```
---
title: "BA Project"
author: " Bharati Malik"
output:
  pdf_document: default
  word_document: default
editor_options:
  chunk_output_type: console
---

```{r}
# You should generally clear the working space at the start of every R session
rm(list = ls())

# Set the directory
setwd("C:/Users/asark/Desktop/Bharati SCU/1st Quarter/OMIS 2392 Econometrics with R (Lecture) Tue Thurs 7.35pm to 9.10 pm/0 BA projet/BOPS")

# install packages
#install.packages("readstata13")

# Load libraries
library(stargazer)
library(gdata)
library(ggplot2)
library(psych)
library(ggeffects)
library(QuantPsyc)
library(readstata13)
library(lmtest)
library(usdm)
library(multiwayvcov)
library(sandwich)
library(foreign)
library(AER)
library(aod)
library(Rcpp)
library(mfx)
library(nnet)
library(reshape2)
library(msm)

# turn off scientific notation.
options(scipen = 9)

```
```

#==
#
Data Loading
#==
#
```{r}

cust_Level=read.dta13("C:/Users/asark/Desktop/Bharati SCU/1st Quarter/OMIS 2392 Econometrics with R (Lecture) Tue Thurs 7.35pm to 9.10 pm/0 BA projet/BOPS/consumer level data.dta")
ODS_prodcat=read.dta13("C:/Users/asark/Desktop/Bharati SCU/1st Quarter/OMIS 2392 Econometrics with R (Lecture) Tue Thurs 7.35pm to 9.10 pm/0 BA projet/BOPS/online daily prod_cat sales-returns data.dta")
ODS_Overall=read.dta13("C:/Users/asark/Desktop/Bharati SCU/1st Quarter/OMIS 2392 Econometrics with R (Lecture) Tue Thurs 7.35pm to 9.10 pm/0 BA projet/BOPS/online daily sales-returns data.dta")
trans_level=read.dta13("C:/Users/asark/Desktop/Bharati SCU/1st Quarter/OMIS 2392 Econometrics with R (Lecture) Tue Thurs 7.35pm to 9.10 pm/0 BA projet/BOPS/transaction level data.dta")

head(cust_Level)
head(ODS_prodcat)
head(ODS_Overall)
head(trans_level) # view the first few rows of the data
```

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

	customer_id	store_number	age_band	est_income_code	homeowner_code	length_of_residence	child	female	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user
1	103465	2	5	7	R	15	Y	0	0	195.30	1	1	0
2	103465	2	5	7	R	15	Y	0	1	179.00	1	1	0
3	106978	2	9	4	O	3	Y	NA	0	72.00	1	1	0
4	106978	2	9	4	O	3	Y	NA	1	90.16	2	4	0
5	107501	2	9	1	O	15	N	0	0	2898.00	2	7	0
6	107501	2	9	1	O	15	N	0	1	1873.01	8	21	0
7	112054	2	0	6	O	2	Y	0	0	169.15	1	1	1
8	112054	2	0	6	O	2	Y	0	1	335.96	4	5	1
9	113914	2	9	7	O	13	Y	1	0	558.41	1	1	0
10	113914	2	9	7	O	13	Y	1	1	143.99	1	1	0
11	113980	2	0	7	O	2	Y	NA	0	89.10	1	1	0

Showing 1 to 12 of 84,420 entries

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

	store_number	year	month	month_index	month_dummy	bops_in_effect	day	salesvalue	returnvalue	salesquantity	returnquantity	avg_
1	2	2012	MAR	32	3	1	583	327516.28	54313.5586	1477	168	
2	2	2010	AUG	13	8	0	1	1290.03	48.9900	11	1	
3	2	2010	AUG	13	8	0	2	226777.23	33090.8086	1701	214	
4	2	2010	AUG	13	8	0	3	164627.44	25445.7695	1054	118	
5	2	2010	AUG	13	8	0	4	202743.41	45929.0586	1031	155	
6	2	2010	AUG	13	8	0	5	168273.48	26244.8496	988	100	
7	2	2010	AUG	13	8	0	6	161634.70	26731.4492	804	106	
8	2	2010	AUG	13	8	0	7	11058.96	1583.1799	77	11	
9	2	2010	AUG	13	8	0	8	241.69	0.0000	3	0	
10	2	2010	AUG	13	8	0	9	247811.56	36102.4297	1581	205	

