Prediction and Multiple Regression

Measurement Methods

Lecture 14

Chapter 8

Today's Class

- Prediction
- Simple linear regression
- Multiple regression

Upcoming Schedule

Our Upcoming Schedule

Note: THE CLASS PROJECT IS DUE ON 5/16

Date	Chapter and Topic
3/30	Chapter 7: Reliability – Applications
4/4	Computer Lab session
4/6,4/11	No Class
4/13	Chapter 8: Prediction and Multiple Regression
4/18, 4/20	Chapter 9: The Common Factor Model
4/25	Chapter 10: Validity
4/27	Chapter 11: Classical Item Analysis
5/2	Chapter 12: Item Response Models
5/4	Chapter 13: Properties of Item Response Models
5/9	Chapter 14: Multidimensional Item Response Models
5/11	Review
5/16	Final Exam (1:30pm – 4:00pm)

Introduction to Regression

Motivation

- Often, scores on a single test can be used to predict some further performance external to the test itself.

- By scores, I mean:
  - Test score
  - Subtest score
  - A set of scores on a battery of tests
  - Item scores

- The key is the differential weighting that is applied by the regression analysis.

Motivation

- Because of these topics, this chapter (Ch. 8) provides a review of multiple regression.

- Some basic terminology:
  – Criterion – the thing begin predicted (AKA the dependent variable, the response, or the outcome).
  – Explanatory variable(s) – the variables being used to predict the criterion (AKA the independent variables).

Bivariate Regression

Simple Linear (Bivariate) Regression

- To introduce regression, consider the case where we have only two variables, X and Y.
- To begin, lets suppose that we know the means of X and Y (μ_X and μ_Y).
- Lets also suppose that we know the variances of X and Y (σ_X^2 and σ_Y^2) and the covariance of X and Y (σ_{XY}).

Linear Regression

- We assume that there is a linear relationship between X and Y:

$$\hat{Y} = \alpha + \beta X$$

- The mean (or expected value) of Y **given** X is given by the functional equation above.

- The **regression coefficient**, β, is the expected change in Y corresponding to a unit change in X.

- The **regression constant**, α, is the intercept – the point where the line crosses the Y axis (where X=0).

Linear Regression Model

- The model for the linear regression is a bit more general:

$$Y = \alpha + \beta X + E$$

- Here, E is the error of prediction of Y given X.
- E is uncorrelated with X.
- E is also referred to as the residual.

Re-expressing the Coefficients

- We can re-express the coefficients of the regression model in terms of our means, variances, and covariances:

$$\beta = \frac{\sigma_{Y\,X}}{\sigma_X^2}$$

$$\alpha = \mu_Y + \frac{\sigma_{Y\,X}}{\sigma_X^2}\mu_X$$

More Fun With Linear Regression

- Because X and E are uncorrelated:

$$\sigma_Y^2 = \beta^2\sigma_X^2 + \sigma_E^2$$

- Which actually becomes:

$$\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_E^2$$

- The last item is a nice decomposition into variance explained by the model, and variance due to error.

Variance Accounted For

- We often describe the efficiency of a regression model in terms we usually attribute to other things in class.

- The variance accounted for by a regression is the ratio of the variance of the predicted value of Y to the variance of the Y:

$$\frac{\sigma^2_{\hat{Y}}}{\sigma^2_{Y}}$$

Example of Simple Linear Regression

- Imagine that X is the number right on a test of aptitude covering topics in 790/791 (linear models) for graduate students in PSYC at KU.
- Image that Y is the grade-point average in the first year of graduate studies
- Here:
  - $\mu_X = 50$
  - $\mu_Y = 3.0$
  - $\sigma_X^2 = 25$
  - $\sigma_Y^2 = 0.64$
  - $\sigma_{YX} = 3.5$

More of the Example

- Here:
  - $\mu_X = 50$
  - $\mu_Y = 3.0$
  - $\sigma_X^2 = 25$
  - $\sigma_Y^2 = 0.64$
  - $\sigma_{YX} = 3.5$
- $\beta = 3.5/25 = .14$
- $\alpha = 3 - .14 \times 50 = -4$
- So, $\hat{Y} = -4 + 0.14X$

But Don't Forget…

$$\sigma^2_{\hat{Y}} = 0.49$$

$$\sigma^2_E = 0.15$$

$$\text{VAC} = \frac{\sigma^2_{\hat{Y}}}{\sigma^2_Y} = 1 - \frac{\sigma^2_E}{\sigma^2_Y} = 1 - \frac{0.15}{0.49} = 0.766$$

Multivariate Regression

Multiple Regression

- The world isn't always simple enough to have a single variable completely explain variance in another.
  - If it was, we could all go hang out on the beach.
- Instead, we have to consider the case when we have multiple predictor/independent (X) variables.

MR Math

- The math behind multiple regression is much more tedious than in bivariate regression.
  - It is best left to your favorite stat package or undergraduate student.

- The main point to take away is that we can do least squares multiple regression very simply with matrices:

$$\beta = (\mathbf{X'X})^{-1}\,\mathbf{X'Y}$$

Multiple Regression Equation

- For two independent variables, X_1 and X_2, predicting Y, the general form of the multiple regression equation is given by:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + E$$

- Here, we again typically suppose that E is uncorrelated with both X_1 and X_2.

Variance Accounted For

- Again, we can describe the variance in Y accounted for by variance in the predicted value of Y.

- It all revolves around taking the squared correlation between the predicted value of Y and the true value of Y.

Other Issues

Properties of the Estimates

- The ordinary least squares estimates of the regression model parameters minimize the sum of the squared distance between the regression line and the points observed in the data.

- In a statistical sense, these estimates are the best unbiased estimates of the parameters of the model.
  – The model R^2 is positively biased, however.

Partial Covariance and Correlation

- We often encounter times where we would like to see the relationship between two variables when the values of a third are held constant.
- The partial correlation is the measure that does just that.
  - It can be computed from correlating residuals of multiple analyses where some variables are used and others not.
  - For example consider taking the residuals of the regression of X_1 on Y and the residuals of X_2 on Y.
    - The partial correlation between X_1 and X_2 holding Y constant would be the correlation between the two sets of residuals.

Partial Correlation in Practice

- Often, the variable(s) we fix in a partial correlation is thought to have a causal influence on the other two variables.

- We want to determine if the correlation between two variables is spurious or not.

An Empirical Example

Example

- To elaborate on an example, consider p. 159 on in Ch. 8.

- The example given describes regression in its full glory.

- What you will find is that the motivation for providing a review of regression is that we are doing regression in FA, but with unobservable variables.
  - Much of the math remains the same, though.

Wrapping Up

- This chapter covers the concepts of regression.
  - Please read this, it is practical information.
- Regression is something that has been around for many years.
  - Many of these applications are classics.
- This is probably my least favorite chapter in the book.
  - It seems somewhat out of place in the context – not enough information tying it to other material presented (although concepts presented are fundamental in some of our ideas about factor analysis).

Next Time

- The common factor model.
  - What you do if your test measures more than one trait.