70208-Regression Analysis

Fall 2019

Instructor

Professor Rebecca Lessem (Mueller)
Office: Tepper 5215
rlessem@andrew.cmu.edu

Course website

Located on Canvas

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. We use regressions to summarize the relationship *in the sample* between a dependent variable and one or more independent variables. We can use the relationship in the sample to estimate the relationship *in the population* and conduct inference about it. This allows us to determine the precision of our estimates and assess the accuracy of the predictions we make. To conduct this inference, we use statistical theory.

Throughout the course, we will look at a number of applications to show how these methods are used. This will involve both in class examples and assignments that give you the opportunity to work with the data on your own. Lectures and assignments will teach how to use these methodologies and to interpret the results, allowing you to develop the skills to be informed users of regression methodologies. We will do a number of business related applications to highlight how these tools will be useful in the workforce.

Learning objectives

By the end of the course, you will be able to do the following:

1. Interpret the coefficients from a linear regression.
2. Judge the goodness of fit of a linear regression.

3. Write down the regression model best suited to answer the question of interest.
4. Determine the accuracy of coefficient estimates and predicted values from a linear regression.
5. Test hypotheses involving parameters of the classical linear regression model.
6. Identify the factors related to the magnitude and precision of parameter estimates from linear regression.

Class topics

Here is a list of the topics covered in the course.

1. Inference about a population mean
    - Population vs sample
    - Sampling distributions
    - Law of Large Numbers
    - Central Limit Theorem
    - Confidence interval for the sample mean
    - Hypothesis testing for the sample mean
2. Multivariate populations
    - Bivariate probability distributions
    - Marginal distributions
    - Conditional expectation and variance
    - Covariance and coefficient of correlation
3. Univariate Regression
    - Least squares line
    - Least squares algebraic properties
    - Correlation vs causality
4. Classical regression model: univariate regression
    - Assumptions

- Estimation by least squares
- Sampling distribution of the least square estimators
- Hypothesis testing and confidence intervals in the classical regression model
- Predictions
- Assessing the fit of a regression

5. Multiple regression
   - Omitted variable bias
   - Multicollinearity
6. Classical regression model: multiple regression
   - Fit of the multiple regression
   - F-tests
   - Interpretation of results: understanding magnitude and precision
7. Functional form issues in the classical regression model
   - Polynomials
   - Transformation of the dependent and/or the independent variables
8. Application: testing the efficient market hypothesis
9. Dummy variables
   - Fixed effects
   - Interaction effects
10. Applications of dummy variables
    - Racial wage gaps, "The role of premarket factors in black-white wage differences", by Derek Neal and William Johnson
    - Recycling behavior, "Promoting Recycling: Private values, social norms, and economic incentives", by W. Kip Viscusi, Joel Huber, and Jason Bell, 2011
    - Management practices, "Do private equity owned firms have better management practices", by Nicholas Bloom, Raffaella Sadun, and John Van Reenen, 2015
11. Experiments to inform decisions

- Policy experiments, "The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment", by Paul Ferraro, Juan Miranda, and Michael Price
- Unemployment duration, "Duration dependence and labor market conditions: Evidence from a field experiment," by Kory Kroft, Fabian Lange, and Matthew Notowidigdo, 2013

12. Probit regression