## BUS256 Marketing Analytics Spring 2017

## Rafael Martins Guimaraes

### Exam #2: Tuesday March 28 2017

**Instructions:**  Please re-save this document including your name within the file name. Then type your answers directly into this document, which you will eventually upload to LATTE. You have the entire class session for this exam, and your work must be your own.

**I. Goodbelly case analysis continued (40 pts)**

**For this one question, you will start by downloading one script and csv file from the Github BUS256 folder called “Exam2 data and r code”. Do your work within R Studio.**

We worked with most of this code in class, and in this part of the test, you should slightly modify the code and interpret the results that the script generates. I’ve removed the stepwise regression code because we didn’t work with it very much.

Before starting, go to line 7 of the code and be sure your directory path is appropriate.   
ALSO: go to line 12 and set a seed of your choosing.

Recall that the dependent variable is unit sales, which is the mean number of units sold per week over a 10-week period.

Some stores had demonstrations in the current week and also had demos 1-3 weeks prior. There might be reason to think that repeated demos could really solidify interest in the product, so that a stores that repeated the demos would magnify the impact of demonstrations.

Modify the model specification (line 26 and following) to include an interaction term to capture possible interaction between Demo and Demo1.3. Your main goal in the problem is to interpret a multiple regression model and respond to these questions:

a. Based on your training data, please report the coefficients of Demo, Demo1.3, and the interaction of Demo & Demo1.3:

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 287.69 25.24 11.40 < 2e-16 \*\*\*

Demo 109.90 16.00 6.87 1.2e-11 \*\*\*

Demo1.3 87.26 10.68 8.17 1.0e-15 \*\*\*

Demo:Demo1.3 30.54 16.69 1.83 0.068 .

b. Which, if any, of the three are statistically significant at a level of .10 or lower?

Demo and Demo 1.3 are highly significant for the regression model because of the p value lower than 0.05. The interaction between Demo & Demo 1.3 may be significant to each other, but this interaction is not significant to the dependent variable of the model.

c. Explain the magnitude of main and interaction effects of demonstrations on unit sales, taking into account only demonstration in the current or prior three weeks.

We can see that the Demo is highly significant to the model because on the higher increase on unit sold than Demo 1.3. This means that people are more affected by it when they are exposed to it prior the release than weeks before. The interaction between them, as said before, is not significant to the because of its higher p value.

**Paste in the relevant lines of your final code here:**

# ---------------------------------------------------------------------------

# Model

#---------------------------------------------------------------------------

model <- (Units.Sold ~ Average.Retail.Price + Sales.Rep +

Endcap + Demo + Demo1.3 + Demo4.5 +

Sales.Rep \* Endcap +Sales.Rep \* Demo +

Sales.Rep \* Demo1.3 + Sales.Rep\*Average.Retail.Price + Demo\*Demo1.3 # interaction

)

fit <- lm(model, data=train) # estimate the model

summary(fit) # report regression results

**Paste in the relevant final regression output here (that is, the output you used to draw your conclusions:**

**II. Market Basket. (30 pts)**

Miller’s original presentation of market basket rules focused on vegetables. I’ve modified that analysis to develop a set of rules related to the purchase of **fruits**. Here is the set of “top 10” rules, sorted by lift:

> inspect(top.fruit.rules)

lhs rhs support confidence lift

[1] {dairy produce,

vegetables} => {fruit} 0.079 0.46 1.9

[2] {bread and backed goods,

vegetables} => {fruit} 0.051 0.44 1.8

[3] {dairy produce,

non-alc. drinks} => {fruit} 0.064 0.42 1.7

[4] {bread and backed goods,

dairy produce} => {fruit} 0.077 0.41 1.7

[5] {cheese} => {fruit} 0.051 0.41 1.6

[6] {vegetables} => {fruit} 0.107 0.39 1.6

[7] {dairy produce} => {fruit} 0.156 0.35 1.4

[8] {sausage} => {fruit} 0.065 0.34 1.4

[9] {bread and backed goods} => {fruit} 0.108 0.31 1.2

[10] {non-alc. drinks} => {fruit} 0.094 0.29 1.2

1. Within this list, use plain language to “translate” and explain the rule that applies most frequently to customer purchases:

