DSO 510: Homework 2

Due: Sunday, November 21 @ midnight

General Instructions

You must submit 1 word/pdf document with your screenshots and answers. Even if you have answered the questions in the statistical software (Excel, R, Python, etc.) you are using, you should summarize your answers in the document. You should also submit the file where you conducted your statistical analysis (Excel, R, Python, etc.)

PART 1

Consider yourself working for a global retailer that over the years has added a web-based channel to their physical store locations. Now, after learning more about mobile-led changes in retailing, they are excited about what the mobile ecosystem offers. They are seeking your help as they embark on using mobile as a channel. They want to commission an app development team to deploy a presence on iOS and Android. However, several questions arise about the deployment of the app. Your job is to provide data driven insights to help them navigate this complex landscape.

Specifically, you are tasked with:

1. Using the data, estimate a linear model for the relationship between demand and price. For this you have access to a large volume of app level data (in a file called **hw2_1.csv**), including information about the 'rank' of the app on the app store. Assume Sales = (1/rank)*1,000,000 (don't worry about the details behind this assumption, just make the assumption). Specifically, estimate a univariate regression where the dependent variable is sales and the independent variable is price: $Sales = \beta_0 + \beta_1 * Price$
   a. Report the estimated intercept and the estimated slope coefficient.
   b. Test the following null hypothesis: $\beta_1 = 0$. Use a 5% significance level. Provide an explanation of your answer.

2. Create a dummy/binary variable for region. This variable should have a value of 0 if the region is CN (China) and 1 if the region is US (USA). Estimate a univariate regression of sales on this newly created variable. Provide a screenshot and an **interpretation** of both estimated coefficients. Be specific.

3. Create another dummy/binary variable for in app advertisements (in_app_ads). This variable should have a value of 1 if the device has in app advertising and a value of 0 if the device does NOT have in app advertising. Estimate a regression of sales on the dummy variable created in part 2 and this newly created dummy variable (all in the same model). Provide a screenshot of the results and provide an interpretation of **all** the coefficients. Be Specific.

4. Estimate a univariate regression of sales on rank (similar to part 1) except in this case your model should able to speak in terms of elasticity. By elasticity you want to speak to your management in percentage terms – what is the % change in sales for a % increase in price? (Tip: we do this using log-log-regression models.) Since price can have a value of 0, you will have to adjust the variable. You can do this by adding 1 to each price and then taking the log. Provide a screenshot of the results and provide an **interpretation** for all the coefficients. Be specific. (https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-do-i-interpret-a-regression-model-when-some-variables-are-log-transformed/) .

5. The app retailer believes that other factors, specifically the filesize, the number of screenshots, and the average rating may also be associated with both sales and price. The retailers want a model that estimates the relationship between price and sales (similar to 4) except they want the impact of the above-mentioned factors to be controlled for. Estimate a model that accomplishes this. Your model should speak in terms of elasticity (same as part 4). Provide screenshots of your results and discuss how this model achieves what the retailers want. Provide an interpretation of **all** the estimated coefficients.

6. The retailer is also interested in understanding the impact of the in-app purchase option. Specifically, the retailer believes that the relationship between price and sales is different for apps with an in-app purchase option and apps without an in-app purchase option. To do this, estimate the same model that you estimated in part 5 except add an interaction term between price and in app purchase option (dummy variable). Provide the results and an interpretation of **all** the estimated coefficients. Be specific.

PART 2

Exercise 1. [2 points]

You are interested in examining whether visitors to your website spend, on average, more than 12 minutes browsing the website. While you had not previously kept track of website visitors, you start tracking after deciding that you want this information. In the file ***exercise_1.csv***, you will find the minutes spent browsing for a sample of 24 website visitors. Use this data to statistically evaluate whether, on average, website visitors spend **more** than 12 minutes browsing on your website. Be specific about your approach (set your alpha-level at 0.05). State the null/alternative hypothesis and your conclusion.

Exercise 2. [8 points]

An online retailer has implemented a new website design which they believe will increase sales. This includes digital features such as more product pictures and a zoom feature so that consumers can better examine the product before purchasing it. For a short period of time, the company randomly sends 10% of website traffic to the newly designed website, while the remaining 90% are sent to the original website. A sample of visitor behavior is available in the file ***exercise_2.csv***. The table below provides a description of the data in this file.

Data Description

website_design	1: if the visit corresponds to a visit to the new website design. 0: if the visit corresponds to a visit to the original website design.
member	1: if the visitor has signed up for membership. 0: if the visitor has not signed up for membership. The membership program is designed for frequent users to store their credit card, shipping, and other preferences for an expediated experience.
sale_1_0	1: if the visitor made any purchase on this visit. 0: if the visitor did not make any purchases on this visit.
return_1_0	1: if an item purchased during this visit was eventually returned. 0: if no items related to this visit were returned.
minutes_spent	The total minutes spent browsing the website during this visit.

a. Evaluate whether the new website design visits are statistically more likely to end in a sale than the visits to the original website design. You do not need to include other variables in your model since the customers were randomly assigned. Be specific about your approach (set your alpha-level at 0.05). State the null/alternative hypothesis and your conclusion.

b. Examine whether there is a statistical difference between the mean of *minutes_spent* for the subset of consumers that were sent to the new website design and the subset that were sent to the original website design. You do not need to include other variables in your model since the customers were randomly assigned. Be specific about your approach (set you alpha-level at 0.05). State the null/alternative hypothesis and your conclusion.

c. There is concern that there may have been a programming error regarding the random assignment of consumers. Specifically, it may be that the selection of the 10% of customer traffic that was directed to the newly designed website was not random. Does the data suggest that this concern is legitimate? Even if it was not random, does the data suggest that our conclusions about the new website design observed in **part a** should change? Be specific about your approach (set you alpha-level at 0.05). State the null/alternative hypothesis and your conclusion.

d. The company is worried that the introduction of the new features (more product pictures and zoom feature) is having an impact on customers returning the products they purchased. Statistically examine the impact of the new website design on returns.