Homework #3: Causality

Raveena Kamal

Overview & instructions

For homework, you will first analyze a panel dataset from a new product category. Then you will evaluate the effectiveness of an online display ad campaign.

Instructions

1. Your task is to fill in all R code blocks that currently contain "#TBD" comments. Similarly, insert text responses wherever you see *TBD* in the markdown file.
2. PLEASE ADD YOUR NAME TO THE AUTHOR LINE ABOVE

Task 1 description

Your task is to construct a linear model of demand for a common packaged good (a laundry detergent). The effect of interest is the average effect of `price` on `sales` (units sold).

You have access to scanner data across a set of stores of a retail chain in the Chicago metro region. The data are in the file `detergent_data.csv` here. The variables in the data set are:

`store`	Store id number
`week`	Week number
`promoflag`	= 1 if any product in the category was on promotion
`sales`	Tide 128oz laundry detergent: unit sales
`price`	Tide 128oz laundry detergent: price ($)

1) Data Description

1.1) Read in the data

Read in the data from the csv file, and store to dataframe `DF1`.

Also, report (print):

1) the number of unique stores in the dataset, and
2) number of unique time periods in the dataset

Hint: the functions `unique()` and `length()` can be useful in calculating the number of unique observations

```
DF1<-read.csv("/Users/raveena/Desktop/Classroom - R/Marketing Analytics/data/detergent_data.csv")

length(unique(DF1$store))
```

```
## [1] 86
```

```r
length(unique(DF1$week))
```

```
## [1] 224
```

What is the total number of observations? 14744

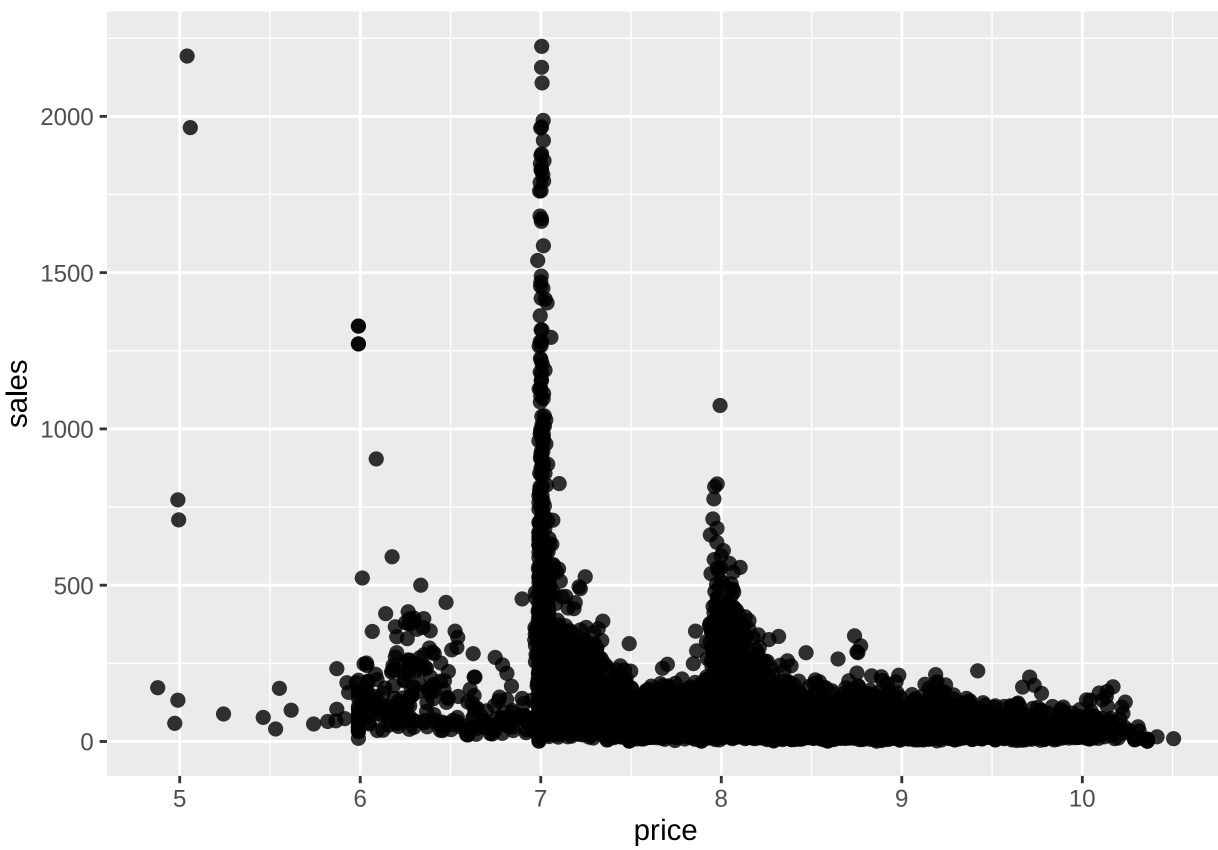
Is the data a balanced or unbalanced panel? Why? The data is unbalanced panel. This is beacaue each store should have same number of time periods. If we group the data by store and count the weeks for each store it should be same for all the stores for the data to be a balanced panel.

