

DSO 510: Homework 1

Due: Sunday, November 7 @ 11:59 PM

[10 Points]

Instructions: Please Read Carefully.

Accompanying this homework you will find a csv file named hw1.csv.

This file contains sales data for a grocery store chain operating in California. Each row corresponds to the monthly in-store sales of each store. You are provided monthly data for each store from January 2015 – December 2016.

The grocery store partnered with a company that offered online delivery services starting in January 2016. This means you have access to data before and after the online delivery service was introduced.

For a fee, customers could order their groceries online and have the 3rd party company deliver the groceries. All deliveries had to be made to addresses in the same zipcode as the grocery store. The fee was not insignificant. The online delivery service was not offered at all stores. The data is provided at the zipcode level and each zipcode has only one physical store. We do not have data on online sales, but only in-store sales. Importantly, in-store sales do not include sales made through the online delivery service. Below is a description of the columns in the data set:

Zipcode	The zipcode in which the store resides.
Date	The month/year.
Online Available	Variable indicating whether the store was part of the subset of stores that offered online services. 1: offered online delivery ; 0: did not offer online delivery.
Pharmacy	Variable indicating whether the store has a pharmacy. 1: the store has a pharmacy ; 0: the store does NOT have a pharmacy
In Store Sales	The sales made in the grocery store for the month. These are only sales made in the store. They DO NOT include online delivery sales.
Median Income	The median income of the zipcode the store is in. Does not change monthly.
Population	The population of the zipcode the store is in. Does not change monthly.

Answer the following questions regarding the data with visualizations or descriptive statistics. Do **NOT** use t-tests or regressions. You may use any software you want. Please submit a document that answers the following questions. This should include the visualizations (i.e., screenshots) in a manner that clearly indicates the question associated with each visualization. Please also submit all accompanying files (i.e., Tableau file, Excel file, R code, etc.)

- 1) Does the online delivery option appear to have an impact on the sales made in the store? If your answer is yes, provide evidence that this is not simply a result of less grocery store demand in California in 2016. **[2 points]**
- 2) In Part 1, you examined the relationship between grocery stores introducing an online delivery option and the in-person sales made by grocery stores. Now you want to understand whether that relationship is the same for stores that have a pharmacy and those that don't. Specifically, does the introduction of the online delivery option appear to have the same impact on grocery stores with a pharmacy and those without a pharmacy. Provide visualization(s) or descriptive statistics that support your answer. Also provide a rationale for why you think that may be the case. Do not just explain the visualization. Provide your logical assessment of why you think the result is the way it is. **[2 points]**
- 3) Repeat question 2, except now you are interested in the median income of the zipcode. Specifically, does the introduction of the online delivery option appear to have the same impact on grocery stores in areas where residents have high/low median income. Provide visualization(s) or descriptive statistics that support your answer. Also provide a rationale for why you think that may be the case. Do not just explain the visualization. Provide your logical assessment of why you think the result is the way it is. **[1.5 points]**
- 4) Repeat question 3, except focus on the zipcode population. **[1.5 points]**
- 5) Using the data provided, assess the selection of the stores that were chosen for the online delivery option. Think about this in the context of geographic location (southern California, northern California, etc.) and location attributes (income, population, etc.). Does the data suggest that the stores that had the online delivery option were selected randomly? Provide visual evidence to support your claim. Regardless of whether the selection of stores for the online option appears random, discuss the advantages of a scenario where the stores with the online delivery option are randomly selected. **[2 points]**
- 6) The sales that were made through the online 3rd party are not available in the provided data. Thus far we have analyzed the relationship between introducing an online delivery channel and the in-store sales of a grocery store. The grocery store is thinking about making the online option available at all stores. Discuss why the data limitation mentioned (a lack of online delivery option sales data) limits what we can conclude about the **overall** effect of a grocery store introducing an online delivery option. **[1 point]**