Showing 1 to 12 of 2,858 entries

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

y	salesvalue	returnvalue	salesquantity	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner
583	327516.28	54313.5586	1477	168	0.5238829	4.363883	5.083560	0.5835598	6.724864	0.3559783
1	1290.03	48.9900	11	1	0.4444444	3.181818	4.545455	0.7272727	8.272727	0.2727273
2	226777.23	33090.8086	1701	214	0.5964794	4.982840	5.563905	0.6390532	7.076331	0.3396450
3	164627.44	25445.7695	1054	118	0.5744456	4.905033	5.462488	0.6201330	6.236467	0.3789174
4	202743.41	45929.0586	1031	155	0.5719063	5.119534	5.565598	0.6209912	7.001944	0.3624879
5	168273.48	26244.8496	988	100	0.5875706	5.084093	5.530902	0.6727457	7.467072	0.3535967
6	161634.70	26731.4492	804	106	0.5497954	5.022444	5.524938	0.6321696	7.008728	0.3678304
7	11058.96	1583.1799	77	11	0.5652174	4.578948	5.644737	0.5263158	6.736842	0.4210526
8	241.69	0.0000	3	0	0.0000000	3.333333	4.333333	0.6666667	4.000000	0.0000000
9	247811.56	36102.4297	1581	205	0.6028268	4.953768	5.620646	0.6846105	7.150728	0.3647878

Showing 1 to 12 of 2,858 entries

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

	store_number	year	month	month_index	product_category	month_dummy	bops_in_effect	day	salesvalue	returnvalue	salesquantity
1	2	2012	MAR	32	5	3	1	583	39173.61	4828.76	300
2	2	2010	AUG	13	4	8	0	1	201.18	0.00	1
3	2	2010	AUG	13	5	8	0	1	836.58	0.00	7
4	2	2010	AUG	13	13	8	0	1	203.28	0.00	2
5	2	2010	AUG	13	2	8	0	1	48.99	48.99	1
6	2	2010	AUG	13	12	8	0	2	25299.30	2745.30	367
7	2	2010	AUG	13	2	8	0	2	23478.47	3603.17	178
8	2	2010	AUG	13	4	8	0	2	51394.78	8993.87	236
9	2	2010	AUG	13	6	8	0	2	8042.66	2386.88	42
10	2	2010	AUG	13	21	8	0	2	19178.07	3335.25	246

Showing 1 to 12 of 30,475 entries

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

esvalue	returnvalue	salesquantity	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner
39173.61	4828.76	300	29	0.5092251	4.993311	5.612794	0.6296296	7.037037	0.36363637
201.18	0.00	1	0	1.0000000	0.000000	8.000000	1.0000000	4.000000	0.00000000
836.58	0.00	7	0	0.5000000	3.714286	4.428571	0.8571429	8.428572	0.28571430
203.28	0.00	2	0	0.0000000	2.500000	4.500000	0.5000000	10.000000	0.00000000
48.99	48.99	1	1	0.0000000	4.000000	2.000000	0.0000000	8.000000	1.00000000
25299.30	2745.30	367	34	0.6012461	5.046448	5.691257	0.6229508	7.628415	0.39071038
23478.47	3603.17	178	25	0.6894410	5.393259	4.966292	0.5842696	5.707865	0.34269664
51394.78	8993.87	236	38	0.4747475	4.474576	5.987288	0.6271186	7.495763	0.34745762
8042.66	2386.88	42	8	0.6842105	5.571429	5.047619	0.5714286	5.309524	0.28571430
19178.07	3335.25	246	43	0.6574074	4.630705	5.634855	0.6970955	7.419087	0.33195022