Interpret the mean value of the variable `promoflag` Mean is 0.81 which means that 81% of the panel dataset have used promotions.

1.2) Scatterplot of `sales` vs. `price`

Generate a scatterplot of sales vs. price. Add a linear regression line.

```r
library(ggplot2)
ggplot(data = DF1, aes(x = price, y = sales)) +
  geom_point(alpha = .8, size = 2)
```



Comment on the distribution of sales and prices. What patterns do you notice? The sales is high for some price points like 7 and then falls. It shows a fluctuation where sales is increasing with increase in price (between $7-$8) and then decreasing as price further increases.

Is the (sign of the) fitted regression line slope as expected? Why or why not? Yes. The sign is -ve which is expected from the model. It does show that as price increases unit sold decreases.

2) Regresion models where DV = sales

2.1) regressors: `price`, `promoflag`

Use `lm()` to estimate a model of sales with regressors: `price`, `promoflag`. Use `summary()` to summarize the results.

```
summary( lm(sales ~ price + promoflag, data=DF1))
```

```
##
## Call:
## lm(formula = sales ~ price + promoflag, data = DF1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -259.94  -47.25  -15.51   19.36 2045.53
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  636.795     11.573  55.024   <2e-16 ***
## price        -68.622      1.336 -51.347   <2e-16 ***
## promoflag     22.379      2.635   8.492   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 123 on 14741 degrees of freedom
## Multiple R-squared:  0.1588, Adjusted R-squared:  0.1587
## F-statistic:  1391 on 2 and 14741 DF,  p-value: < 2.2e-16
```

