Question Set

Question One

> *(a) In your own words, explain how the product information was recorded, and the measures that researchers took to ensure that the data was accurate and representative of the treatment group. What were some of the data collection issues that they encountered?*

Product information was initially collected using both tablet computers and paper forms. In March 2016, they then switched to just paper forms due to the inadequate speed of digital collection. Once collected, paper forms were inputted into a database of all variables measured including both pre-tax and post-tax data. This was then '*double-entered by trained research project assistants*' to minimise the possibility of human error and increase data accuracy.

To improve representation, the panel included less known brands as stock differed between stores which were substitutes to more well known brands in other stores. Consequently, data for every beverage was not collected for every store.
Drinks such as Water and Milk were collected in addition to sugary beverages to act as a control group to locate trends in the data set.

On top of accommodating for different beverages, the researchers chose a number of different types of stores and locations to ensure the whole bay area was represented. This included large and small chain supermarkets, chain petrol stations, pharmacies, and independent corner stores and petrol stations. These stores also differed in location serving commercial neighborhoods with businesses and BAME communities also.

> *(b) Instead of using the name of the store, each store was given a unique ID number (recorded as store_id on the spreadsheet). Using Excel's filter function, verify that the number of stores in the dataset is the same as that stated in the 'S1 Text' (26). Similarly, each product was given a unique ID number (product_id). How many different products are in the dataset?*

Using the `COUNT(UNIQUE(range))` formula we can find all distinct values for both stores and products. As shown in analysis sheet: the final result is `store-id = 26` and `product-id = 247`.

Question Two

> *A frequency table showing the number (count) of store observations (store type) in December 2014 and June 2015, with 'store type' as the row variable and 'time period' as the column variable. For each store type, is the number of observations similar in each time period?*

Question 2.a

Count of time	Column Labels		
Row Labels	DEC2014	JUN2015	Grand Total
1	177	209	386
2	407	391	798
3	87	102	189
4	73	96	169
Grand Total	**744**	**798**	**1542**