The objective of this analysis is to find what kind of product goes together, providing information to guide the product placement in stores. Association rule model finds interesting associations and correlation relationships among large sets of data. Association rules show attribute value conditions that occur frequently together in a given data set, it ask, “what goes with what?". Is the division of each item set into to 2 subsets, with one subset called the antecedent, thought of as preceding the other subset, called the consequence. This will help you to optimize the cross selling and bundling for example.

b. Within the list, the first rule has the largest confidence as well as the highest lift. Using the example of the items in that rule, explain what confidence and lift mean.

Confidence estimates the consequence given the antecedent. In the model, the support Determines the proportion of time it occurs in the data set. A support criteria of 0.01 implies that one in every one hundred market basket must contain the item set . Consequence values go to zero to one. A higher value in confidence is what you want to achieve.

Lift, is the number of transactions that include the consequent divided by the total number of transactions. A lift ratio larger than 1.0 implies that the relationship between the antecedent and the consequent is more significant than would be expected if the two sets were independent. The larger the lift ratio, the more significant the association.

1. Of the ten rules, which one do you think might be *least* useful to managers in a grocery store? Explain your thinking.

[10] {non-alc. drinks} => {fruit} 0.094 0.29 1.2

Because of the low level of confidence. A higher value in confidence is what you want to achieve, you want to know the reliability of the consequence, so drinks with fruits have the lowest confidence level.

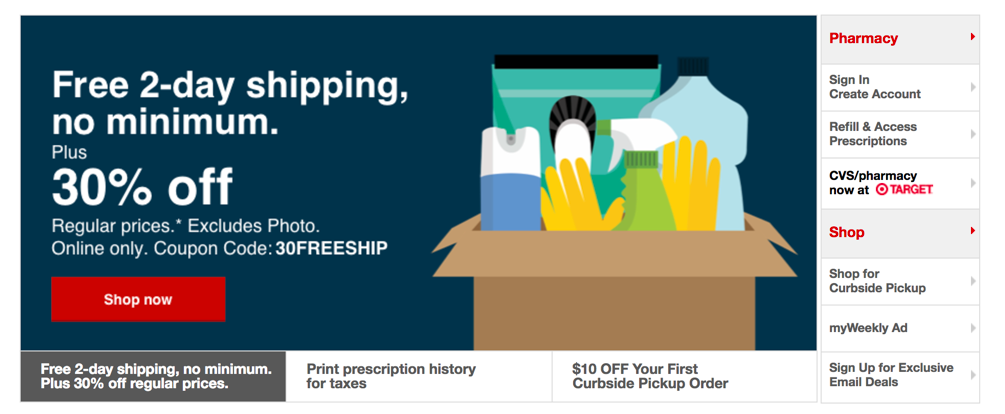
**III. Experimental Design (30 pts)**

I have mentioned in class that CVS has been a leader in using experimental design to gather insight into consumer behavior. Navigate to [www.cvs.com](http://www.cvs.com) now, and examine their main landing page. Find at least *four* areas or items on the web page where you think they might be conducting an on-line experiment right now. In other words, identify four specific places on the screen where they might well be showing different information or images to different viewers.

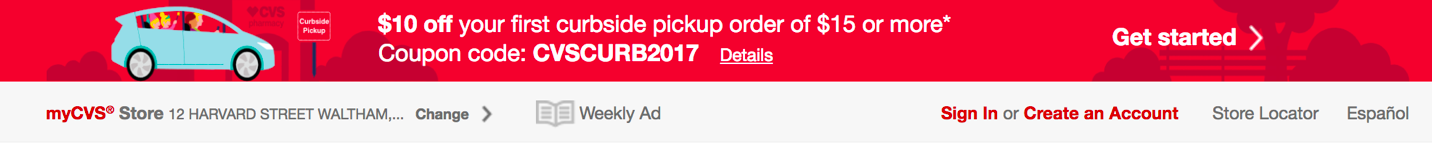
Explain your decision to choose those places, and explain how they could use the data gathered in connection with other available data to gain deeper understanding of consumer behavior.