2.2) regressors: `price`, `promoflag`, time (week) fixed effects

Use `lm()` to estimate a model of sales with regressors: `price`, `promoflag` and time (week) fixed effects. Use `summary()` to summarize the results.

```
summary( lm(sales ~ price + promoflag + week, data=DF1))
```

```
##
## Call:
## lm(formula = sales ~ price + promoflag + week, data = DF1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -262.39  -47.39  -15.66   19.46 2044.02
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 633.94646   11.73102  54.040  < 2e-16 ***
## price       -68.54153    1.33747 -51.247  < 2e-16 ***
## promoflag    21.59894    2.68730   8.037 9.87e-16 ***
## week          0.02843    0.01918   1.482    0.138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 123 on 14740 degrees of freedom
## Multiple R-squared:  0.1589, Adjusted R-squared:  0.1587
## F-statistic: 928.2 on 3 and 14740 DF,  p-value: < 2.2e-16
```

```r
summary( lm(sales ~ price + promoflag + factor(week), data=DF1))
```

```
##
## Call:
## lm(formula = sales ~ price + promoflag + factor(week), data = DF1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1191.94   -18.41    -3.96    12.65  1878.07
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      179.2997    12.7714  14.039  < 2e-16 ***
## price            -17.3808     1.2545 -13.854  < 2e-16 ***
## promoflag         -7.2057     3.0677  -2.349 0.018845 *
## factor(week)2     -4.2591    10.8031  -0.394 0.693406
## factor(week)3      7.1346    10.7676   0.663 0.507599
## factor(week)4    105.6222    10.6652   9.903  < 2e-16 ***
## factor(week)5     67.2293    10.2984   6.528 6.88e-11 ***
## factor(week)6     21.1420    10.3973   2.033 0.042030 *
## factor(week)7     30.3143    10.4385   2.904 0.003689 **
## factor(week)8     31.5211    10.2880   3.064 0.002189 **
## factor(week)9     22.2312    10.1885   2.182 0.029127 *
## factor(week)10    24.1935    10.7717   2.246 0.024718 *
## factor(week)11     5.3013    10.9161   0.486 0.627227
## factor(week)12    23.7954    10.6671   2.231 0.025715 *
## factor(week)13    33.6070    10.4728   3.209 0.001335 **
## factor(week)14    52.7934    10.2992   5.126 3.00e-07 ***
## factor(week)15    48.3027    10.2661   4.705 2.56e-06 ***
## factor(week)16    63.8106    10.3656   6.156 7.66e-10 ***
## factor(week)17   368.5436    10.2499  35.956  < 2e-16 ***
## factor(week)18    21.7282    10.7363   2.024 0.043010 *
## factor(week)19    48.4430    10.7443   4.509 6.57e-06 ***
## factor(week)20   333.7572    10.5487  31.640  < 2e-16 ***
## factor(week)21    44.5825    10.8911   4.093 4.27e-05 ***
## factor(week)22    18.5942    10.8013   1.721 0.085185 .
## factor(week)23    22.5426    10.8458   2.078 0.037685 *
## factor(week)24    16.7553    11.1806   1.499 0.134000
## factor(week)25    20.5327    11.3084   1.816 0.069437 .
## factor(week)26     7.4717    11.3090   0.661 0.508823
```

```
## factor(week)27    13.3017    11.6735   1.139 0.254522
## factor(week)28   126.9142    11.3049  11.226  < 2e-16 ***
## factor(week)29    57.6310    11.5619   4.985 6.28e-07 ***
## factor(week)30    25.4194    11.8501   2.145 0.031962 *
## factor(week)31    27.3902    11.5488   2.372 0.017720 *
## factor(week)32    32.7177    11.3413   2.885 0.003922 **
## factor(week)33    21.5342    11.1378   1.933 0.053201 .
## factor(week)34    19.9710    11.2802   1.770 0.076673 .
## factor(week)35    23.5177    11.2427   2.092 0.036473 *
## factor(week)36    42.7294    11.0479   3.868 0.000110 ***
## factor(week)37    44.6859    10.2889   4.343 1.41e-05 ***
## factor(week)38    44.1526    10.2205   4.320 1.57e-05 ***
## factor(week)39    35.9987    10.2099   3.526 0.000423 ***
## factor(week)40    40.3781    10.3262   3.910 9.26e-05 ***
## factor(week)41    37.9819    10.2181   3.717 0.000202 ***
## factor(week)42    36.8931    10.1085   3.650 0.000263 ***
## factor(week)43    31.6829    10.0770   3.144 0.001669 **
## factor(week)44    12.7436    10.5854   1.204 0.228653
## factor(week)45    21.4324    10.1278   2.116 0.034346 *
## factor(week)46    20.7015    10.1558   2.038 0.041529 *
## factor(week)47    20.5786    10.2146   2.015 0.043962 *
## factor(week)48     5.5088    10.6186   0.519 0.603914
## factor(week)49     8.8089    10.6693   0.826 0.409030
## factor(week)50    14.7092    10.0570   1.463 0.143602
## factor(week)51    19.0112    10.0468   1.892 0.058476 .
