

### Independent Project

The goal of this project is to describe the relationship between a dependent variable (Y) and at least two independent variables ( $X_1$  and  $X_2$ ) in a multiple regression context using the format of a typical econometric study. That is you will develop and estimate a model that describes how one variable is determined by two or more independent variables.

Pick a topic that interests you -- three restrictions apply.

1. You must use at least 30 sample observations.
2. No two students can use the same data.
3. The data must have some relevance to health care – and preferably would be related to your job in some way.

Turn in the variables' names and data sources; identify which variable is the dependent variable, and which variables are the independent variables. The variables and data sources must be approved prior to the Preliminary Report

### Preliminary Report (15 percent of HCAI 5220 final grade)

You will submit a preliminary report that addresses the following questions by **Monday October 12<sup>th</sup>**. This can be in presentation form. Remember to include citations for the data as well as other appropriate references. You do not need to estimate any regressions at this point; the preliminary report is primarily descriptive.

1. What theory are you testing and why is it interesting? Write out the equation you are estimating in the form:  $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$   
Should  $\alpha_1$  (the marginal influence of  $X_1$  on Y) be positive or negative and why?  
Should  $\alpha_2$  (the marginal influence of  $X_2$  on Y) be positive or negative and why?
2. What data are you using? Are these data adequate for testing your theory? Are there variables you are leaving out whose omission may confound your results? What would the ideal data set look like? Explain.
3. Provide some basic descriptive statistics for your data:
  - overall average and standard deviation for all variables
  - mean and standard deviation of the dependent variable by group (for interesting/relevant groups)
4. Graph your sample data for Y and  $X_1$  and do a second graph for Y and  $X_2$  (use horizontal and vertical reference lines to represent the variables' means). Do the variables appear to be linearly related? Do the variables appear to be positively or negatively related? Do the variables appear to be strongly related? **If you are estimating a model with more than 2 independent variables, pick 2 to do the analysis in questions 4, 5, and 6.**
5. Estimate and interpret two correlation coefficients (simple correlations, not regression): the correlation between Y and  $X_1$  and the correlation between Y and  $X_2$ . For each pair, test the null hypothesis that the population correlation coefficient equals zero against the relevant alternative

hypothesis. Do the variables appear to be linearly related in the manner (sign, strength) that you expected?

6. Perform two chi-square tests of independence: one for Y and X<sub>1</sub> and one for Y and X<sub>2</sub>. Divide each variable into three categories and explain the bases for your categorizations. Do the variables appear to be related?
7. Are the graphical analyses, correlation tests, and chi-square tests consistent with each other and with your theory? If not, what could explain the inconsistency?

**Written Term Paper** (30 percent of your final grade in 5221)

Turn in a typed paper that follows the outline below by **5 p.m. on Friday December 11<sup>th</sup>**.

**Presentation** (15percent of final grade in 5221)

In our last teleconference (**December 5<sup>th</sup>**) each student will make a 10 minute presentation to the class. The presentation should provide the motivation for the study, the hypotheses being tested, a presentation and interpretation of the results, and recommendations for further research.

### **Term Paper Outline**

#### **Executive Summary**

This is a one-paragraph description of your conclusions and how you reached them. Don't summarize everything that you did. Concisely state what you did, what you found, and what it means.

#### **I. Introduction**

Introduce the topic by describing what you are doing and why it is important or unique. How will the findings of the study help decision making at your organization? **The introduction (and conclusion) should be devoid of any statistics, equations, etc. These sections should be readable to someone with little to no statistical knowledge and they should have a good idea of what you did, what it means, and what the pros and cons of the analysis are.**

#### **II. Model Development**

Write out the model you are estimating in the form of the following equation:

$$(1) \quad Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$$

Carefully develop the theory that you are testing in the context of a linear regression model as given in equation (1). Clearly explain which variables are the independent variables, which is the dependent variable, and any signs (positive or negative) or numerical values for the y-intercept and the slope coefficients that the theory suggests.

#### **III. Discussion of the Data**

Are the data appropriate to test the theory? How closely do your data match your theory? Describe the data sources, the sampling procedures used to collect the data, and possible nonsampling errors.

Summarize your data with appropriate sample statistics (mean, median, percentiles, standard deviation, CV, etc.). What do the statistics tell you about the sample (*e.g.*, the degree of dispersion or the symmetry

of the data)? Do you have any interesting observations or observations with high leverage? Might those observations affect your analysis and, if so, how?

#### IV. Regression Analysis

In your discussions, do not discuss the mechanics of the calculation but, instead, emphasize the interpretation and implications of the estimates. Do not simply paste the statistical results as presented in *Excel*; instead, create more informative tables.

