**Marketing 453-653 Spring 2021 Project**

**(Due, midnight, 5/21/2021, Total Points = 50)**

You can either complete a standard project using data from the Dominicks database, or do an alternative project using survey data collected using Qualtrics. It will be easier to complete the standard project. For both types of projects, I clearly specify requirements. As long as you complete the requirements satisfactorily, you will get full points. Data for both the standard project and the alternative project are posted in the Project folder in Blackboard Content.

**Marketing 453-653 Spring 2021 Standard Project**

In the “Data Sets for Standard Project” in the project folder, data for three product categories from the Dominicks data-base are posted: orange juice, sports drinks, and analgesics.

Please select one of the three product categories (analgesics, orange juice, and sports drinks) posted on Blackboard. The data are acquired from the Dominicks data base, and for each product category, movement data for all stock keeping units, and for only the high movement stock keeping units are provided separately. Once you choose the product category (for example, orange juice, high movement), you have to work with three data sets:

1. **Weekly movement data.** These data include:

* Storeweek: This is 1000\*store + week and, for a given UPC, uniquely identifies each data point.
* UPC
* Store
* Week (1:400, Week 1 started on 9/14/89, week 400 started on 5/8/97)
* Move (number of units sold at a store in a given week)
* Logmove (natural logarithm of Move)
* Price (unit price at the store in the given week)
* Logprice (natural logarithm of price)
* Profit (this is percent contribution margin, that is, ((P – VC)/P) × 100).
* Feat (1 if the product was on sale in the store that week, 0 if not)

1. **UPC and product description.** These include

* UPC
* Brand name
* Description
* Size

1. **Store demographics.** These are census data for the area where a store is located. I suggest you work with the limited store demographics data included in the Excel file DEMOSMALL.xls

The project consists of two parts.

**Part 1. Data Preparation**

Please complete the following items:

1. Choose one of the three product categories for which data are provided (analgesics, sports drinks, orange juice).
2. For the product category you chose, only look at the high movement data file. This file gives store, week, upc, move, price and Feat for the highest selling brands.
3. For this data file, use pivot table to find average prices for the brands for Feat = 0 and Feat = 1, and total counts of cases for which you have data. Select three UPC’s such that two are of the same price category (high or low), and the third is of the other type. For example, you can choose two high priced UPC’s and one low-priced UPC. Make sure package size is same for all three UPC’s. Only choose UPC’s such that you have at least 2000 cases for each UPC in the data.
4. Use filter in Excel or Access to create a new data set that has the three UPC’s selected.
5. Using Access, select a random sample of 4000 from this data set.
6. In this data set (of 4000 cases), create the new variables STOREWEEK, logmove, logprice, and Season. (Some of these may already be there.)
7. Join your data set to the Demo Small data. Create a data set that has Store, Week, STOREWEEK, Season, move, price, logmove, logprice, Feat, and any ten demographic variables you choose.

**Part 2. Data Analysis**

For the dataset you prepared, develop research questions and answer them using statistical methods you learned in the class, such as dummy variable regression and logit. Briefly discuss your findings.

**Examples of research questions:**

* How does the demand for a brand depend on price? What is the price elasticity of demand of a brand?
* Is price elasticity different for different brands?
* How does demand depend on whether the product is on sale (Feat =1)? Is this dependence same for all brands?
* How does the demand for a brand depend on the price of another brand? To do this analysis, you have to prepare a worksheet with same store data using Access.
* What demographic factors affect demand?
* How does price vary across brands?
* How does the proportion of times a brand is on sale vary across brands?
* How does the demand for a brand or a product category vary over time?

**Specific Examples:**

Example 1. Estimate a regression model with logmove as the dependent variable, and the following independent variables: BRAND, Feat, Season, and the demographic variables you selected. Using linear hypothesis test, test:

* Is the price elasticity of demand same for all brands?
* Is the price elasticity of demand same for the lower priced brands?
* Is demand same in all seasons?
* Is demand same in a subset of seasons (such as Summer and Fall)?
* Can you drop a subset of demographic variables?

Interpret the final model. (What are the price elasticities of demand? How does demand vary across seasons? How does demand depend on demographic variables?)