## factor(week)52    14.0450    10.0528   1.397 0.162398
## factor(week)53    24.2244    10.0297   2.415 0.015736 *
## factor(week)54   188.1124    10.0515  18.715  < 2e-16 ***
## factor(week)55    38.3523    10.0777   3.806 0.000142 ***
## factor(week)56    -1.1565    10.0017  -0.116 0.907944
## factor(week)57    -4.7928    10.4569  -0.458 0.646718
## factor(week)58     2.4062    10.1506   0.237 0.812619
## factor(week)59     3.9848    10.3925   0.383 0.701408
## factor(week)60    23.2274    10.0277   2.316 0.020554 *
## factor(week)61   121.1052    10.0801  12.014  < 2e-16 ***
## factor(week)62    44.5783    10.0692   4.427 9.62e-06 ***
## factor(week)63     5.3834     9.9838   0.539 0.589745
## factor(week)64     5.8605    10.4604   0.560 0.575315
## factor(week)65     1.2656     9.9762   0.127 0.899050
## factor(week)66    13.1524     9.9120   1.327 0.184556
## factor(week)67    31.8729     9.9468   3.204 0.001357 **
## factor(week)68    44.3713    10.0841   4.400 1.09e-05 ***
## factor(week)69    45.7043    10.0262   4.559 5.19e-06 ***
## factor(week)70    41.1927    10.0887   4.083 4.47e-05 ***
## factor(week)71    41.4639     9.9849   4.153 3.30e-05 ***
## factor(week)72    58.9130    10.0208   5.879 4.22e-09 ***
## factor(week)73   303.8825    10.0267  30.307  < 2e-16 ***
## factor(week)74    40.2134     9.9938   4.024 5.76e-05 ***
## factor(week)75    33.8421    10.0237   3.376 0.000737 ***
## factor(week)76    38.3706    10.0230   3.828 0.000130 ***
## factor(week)77    35.4360    10.0218   3.536 0.000408 ***
## factor(week)78    54.3113    10.0846   5.386 7.33e-08 ***
## factor(week)79   171.1303    10.1337  16.887  < 2e-16 ***
## factor(week)80    49.6456    10.1263   4.903 9.56e-07 ***
```

```
## factor(week)81    55.1670    10.0225   5.504 3.77e-08 ***
## factor(week)82    50.9635    10.0985   5.047 4.55e-07 ***
## factor(week)83    48.9224    10.0128   4.886 1.04e-06 ***
## factor(week)84   293.9316    10.0249  29.320  < 2e-16 ***
## factor(week)85    46.1738    10.0083   4.614 3.99e-06 ***
## factor(week)86    53.7184     9.9987   5.373 7.88e-08 ***
## factor(week)87    62.5244    10.0427   6.226 4.92e-10 ***
## factor(week)88   103.7413    10.0139  10.360  < 2e-16 ***
## factor(week)89    48.6748    10.1971   4.773 1.83e-06 ***
## factor(week)90    56.5608    10.0537   5.626 1.88e-08 ***
## factor(week)91    31.4695    10.0184   3.141 0.001686 **
## factor(week)92    37.5773     9.9300   3.784 0.000155 ***
## factor(week)93    34.4123     9.9792   3.448 0.000566 ***
## factor(week)94    36.4961     9.9851   3.655 0.000258 ***
## factor(week)95    24.3166    10.0267   2.425 0.015312 *
## factor(week)96    20.3438    10.0312   2.028 0.042574 *
## factor(week)97    20.4977    10.0303   2.044 0.041014 *
## factor(week)98    15.7109    10.4708   1.500 0.133519
## factor(week)99    15.9032    10.4724   1.519 0.128890
## factor(week)100   16.7480     9.9992   1.675 0.093968 .
## factor(week)101   15.4032    10.5299   1.463 0.143542
## factor(week)102   21.1013    10.0264   2.105 0.035345 *
## factor(week)103   17.9604    10.5010   1.710 0.087222 .
## factor(week)104   23.0040    10.0535   2.288 0.022143 *
## factor(week)105  236.1183     9.9571  23.714  < 2e-16 ***
## factor(week)106  163.0363     9.9672  16.357  < 2e-16 ***
## factor(week)107   32.9097     9.8963   3.325 0.000885 ***
## factor(week)108   47.8498    10.2052   4.689 2.77e-06 ***
## factor(week)109   23.7042     9.9779   2.376 0.017530 *
## factor(week)110   24.2984    10.3804   2.341 0.019257 *
## factor(week)111   39.7323     9.9097   4.009 6.12e-05 ***
## factor(week)112   25.5935     9.9247   2.579 0.009925 **
## factor(week)113  221.0864     9.9092  22.311  < 2e-16 ***
## factor(week)114  119.2756     9.9143  12.031  < 2e-16 ***
## factor(week)115   39.7831     9.9662   3.992 6.59e-05 ***
## factor(week)116   24.4513     9.9390   2.460 0.013900 *
## factor(week)117   26.6132     9.9392   2.678 0.007424 **
## factor(week)118   26.5014     9.9095   2.674 0.007496 **
## factor(week)119   27.9500     9.9323   2.814 0.004899 **
## factor(week)120   36.6042     9.9109   3.693 0.000222 ***
## factor(week)121   35.0496     9.9317   3.529 0.000418 ***
## factor(week)122  125.9602    10.0031  12.592  < 2e-16 ***
## factor(week)123   25.5963     9.9056   2.584 0.009775 **
## factor(week)124   17.5425     9.9489   1.763 0.077876 .
## factor(week)125   20.1312     9.9533   2.023 0.043137 *
## factor(week)126   27.9608     9.9188   2.819 0.004824 **
## factor(week)127  201.9281    10.0769  20.039  < 2e-16 ***
## factor(week)128   13.9993     9.9048   1.413 0.157563
## factor(week)129   28.6200     9.9021   2.890 0.003855 **
## factor(week)130   16.1986     9.9103   1.635 0.102170
## factor(week)131   18.3100     9.9058   1.848 0.064565 .
## factor(week)132   17.8786     9.9652   1.794 0.072816 .
## factor(week)133   20.0554     9.9378   2.018 0.043599 *
## factor(week)134   19.4797     9.9176   1.964 0.049532 *
```

```