Showing 1 to 12 of 30,475 entries

BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

	customer_id	purchase_date	transaction_id	store_number	price	sku	return	age_band	est_income_code	homeowner_code	length_of
1	308421329614	05MAR2012:00:00:00	50002	2	386.55	12987079	0	7	8	O	
2	308421329614	05MAR2012:00:00:00	50002	2	1236.06	14550214	0	7	8	O	
3	12311873	01AUG2010:00:00:00	2038802	2	201.18	18094953	0	0	8	O	
4	822140002460	01AUG2010:00:00:00	2038804	2	199.00	18012591	0	5	3	O	
5	308431622663	01AUG2010:00:00:00	2038806	2	28.98	18095406	0	0	3	O	
6	25133466	01AUG2010:00:00:00	2038808	2	217.15	17812314	0	5	6	O	
7	22604254	01AUG2010:00:00:00	2038811	2	86.27	17581620	0	5	3	O	
8	22604254	01AUG2010:00:00:00	2038811	2	47.21	18095174	0	5	3	O	
9	22604254	01AUG2010:00:00:00	2038811	2	86.28	18072892	0	5	3	O	
10	26196930	01AUG2010:00:00:00	2038818	2	163.84	16928277	0	6	7	O	

Showing 1 to 12 of 1,671,502 entries

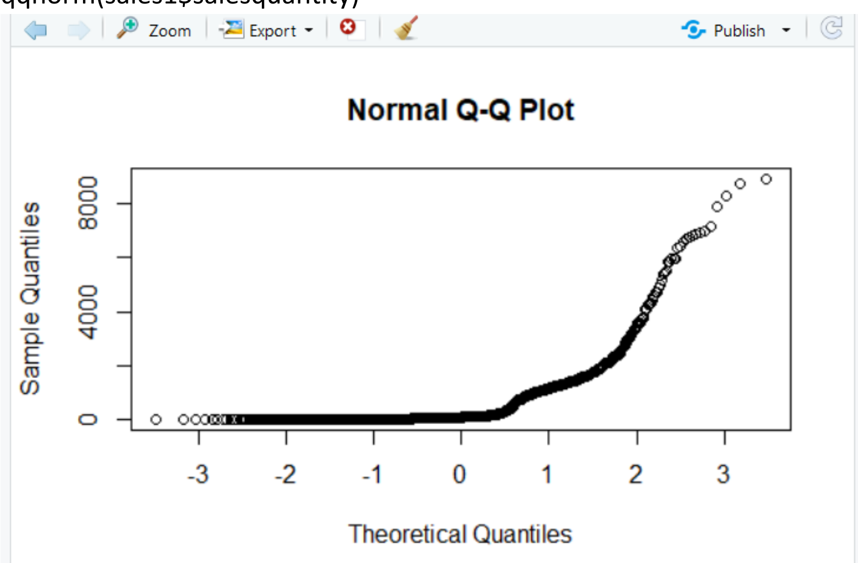
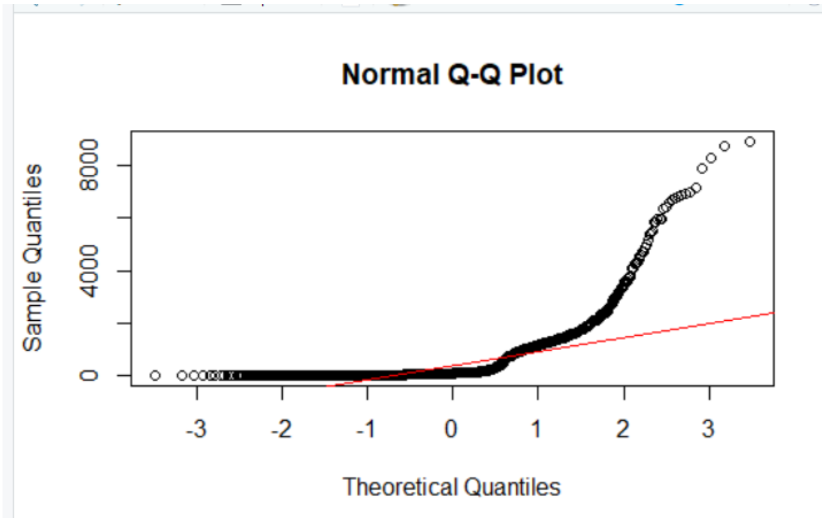
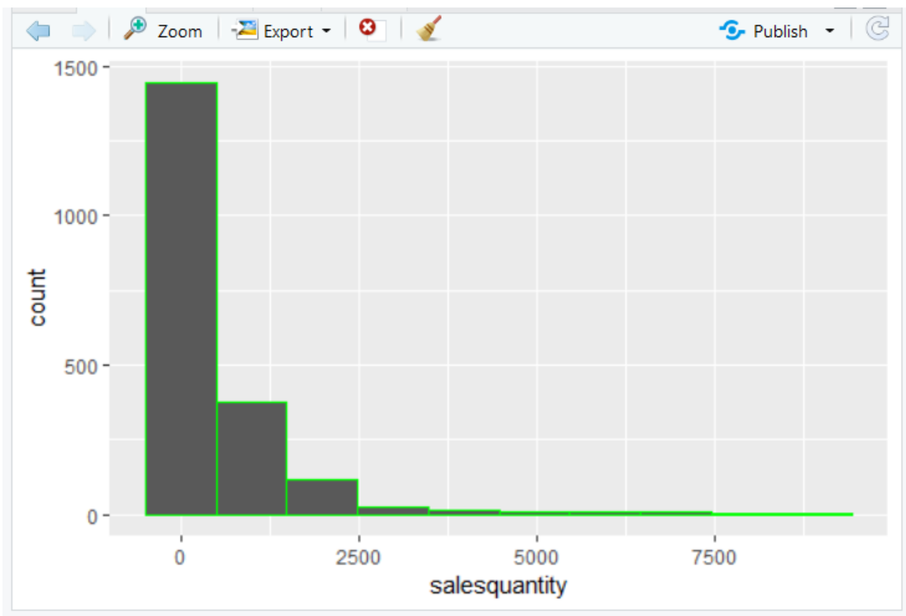
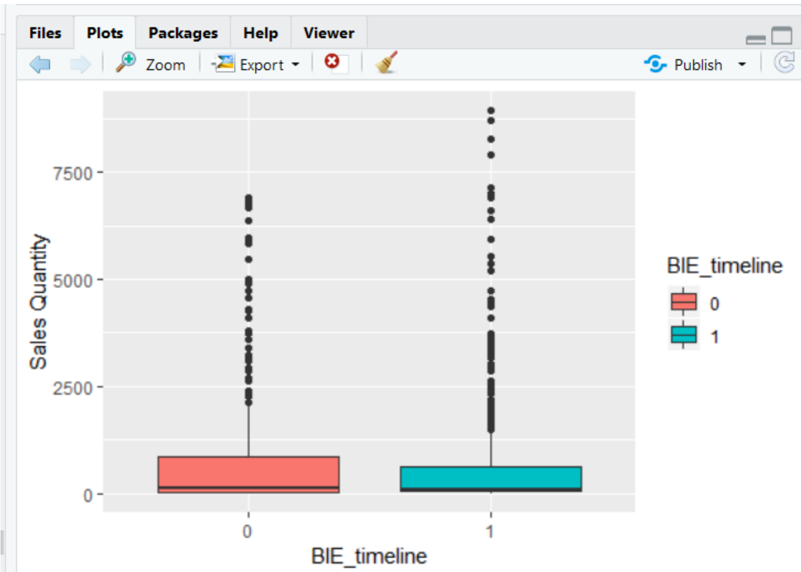
