines, and not substitutes for attending each class. You are expected to attend class and take notes.

Grading

The total percentage grade is weighted as follows:

Problem Sets	25%
Midterm	25%
Final Exam	40%
Participation	10%

The final letter grade is determined according to the total percentage grade:

A	$\geq 90\%$
B	$< 90\%$ and $\geq 80\%$
C	$< 80\%$ and $\geq 70\%$
D	$< 70\%$ and $\geq 60\%$
R	$< 60\%$

The grading bounds may be lowered, but they will not be made higher.

Problem Sets

There are 5 problem sets in this course. You may work in groups to solve them. If you do work in a group, you must still submit your own copy of the solutions, and write the names of the people that you worked with on your assignment. Problem sets are aimed to help you learn the material, and therefore it is important that you think through each problem and understand them to help prepare you for the exams.

Problem sets will be posted on Canvas on Tuesdays, and are due on Tuesday the following week. The problem sets must be turned in during lecture. Paper copies must be turned in, and no late assignments will be accepted. Solutions will be posted on Blackboard after the problem sets are turned in.

For the final grade in the course, the lowest one problem set grade will be dropped.

Exams take place in class. If you cannot make an exam, contact me within 2 days of the test to justify your absence. Excused absences include CMU activities, serious illness, and family emergencies. For medical emergencies, we need documentation of your illness from a doctor.

If you need extra time for an exam due to medical reasons, please turn in the documentation a week before the exam.

If you think there was a mistake in the grading of a homework or exam, email Kole and state the question that was graded incorrectly, and why it was graded incorrectly.