## factor(week)135   26.7934     9.9058   2.705 0.006842 **
## factor(week)136   94.8138     9.8966   9.580  < 2e-16 ***
## factor(week)137   17.2997     9.8579   1.755 0.079296 .
## factor(week)138   21.0836     9.8960   2.131 0.033145 *
## factor(week)139   42.8556     9.9132   4.323 1.55e-05 ***
## factor(week)140   32.7314     9.8574   3.320 0.000901 ***
## factor(week)141   23.3636     9.8689   2.367 0.017927 *
## factor(week)142   28.2204    10.0108   2.819 0.004824 **
## factor(week)143   95.4180    10.0084   9.534  < 2e-16 ***
## factor(week)144   17.5950     9.9636   1.766 0.077430 .
## factor(week)145   20.7300    10.0328   2.066 0.038825 *
## factor(week)146   20.4586     9.9451   2.057 0.039689 *
## factor(week)147   25.1984     9.9487   2.533 0.011324 *
## factor(week)148   88.0258     9.9367   8.859  < 2e-16 ***
## factor(week)149   32.6844     9.8952   3.303 0.000959 ***
## factor(week)150   13.7384    10.0233   1.371 0.170505
## factor(week)151   18.8133     9.9274   1.895 0.058101 .
## factor(week)152   19.8200     9.8978   2.002 0.045253 *
## factor(week)153   11.9356    10.4270   1.145 0.252362
## factor(week)154   57.4455     9.8579   5.827 5.75e-09 ***
## factor(week)155   18.8106    10.0070   1.880 0.060163 .
## factor(week)156   16.2379     9.9254   1.636 0.101862
## factor(week)157  230.4539     9.8825  23.319  < 2e-16 ***
## factor(week)158  120.2192     9.8559  12.198  < 2e-16 ***
## factor(week)159   17.9834     9.9079   1.815 0.069537 .
## factor(week)160   18.8139     9.8843   1.903 0.057008 .
## factor(week)161 1174.6969     9.9166 118.458  < 2e-16 ***
## factor(week)162   26.6810     9.8638   2.705 0.006840 **
## factor(week)163   10.3208     9.9157   1.041 0.297961
## factor(week)164   13.5380     9.8832   1.370 0.170772
## factor(week)165    8.2425    10.0769   0.818 0.413395
## factor(week)166    9.2891     9.9148   0.937 0.348828
## factor(week)167   13.9235     9.9152   1.404 0.160262
## factor(week)168    0.3635     9.8640   0.037 0.970605
## factor(week)169   21.5975     9.8607   2.190 0.028521 *
## factor(week)170    2.1990     9.8916   0.222 0.824079
## factor(week)171    2.6709     9.9760   0.268 0.788912
## factor(week)172    1.2012     9.8916   0.121 0.903349
## factor(week)173    6.4766     9.8905   0.655 0.512590
## factor(week)174   81.0736    10.2697   7.894 3.12e-15 ***
## factor(week)175   36.1773    10.2702   3.523 0.000429 ***
## factor(week)176   -1.6349     9.8931  -0.165 0.868743
## factor(week)177   -0.2837     9.8906  -0.029 0.977117
## factor(week)178  611.0807     9.9709  61.286  < 2e-16 ***
## factor(week)179   92.3112     9.9782   9.251  < 2e-16 ***
## factor(week)180  -10.2435    10.3341  -0.991 0.321591
## factor(week)181   -2.6761     9.9183  -0.270 0.787308
## factor(week)182   -3.9632     9.8910  -0.401 0.688654
## factor(week)183   -3.3458     9.8892  -0.338 0.735117
## factor(week)184   -6.3216    10.0039  -0.632 0.527457
## factor(week)185   -7.9301     9.9212  -0.799 0.424124
## factor(week)186   39.1561    10.1801   3.846 0.000120 ***
## factor(week)187   -1.7573     9.9199  -0.177 0.859392
## factor(week)188  -14.6410    11.9202  -1.228 0.219373
```

```
## factor(week)189  -11.0933    24.3997  -0.455 0.649367
## factor(week)190  579.7828    31.4204  18.452  < 2e-16 ***
## factor(week)191  -20.0315    43.7986  -0.457 0.647423
## factor(week)192  -13.1714    43.7910  -0.301 0.763588
## factor(week)193  -21.7162    61.5765  -0.353 0.724341
## factor(week)194  -28.7832    61.4996  -0.468 0.639776
## factor(week)195  -34.5086    61.4989  -0.561 0.574721
## factor(week)196  -32.7168    61.5009  -0.532 0.594753
## factor(week)197  -12.6022    61.5125  -0.205 0.837675
## factor(week)198  -29.9069    61.5114  -0.486 0.626832
## factor(week)199  -35.8126    61.5777  -0.582 0.560855
## factor(week)200  -27.6635    61.6200  -0.449 0.653484
## factor(week)204   31.9950    61.5148   0.520 0.602989
## factor(week)205   -3.1236    61.5444  -0.051 0.959523
## factor(week)207   18.8311    61.5445   0.306 0.759628
## factor(week)210  -10.9277    61.5496  -0.178 0.859084
## factor(week)212  -30.2741    61.5208  -0.492 0.622659
## factor(week)229   24.3212    61.5159   0.395 0.692580
## factor(week)255  -42.7601    61.5990  -0.694 0.487587
## factor(week)256  -38.7601    61.5990  -0.629 0.529206
## factor(week)277   16.2798    43.8825   0.371 0.710653
## factor(week)278 1232.5170    43.8680  28.096  < 2e-16 ***
## factor(week)279  -28.5465    43.8297  -0.651 0.514860
## factor(week)280   27.8266    43.9098   0.634 0.526272
## factor(week)281  -19.5654    36.1053  -0.542 0.587897
## factor(week)282 1232.5170    43.8680  28.096  < 2e-16 ***
## factor(week)283  -28.5465    43.8297  -0.651 0.514860
## factor(week)284  -17.7229    43.8128  -0.405 0.685840
## factor(week)285  -11.2128    43.8128  -0.256 0.798011
## factor(week)286    8.1879    43.8130   0.187 0.851755
## factor(week)287  655.6741    43.9602  14.915  < 2e-16 ***
## factor(week)288  -30.8935    43.8234  -0.705 0.480850
## factor(week)289  -18.6022    61.5125  -0.302 0.762342
## factor(week)298  -36.9830    61.5528  -0.601 0.547960
## factor(week)299   -4.9830    36.0965  -0.138 0.890206
## factor(week)300  -18.9830    61.5528  -0.308 0.757781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.06 on 14518 degrees of freedom
## Multiple R-squared:  0.7959, Adjusted R-squared:  0.7928
## F-statistic: 251.7 on 225 and 14518 DF,  p-value: < 2.2e-16
```