When you present the regression analyses, be certain to discuss the following points.

- a. Interpret the estimated regression coefficients for the population model of equation (1) in their context. Do they make sense intuitively and why or why not?
- b. How accurate is the estimation of the linear representation given in equation (1)? Look at the coefficient of determination and the standard error of the estimate and interpret. What does this tell you in terms of your model?
- c. Compare the estimates derived from equations (2) and (3) to the corresponding estimates derived from equation (1). Note that equations 2 and 3 below are simple regressions which only include one of your X variables.
  - (2)  $Y = \beta_0 + \beta_1 X_1 + \varepsilon$
  - (3)  $Y = \gamma_0 + \gamma_2 X_2 + \varepsilon$
  - i. Are there any significant changes in the signs or values of the estimated slope coefficients (estimated  $\alpha_1$  compared to estimated  $\beta_1$  and estimated  $\alpha_2$  compared to estimated  $\gamma_2$ )?
  - ii. Are there any significant changes in the adjusted  $R^2$  or the standard error of the estimate in moving from the estimated equations (2) and (3) to the estimated equation (1)?
  - iii. Clearly explain why the changes in (i) and (ii) occurred and what they mean. Again, remember that this point is related to the strength of the relationship between  $X_1$  and  $X_2$  (the extent of multicollinearity) as well as to the strength of the marginal influence of each variable upon Y.
- d. Perform the appropriate hypothesis tests to determine if there is a linear relationship as specified in equation (1). Clearly interpret the results. Did multicollinearity affect your t-test results? The following null hypotheses should be tested and interpreted.

$$H_0: \alpha_1 = \alpha_2 = 0,$$

$$H_0: \alpha_0 = \mu_Y \text{ (the intercept equals the mean value of Y)}$$

$$H_0: \alpha_1 = 0$$

$$H_0: \alpha_2 = 0$$

**Your regression results should be summarized in table form in the paper. Do not simply cut and paste the Excel output, but create a table that makes the interpretation of the results easy for the reader.**

#### V. Conclusion

Discuss the implications of the analyses for your theory. Focus on the multiple regression analysis and the interpretation of your sample results in that context. Clearly explain if the data support your theory as specified in equation (1) or if the data more strongly support equation (2) or equation (3). Are the conclusions drawn from your different tests consistent, and why or why not? These conclusions might

relate to validity of the theory, to the testability of the theory, or the quality of your data. Give suggestions for future, improved research in this area.

## **VII. Appendices**

Include your *Excel* program (as an email attachment not as part of the paper) with appropriate comments in it (so that a reader can follow it more easily).

### **Previous Topics**

To give you an idea of what kinds of projects I am looking for, the following is a list of some of the topics from previous classes.

Predicting Emergency Department Utilization  
The Effect of Patient Age and Surgical Time on Postoperative Temperature  
Assessing Variables in the Risk for Infection in the Postoperative Period  
The Value of Apgar Scores: The Relationship between payor source, race, and gestational age  
The Effects of Age, ASA functional status, Duration of Surgery, Body Mass Index and Emergency Operations in Determining Pharmacy Costs of Anesthesia Drugs  
The Effects of Age, Severity of Illness and Length of Stay on the Cost of Coronary Artery Bypass Graft Procedures  
The Determinants of CBG Mortality Rates  
The Determinants of Patient Satisfaction  
The Effect of a New Triage Process on ED Turnaround Time  
The Effects of ancillaries on the day of discharge and subspecialist on the time to discharge  
The Determinants of Missed Appointments at VA Hospitals  
The Effect of BMI on Worker's Compensation Costs  
Reducing Post Operative Pain in an ASC Setting  
The Effect of Pressure Ulcers and Diabetes on Length of Stay  
The Determinants of Patient Satisfaction  
Effects of Pre-admitting and Insurance Coverage on Upfront Cash Collections  
The Effects of Age and Cefazolin Administration on Infection Rate  
Analyzing the Effect of Amniotic Fluid Levels on Fetal Lung Maturity  
Rounding Hospitalists Model and Its Impact (or lack thereof) on Patient Length of Stay  
The Influence of Provider, Weekend Day, Census and Admissions On Emergency Department Turnaround Time  
The Effects of Pharmacy and Inhalation Therapy Interventions on LOS of Skilled Nursing Facility Patients  
An Analysis of Wait Time from Registration to Triage  
The Determinants of Turnover Rates among Hospital Employees  
The Relationship between Physician Satisfaction and Hospital Employee Satisfaction  
The Effect of Day of Admission and Age on Length of Stay  
The Effects of Housekeeping on Patient Satisfaction in Hospitals