Problem Set 4

Introduction

You work in the shipping and logistics department for Beast Buy, an American mail order company that specializes in pet food for VERY exotic pets. Beast Buy has four warehouses/distribution centers that provide product to each of four different regions in the US; Atlanta services the southeast, Boston services the northeast, Cleveland services the midwest, and Denver services the west. Your manager has recently asked you to analyze the efficiency of these distribution centers.   
The file [beastbuyfa10.xlsx](http://polaris.umuc.edu/~mdobra/fall10/beastbuyfa10.xlsx) contains data from each of the last 26 weeks for each of the four distribution centers. The first column simply denotes the week in which the data were collected. The second column indicates which warehouse the data are from (1=Atlanta, 2=Boston, 3=Cleveland, 4=Denver). The third column contains the distribution cost (in thousands of US dollars) associated with the particular warehouse in each week, and the final column contains data on the number of orders routed through each warehouse each week. Use this data to answer the following questions. 4.3 is extra credit.

Problem 4.1

The first issue your boss has asked you to address is whether or not there are differences in distribution cost between each of the four warehouses. Use the 0.05 level of significance to:  
a) Perform a one-way ANOVA to look for differences in distribution costs between warehouses.  
b) If the results in (a) indicate that it is appropriate, use the Tukey-Kramer procedure to determine which distribution centers differ in mean distribution costs.  
c) Briefly summarize (in plain English) your procedures and the results of (a) and (b) for your manager.

Excel Tips: When using the Data Analysis ToolPak, Excel requires that your data be formatted differently for ANOVA than for regression. The data as downloaded is formatted correctly for regression analysis, so you will have to transform your data prior to estimating the ANOVA.

Problem 4.2

In addition to looking at differences between distribution centers, your manager also wants to know the relationship between the number of orders routed through each center and the distribution cost. Thus, the number of orders is your independent variable and the cost is your dependent variable.   
a)Construct a scatter plot of the two variables.  
b)Estimate a simple linear regression between these two variables.  
c)Interpret the meaning of &beta 0 and &beta 1.  
d)Predict the mean distribution costs of 500 orders, 1000 orders, and 1500 orders. Are these appropriate predictions?  
e)Comment briefly on the predictive power/statistical significance of your estimates.

Problem 4.3

EXTRA CREDIT. To answer this question, you will need to read the materials for week 11 of the class. Here, your task is to combine 4.1 and 4.2 into a single multiple regression model using dummy variables. To put the data in an Excel-friendly format, you will first need to create dummy variables for each of the different processing centers. a) Estimate a multiple regression model, again using cost as the dependent variable, however for your independent variables you will want to use orders AND your set of dummy variables (see technical note below).   
b) Comment on the results from (a) in light of your results in 4.1 and 4.2. Does the coefficient on orders here match your results from 4.2? Do the coefficients on your dummy variables, and differences between the coefficients on your dummy variables, match up with your results from 4.1?