2.3) regressors: `price`, `promoflag`, store fixed effects

Use `feols()` to estimate a store fixed-effects model of sales with regressors: `price`, `promoflag`. Use `summary()` to summarize the results.

```r
library(fixest)
summary(feols(sales ~ price + promoflag | store , data = DF1))
```

```
## OLS estimation, Dep. Var.: sales
## Observations: 14,744
```

```
## Fixed-effects: store: 86
## Standard-errors: Clustered (store)
##           Estimate Std. Error  t value  Pr(>|t|)
## price     -76.4248    3.30935 -23.0936 < 2.2e-16 ***
## promoflag  21.8063    1.39980  15.5782 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 118.8     Adj. R2: 0.211508
##               Within R2: 0.179565
```

2.4) regressors: `price`, `promoflag`, time (week) fixed effects, store fixed effects

Use `feols()` to estimate a store fixed-effects model of sales with regressors: `price`, `promoflag`, and time (week) fixed effects. Use `summary()` to summarize the results.

```r
summary(feols(sales ~ price + promoflag | store + week , data = DF1))
```

```
## OLS estimation, Dep. Var.: sales
## Observations: 14,744
## Fixed-effects: store: 86,  week: 224
## Standard-errors: Clustered (store)
##             Estimate Std. Error   t value   Pr(>|t|)
## price     -22.499122    6.39858 -3.516269 0.00070476 ***
## promoflag   0.219081    1.65204  0.132613 0.89481311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 54.0     Adj. R2: 0.834627
##              Within R2: 0.015512
```

Discussion questions

What patterns do you notice? Adding which controls (time trends, time fixed effects, store fixed effects) leads to the greatest change in the price and promotion parameter estimates? Variety of patterns are noticed. On adding time fixed effects the value of price and promotion coefficient estimate changes and standard error also changes a little bit nit much though whereas on just adding time trend we can see that the effect of "week" is not significant.

However on adding store fixed effects through feols() we observe the greatest change in price estimate and for the promotion estimate it's the same as we get for other models (except when we consider fixed effects of both week and store).

When we take away the fixed effect of both week and store together then we see the lowest change in promotion estimate, however since the p value of the promotion store is high this indicates that relationship is not statistically significant and is out of randomness. Additionally the standard error has increased for price estimate when both the fixed effects are removed compared to when only store fixed effect is considered.

What does this suggest about sources of omitted variable bias? E.g., are omitted factors more likely associated with cross-sectional units (stores) or time periods? Omitted factors are likely associated with stores because when the fixed effect of stores was taken away we observed a larger decrease in sales of units when price increases by 1 unit.

Which estimate would you report as your "best" estimate of the demand response to price? Why? store. Because it shows greatest decrease in units with price increase.

Task 2

Your second task is to evaluate the effectiveness of an online display ad campaign.

You have data from an experiment designed to measure the effectiveness of an online display advertising campaign. The experiment involves randomly assigning Internet users to a test or a control group based on cookies that uniquely identify each user visiting a site where the ad exchange (Rocket Fuel) can place an ad. Users in the test group see an ad for a newly released handbag by TaskaBella, Rocket Fuel's client. Users in the control group are shown a public service announcement that is unrelated to the advertised product. Based on the unique IDs, Rocket Fuel is able to track which users eventually purchased a handbag from TaskaBella, allowing the analyst to discern the effectiveness of the campaign.

Each row in the CSV file data set `rocketfuel_data.csv` here represents a uniquely identified user in the ad campaign. For each user, the following six variables are provided:

`user_id`	Unique identifier of the user
`test`	1 if the user was exposed to the real ad 0 if the user was in the control group and was shown a PSA
`converted`	1 if the user made a purchase, 0 otherwise
`tot_impr`	Total number of ad impressions the user encountered (treat=ad, control=PSA)
`mode_impr_day`	Day of the week on which the user encountered the most impressions (1=Mon,. . . ,7=Sun)
`mode_impr_hour`	Hour of the day (0-23) in which the user encountered the most impressions

For these data, `converted` is the outcome, and `test` is the treatment indicator. `user_id` uniquely identifies users (and rows). The remaining variables provide additional information observed during the experiment.

The client firm TaskaBella estimates that a conversion generates approximately $40 in incremental profit for the firm. The cost to serve ads in the experiment was $9 CPM ($9 per 1000 impressions).

3) Exploratory analysis

Read the data into R and perform some exploratory analysis. Show your work in the R chunks below, and provide text answers following the R chunk.

```r
RFD <- read.csv("~/Desktop/Classroom - R/Marketing Analytics/data/rocketfuel_data.csv")
sum(RFD$test)
```

3.1) How many users are in the test and control conditions?

```
## [1] 564577
```

```r
sum(RFD$test == 0)
```

```
## [1] 23524
```

Users in test = 564577

Users in control = 23524

3.2) Conversion rates The conversion rate for a group is the fraction of users that purchase (`converted`==1) in that group.

What is the conversion rate (in percentage) for the – a) test group and b) control group?

```
mean(RFD[RFD$test == 0, ]$converted)
```

```
## [1] 0.01785411
```

```
mean(RFD[RFD$test == 1, ]$converted)
```

```
## [1] 0.02554656
```

a) test group = 2.55%

b) control group = 1.8%

4) Randomization checks

Verify that Rocketfuel implemented the randomization correctly by examining whether the distributions of the variable `tot_impr` for the test and control groups are the same. If the average number of impressions that users see in each group is different, then the differences in their response rate can be (potentially) attributed to this instead of the ads that they see. We can examine the distribution of `tot_impr`for the two groups in three ways: (simple) mean comparison, distribution (histogram) comparison, and formal difference in means t-tests.