Example 2. Estimate a regression model with price as the dependent variable and independent variables Feat, Season and the demographic variables.

Can you drop of the demographic variables? (Use linear hypothesis test to check if a subset of variables can be dropped.) From the final model, examine how price depends on season and demographics.

Example 3. Estimate a logit model with Feat as the dependent variable, and independent variables BRAND, Season, and the demographic variables.

Identify the demographic variables that are not significant at 90% level (that is, P value of .1 or more). Interpret the final model. (Which brand is most likely to be on sale? Which brand is least likely? How does the probability of sale depend on demographics? How does the probability of sale depend on demographics?)

You are **required to** submit the following:

Data: An Excel file that has three worksheets: (1) The raw data for three UPC’s. (2) The random sample you prepared for data analysis. (3) The table/worksheet of the random numbers.

A 5-10 page Report with the following sections:

(1) Executive Summary: A one-page non technical summary of the data used, research questions, and findings.

(2) A brief summary of how you selected the three brands. This should include the pivot table used to find the three brands.

(3) Summary of data analysis and findings.

**The project will be graded as follows:**

|  |  |
| --- | --- |
| Executive Summary | /10 |
| Data Preparation | /10 |
| Summary of brand selection | /10 |
| Data analysis and findings | /20 |

In every item, you get a score of 90% if you do solid, competent work. A score above 90% requires outstanding work.

**Description of Dominicks Data**

Weekly movement data for three product categories compiled from the Dominicks data set are posted:

1. Refrigerated orange juice 64 oz container (Tropicana, Minute Maid, Florida Gold, Florida’s Natural, HH, Dominicks, and Tree Fresh included)
2. Sports drinks (All Sports, Gatorade and Powerade included)
3. Analgesics (Tylenol, Advil and Motrin included)

For each product category, the following files are included:

* Store Demographics (all data)
* Store Demographics (limited data, description provided later)
* Weekly movement data for high movement items.
* UPC codes and product descriptions for high movement items.

**Data Preparation**

**Inclusion:** Only the following data are included

* Only data for week when movement was greater than zero are included.
* Only data with OK=1 are included.
* Only data for which the store is identified is included.
* Only stores for which demographics are provided are included.

**New Variables:** Four new variables are created

* Logmove: The natural logarithm of movement, that is, number of units sold in a week at a store.
* Logprice: The natural logarithm of price.
* Feat: This is 1 if the product was on sale (SALE is B, C, or S), 0 if not.
* Storeweek: This is equal to 1000\*STORE + WEEK. For a given UPC, this uniquely identifies each data point and can be used as a primary key in Access.

**Dominicks Weekly Movement Data (from Dominicks Data manual)**

1. **UPC:** This is the key to use while merging with upc file.

2. **Price, Quantity and Movement:** DFF will sometimes bundle products (E.g., 3 cans of tomato soup for $2). In such occasion, the ‘qty’ variable will indicate the size of the bundle (E.g., 3), the price will reflect the total price of the bundle (E.g., $2), but move will reflect the number of actual item sold, not the number of bundles.

Hence, to compute total dollar sales, one must do the following calculation:

**Sales = Price \* Move / Qty.**

3. **Profit:** This variable indicates the gross margin in percent that DFF makes on the sale of the UPC. A profit of 25.3 means that DFF makes 25.3 cents on the dollar for each item sold. This yields a cost of goods sold of 74.7 cents.

a) Note however that the wholesale costs in the data do not correspond to replacement cost or the last transaction price. Instead we have the average acquisition cost (AAC) of the items in inventory. This, of course, grates against what economists believe to be the relevant cost for rational decision making.

More precisely, the chain sets retail prices for the next week and also determines AAC at the end of each week, t, according to:

**AACt+1 = (Inventory bought in t) Price paidt + (Inventory, end of t-l-sales t) AACt**

There are two main sources of discrepancy between replacement cost and AAC. The first is the familiar one of sluggish adjustment. A wholesale price cut today only gradually works itself into AAC as old, higher priced inventory is sold off. The second arises from the occasional practice of manufacturers to inform the buyer in advance of an impending temporary price reduction. This permits the buyer to completely deplete inventory and then "overstock" at the lower price. In this case AAC declines precipitously to the lower price and stays there until the large inventory acquired at that price runs off. Thus, the accounting cost shows the low price for some time after the replacement cost has gone back up.

