**SCM 651 Fall 2017 Data Analysis Exam Version 6**

**Total Points = 50**

**Due Date: Midnight, Tuesday, December 12, 2017**

**Instructions and Academic Integrity Policy:**

This is an individual exam. By submitting the exam, you are stating that you did not provide help to or take help from anybody else. A failure to meet this condition is a violation of the Academic Integrity Policy of the Whitman School of Management.

1. All the questions can be answered based on the study of the handouts provided in class and posted on Blackboard.
2. There are three data sets attached to this email. All questions are based on these data.
3. There are **multiple versions** of the exam. The data sets emailed to you have names that include the **version of the test** you are asked to answer.
4. Please answer the questions based on the data emailed to you. If your answer to a question **does not match** the data emailed to you, your score on that question will be **zero**.
5. During the exam, I **can only answer** clarifying questions regarding what a question means. I **cannot** help you answer the question.
6. You may email the answers to me, or submit the answers using the button provided under Assignments in Blackboard.
7. Please check your answers carefully before submission. The submission is final.

**Questions**

1(5 points) Please use the file Q1Q2V6.csv to answer this question. This data set provides:

* Sales (move) for three brands light beer: Amstel, Coors and Miller, at a store in a given week.
* The store, week, and Storeweek = 1000\*store + week
* Season (Fall, Winter, Spring, or Summer)
* Brand
* Feat (1 if product is on sale, 0 if not)
* Log of move
* Price
* Log of price
* Demographic variables at the store location: AGE9, AGE60, EDUC, ETHNIC, INCOME, NOCAR, SINGLE, POVERTY, RETIRED, UNEMP, NWHITE

Please answer the following questions using pivot tables. You may submit the answers either as an Excel worksheet, or in Word.

For each season (Fall, Winter, Spring, and Summer), what percentage of time each brand was on sale (Feat = 1)?

2(6+6+4+4 = 20 points) Use the package R and the file Q1Q2V5.csv to answer this question. Please prepare answers as a Word file where the relevant output from R is pasted.

2(a) Fit a regression model with dependent variable log of move and the following independent variables:

* Log of price
* Brand
* Season
* Interaction between Brand and log of price
* Feat
* Demographic variables given in the data set.

Based on this output, what is the price elasticity of demand of each brand?

2(b) To answer each part of 2(b), please start with the model and the results from 2(a).

2(b)(i) From the regression output in 2(a), identify the demographic variables that are not significant at a 90% level of confidence. At a 99% level of confidence, test the null hypothesis that the coefficients of these demographic variables (that is, variables you identified) are all zeros. Provide the relevant information from the output.

2(b)(ii) Starting with the model in 2(a) (that is, using that model as the full model), test the null hypothesis that the price elasticity of demand is same for all three brands at a 99% level of confidence.

2(b)(iii) Starting with the model in 2(a) (that is, using that model as the full model), test the null hypothesis that the price elasticity of demand is same for Coors and Miller.

2(c) Fit a logit model with Feat as the dependent variable, and Brand and Season as independent variables. Very briefly describe how the probability that the product is on sale (Feat = 1) depends on Brand and Season.

2(d) Starting with the logit model in 2(c), test the following null hypotheses at a 99% level of confidence:

2(d)(i) A brand is equally likely to be on sale (Feat = 1) in Fall, Spring and Summer.

2(d)(ii) Season being same, Amstel and Miller are equally likely to be on sale.

3.(4+6 = 10 points) Please use the file Q3V6 to answer this question. This file includes two worksheets, where each worksheet provides, for a brand of beer:

* Store, Week, and Storeweek = 1000\*store + week
* Season (Fall, Winter, Spring, or Summer)
* Brand
* Sales (move)
* Log of move
* Feat (1 if product is on sale, 0 if not)
* Price
* Log of price

3(a) Using Access and the join variable Storeweek, create a file of **same store sales** that gives, store, week, storeweek, season, and, for a given combination of store and week, the move, log of move, price, log of price, and Feat for both brands. (Only include cases where data for both brands are available.)

3(b) Export the data file created in 3(a), save it as csv, and use R to answer the following questions.

For each of the two brands, fit a regression model where the dependent variable is log of move for that brand, and the independent variables are log prices for both brands, Feat for both brands, and Season.

How does the demand of each brand depend on its own price, the price of the other brand, and whether **the other brand** is on sale (Feat =1)?

**Please paste the relevant parts of the R output in your answers and also attach the data file you created as answer to 3(a).**

4.(10 points) Please use the file Q4V6 to answer this question. This file has three worksheets obtained from World Bank data:

* Country: A complete list of country names and country codes.
* Internet: Country code, and the average number of residents out of 100 that had access to the internet in 2015. (The worksheet only includes countries for which the data are available.)
* Service: Country code, and the value added by services as % of GDP in 2015. (The worksheet only includes countries for which the data are available.)

Using Access and country code as the join variable, prepare the data files listed below. Submit the worksheets as answers. (Please note that in some cases, it is possible that the answer is a blank table.)

4(a) A data file that gives country name, country code, internet and service for only the cases where internet is **not missing** and service is **not missing**.

4(b) A data file that gives country name, country code, internet and service for only the cases where internet is **not missing** and service is **missing**.

4(c) A data file that gives country name, country code, internet and service for only the cases where internet is **missing** and service is **not missing**.

4(d) A data file that gives country name, country code, internet and service for only the cases where internet is **missing** and service is **missing**.

5.(5 points) The Excel fine Q5V6 has 150 cases drawn from the Dominicks data base and includes seven fields, ID, Store, Week, Brand, Move, Price and Feat. This worksheet includes one or more duplicates of the same rows. Each combination of the three fields STORE, WEEK and BRAND should be unique. Import the worksheet to Access and copy the table to a new table with structure only. Following the method described in the handout on Access, open the copy in design view, click Indexes under Table Tools, and create a primary key that is the combination of Store, Week and Brand. Save the change. Append the original table to the copy to remove duplicates. Your answer to question 5 is the table you get after removing duplicates.