4.1) Mean comparison Using the describe command, summarize `tot_impr` for two the groups of users (in test and control conditions). What is the mean of this variable each of these groups? Are the means similar?

```
summary(RFD$tot_impr[RFD$test == 1], na.rm = T)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    4.00   13.00   24.82   27.00 2065.00
```

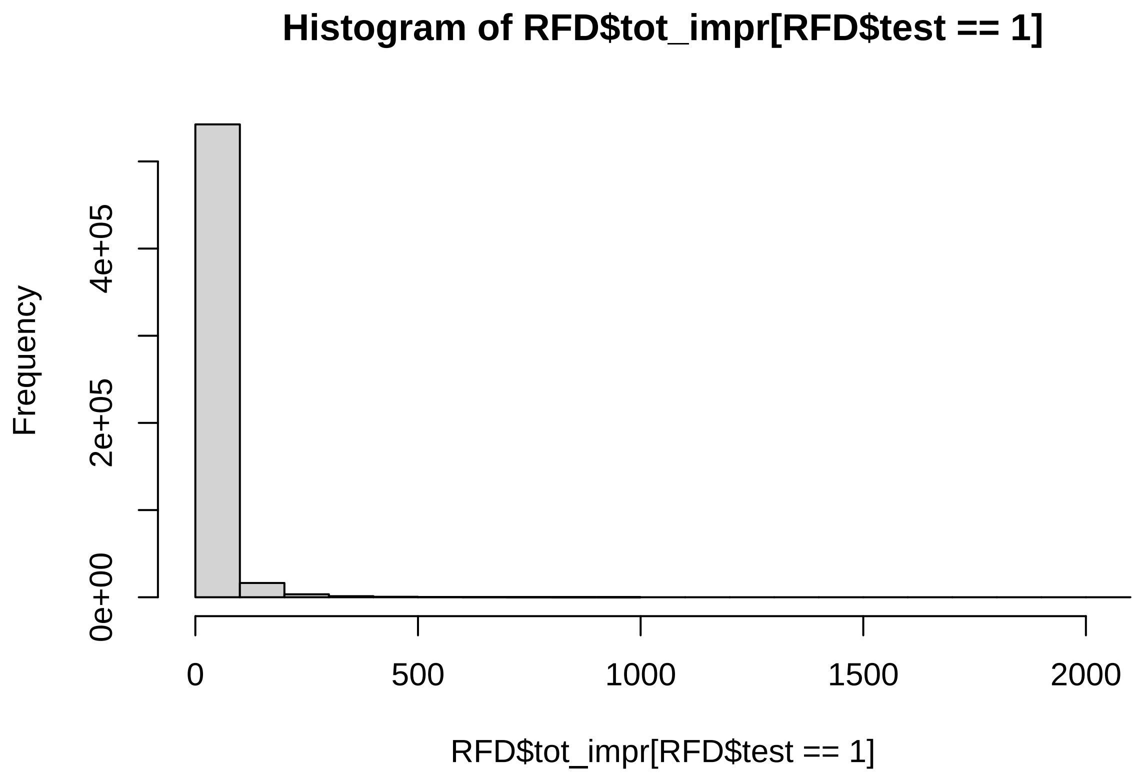
```
summary(RFD$tot_impr[RFD$test == 0], na.rm = T)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    4.00   12.00   24.76   26.00  907.00
```