**Source:** Peltzman, Sam, *Prices Rise Faster Than They Fall*, Working Paper No. 142, The University of Chicago

4. **Sales:** This variable indicates whether the product was sold on a promotion that week. A code of 'B' indicates a Bonus Buy, 'C' indicates a Coupon, 'S' indicate a simple price reduction. Unfortunately, this variable is not set by DFF on consistent basis (I.e., if the variable is set it indicates a promotion, if it is not set, there might still be a promotion that week).

5. **OK**: This is a flag set by us to indicate that the data for that week are suspect. We do not use flagged data in our analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | **Description** | Type | Length |
| upc | **UPC number** | Numeric | 8 |
| store | **Store Number** | numeric | 3 |
| week | **Week Number** | Numeric | 3 |
| move | Number of unit sold | Numeric | 8 |
| price | Retail Price | Numeric | 8 |
| qty | Number of item bundled together | Numeric | 3 |
| profit | Gross margin | Numeric | 8 |
| sale | Sale code (B,C,S) | Character | 8 |
| ok | 1 for valid data, 0 for trash | Numeric | 3 |

**Store Demographics (from Dominicks Manual)**

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| AGE9 | % Population under age 9 |
| AGE60 | % Population over age 60 |
| ETHNIC | % Blacks & Hispanics |
| EDUC | % College Graduates |
| INCOME | Log of Median Income |
| HSIZEAV | Average Household Size |
| HH3PLUS | % of Households with 3 or more persons |
| HH4PLUS | % of households with 4 or more persons |
| HHSINGLE | % of households with1 person |
| HHLARGE | % of households with 5 or more persons |
| HVAL150 | % of households with Value over $150,000 |
| HVAL200 | % of households with Value over $200,00 |
| MORTGAGE | % of households with mortgages |
| NOCAR | % with No Vehicles |
| NWHITE | % of population that is non-white |
| POVERTY | % of population with income under $15,000 |
| RETIRED | % of Retired |
| SINGLE | % of Singles |
| UNEMP | % of Unemployed |
| WORKWOM | % of Working Women with full-time jobs |
| SSTRDIST | Distance to nearest warehouse store |
| SSTRVOL | Ratio of sales of this store to that of nearest warehouse store |
| CPDIST5 | Average distance in miles to the nearest 5 supermarkets |
| CPWVOL5 | Ratio of sales of this store to the average of the nearest five stores |

**Marketing 453-653 Spring 2020 Alternative Project**

Two data sets are posted on Blackboard:

1. **Smartwatch Survey Data**
2. **Sports Drinks Survey Data**

For the Smartwatch Survey and the Sports Drinks Survey, the following types of data are collected:

1. Basic demographics
2. Ownership/usage of products
3. Importance of attributes
4. Conjoint data using hypothetical products and four attributes

Select one of the three data sets, analyze it and prepare a report. The report (about ten pages including appendices) should include the following:

1. Executive summary: This is a one-page non-technical summary of what you did and what you found. Write it as if you conducted the entire study including data collection.
2. Brief overview of the survey.
3. Aggregate analysis: In an appendix, present the frequencies of all variables and the means and standard deviations of the quantitative variables. (You can omit the conjoint data here.) In the body of the paper, write a short description of the main findings.
4. Additional analysis: Develop additional research questions based on the data. You as a group have to decide precisely what you want to analyze. Two examples are given below:

* From the data on importance of attributes, conduct principal components analysis to identify the key dimensions and interpret the dimensions. Use cluster analysis of the factor scores to identify market segments. Describe the segments in terms of demographics, product ownership, etc.
* Using the conjoint estimates, use cluster analysis to identify market segments. Describe the segments you identifies. Develop new products and assess how they will perform in the market.

**The project will be graded as follows:**

|  |  |
| --- | --- |
| Executive Summary | /10 |
| Overview of Survey | /10 |
| Aggregate Analysis | /10 |
| Additional Analysis | /20 |
| Total | /50 |

In every item, you get a score of 90% if you do solid, competent work. A score above 90% requires outstanding work.