After accessing the main webpage of CVS, I have Identify four main places where they migh have conduction a experiment.

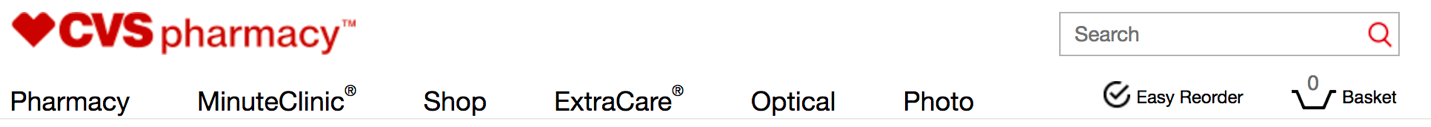
The first, and easiest, is the promotion part below the main menu. The possibility of doing that is high because is the biggest part of the screen. The kind of experiment they my be conducting is a promotion part. In this part of the web page we can see different promotion, and they might want to see which one is the most efficient. They can do that by conducting A/B testing with random users. This is interesting because you can see if price or value is important in an add and measure it by clicks and pageviews.



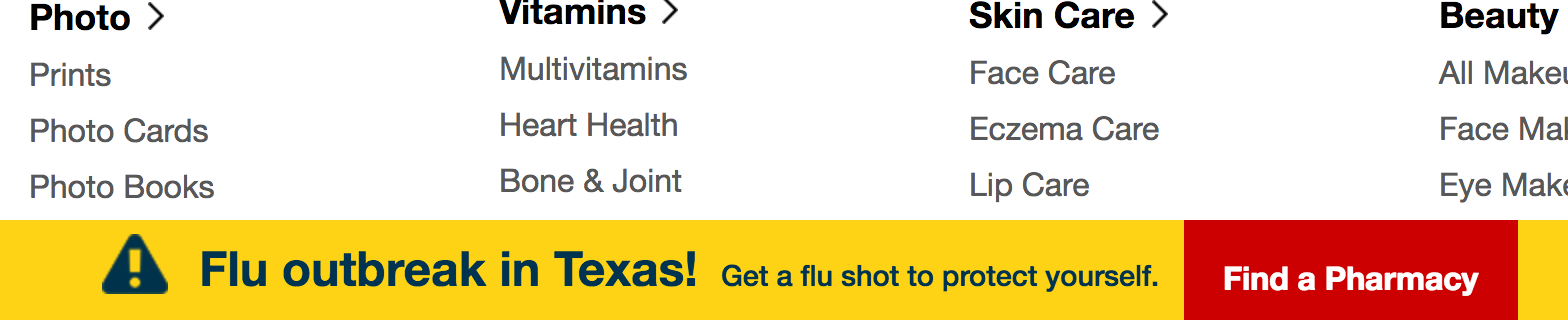
The second place where they might conduct a experiment is in the header in the website. As we can see, the header (the top with the main menu), has an add on it. It is interesting place to put a add, especially if we consider that people start reading the website on the top left, also called F-pattern, by Jakob Nielsein, in his book “Designing Web Usability”. (this image illustrates the F-pattern -> [F\_Pattern](https://media.nngroup.com/media/editor/alertbox/f_reading_pattern_eyetracking.jpg)).



The third place might be the main menu. As we can see there are to option related to services, MinuteClinic and ExtraCare. CVS is know to provide a lot of extras services for their clients, and they might have been conducting an experiment in this area to see if people will use the menu option, the search option. It can also help you to understand how users are navigating in your website



The final part, is maybe the add on the end of the page. Right now it is showing a add for a flu shot in Texas, and they might be doing experiments on it to see how effective it is. Although we usually think that adds or alerts messages are not effective on the end, maybe with a different color pattern it might be efficient and to see how efficient it is, we just need to see the total of clicks and pageviews.



All in all, it is important to do try and error analysis in the company’s website, mostly because what you usually want to achieve is that the client finish his objective in your website as soon as possible, and understand what kind of users you have. You may have people that only use the content and other just to purchase or eve as a reference. That is why experiments are so important, they give you data of what you are doing right and wrong in a real-time situation.