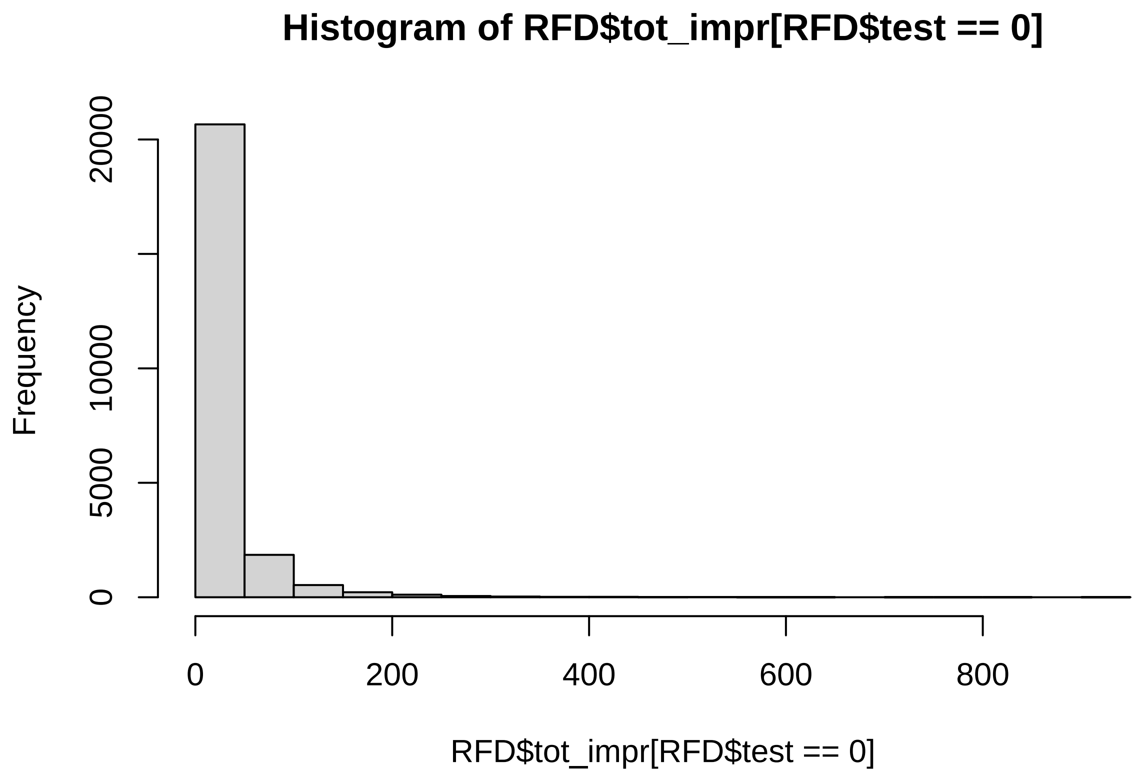
Yes, the means are similar

4.2) Distribution (histogram) comparison To further understand how the distribution of `tot_impr` looks for the two groups, plot the histograms of `tot_impr` for each of the two groups (test and control). Do the two histograms look similar?

```
hist(RFD$tot_impr[RFD$test == 1])
```




```r
hist(RFD$tot_impr[RFD$test == 0])
```



No the histograms don't look similar. The frequency of tot_impr shown to some of test users is way higher than shown to some of control group

4.3) Formal difference in means t-test Finally, conduct a t-test to examine whether the differences (if any) in `tot_impr` across the two groups is statistically significant?

```
t.test( tot_impr ~ test, data = RFD)
```

```
##
##  Welch Two Sample t-test
##
## data:  tot_impr by test
## t = -0.218, df = 25608, p-value = 0.8274
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.6217286  0.4972735
## sample estimates:
## mean in group 0 mean in group 1
##        24.76114        24.82337
```

Since P value is very high at 82% hence the relationship is not statistically significant.

Based on the above analyses, can you conclude that the randomization was done correctly?
The mean and the p-value suggestions that the difference between control group and test group is not statistically significant and that the randomization was done correctly.

5) Average treatment effect (ATE) estimation & application

5.2) Compute the treatment effect "by hand" Calculate the ATE as the difference in mean outcomes across the treatment and control conditions. Report your ATE estimate as a percentage. Was the campaign effective?

```
ATE<-mean(RFD[RFD$test==1, "converted"]) -
  mean(RFD[RFD$test==0, "converted"])
```

0.77%

5.3) Compute the treatment effect by regression Use a regression to estimate the treatment effect (ATE). Does your estimate match the "by hand" calculation? What is the standard error of the ATE?

```
summary(lm(converted ~ test, data = RFD))
```

```
##
## Call:
## lm(formula = converted ~ test, data = RFD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.02555 -0.02555 -0.02555 -0.02555  0.98215
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.017854   0.001023   17.46  < 2e-16 ***
## test        0.007692   0.001044    7.37  1.7e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1568 on 588099 degrees of freedom
## Multiple R-squared:  9.236e-05,  Adjusted R-squared:  9.066e-05
## F-statistic: 54.32 on 1 and 588099 DF,  p-value: 1.703e-13
```

Yes, the estimate matches the one by hand. The standard error is 0.0010, very low.

6) Return on investment (ROI)

We did not do this in class. But for extra points, give it a try.

6.1) Campaign incremental conversions For the users in the test group, how many extra conversions can be attributed to the ad campaign? In other words, what is the incremental number of conversions from the ad campaign?

Hint: the ATE is incremental (causal) effect of the campaign on conversion for each user. The total effect of the campaign on conversion is the number of users in the treatment condition times the ATE.

```
incremental_conv<-(sum(RFD$test == 1) * ATE)
```

4342

6.2) Campaign incremental profit Recall from the overview above that TaskaBella gets on average $40 for each conversion.

How much more money did TaskaBella make by running the campaign (excluding advertising costs)? In other words, what is the incremental profit from the ad campaign?

```
incremental_prof<-incremental_conv* c(40)
```

$173719

6.3) Campaign cost What was the cost of the campaign?

Hint: the relevant number of impressions is contained in the `tot_impr` variable.

```
total_impressions<-sum(RFD$tot_impr)
campaign_cost<- total_impressions*c(0.009)
```

$131375

6.4) ROI calculation Calculate the ROI of the campaign. Percentage ROI is defined as: 100*(incremental_profit - campaign_cost)/campaign_cost.

```
result <- incremental_prof - campaign_cost
ROI<- result/campaign_cost
print(ROI)
```

```
## [1] 0.3223198
```

32% is the ROI

6.5) Control group opportunity cost If the ad campaign had been shown to the control group as well, how much additional profit would have been generated? Explain your answer.

```
incremental_conv_control<-(sum(RFD$test == 0) * ATE)
incremental_prof_control<-incremental_conv_control*c(40)
print(incremental_prof_control)
```

```
## [1] 7238.291
```

Incremental profit generated would be $7238.
If the control group is given the same treatment the effect would be same as that observed for test group which is because if Average treatment effect. Hence we can apply the same ATE for the control group and find incremental profit by times $40.