BA Project_BOPS_Final Submitted.Rmd | cust_Level | ODS_Overall | ODS_prodcat | trans_level

Filter

	homeowner_code	length_of_residence	child	year	month	month_index	product_category	month_dummy	week_index	bops	female
8	O	2	Y	2012	MAR	32	5	3	84	0	0
8	O	2	Y	2012	MAR	32	5	3	84	0	0
8	O	4	N	2010	AUG	13	4	8	1	NA	1
3	O	4	Y	2010	AUG	13	5	8	1	NA	0
3	O	12	N	2010	AUG	13	5	8	1	NA	1
6	O	14	Y	2010	AUG	13	5	8	1	NA	1
3	O	5	N	2010	AUG	13	5	8	1	NA	0
3	O	5	N	2010	AUG	13	5	8	1	NA	0
3	O	5	N	2010	AUG	13	13	8	1	NA	0
7	O	4	N	2010	AUG	13	5	8	1	NA	1

Showing 1 to 12 of 1,671,502 entries

```
#====================================================================================#
## Question1: What is the impact of implementing BOPS strategy on online channel sales?
#====================================================================================#
```{r}

#Descriptive Statistics
stargazer(ODS_Overall, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")
```

```
Descriptive Statistics
=====================================================================================
Statistic        N      Mean    St. Dev.   Min  Pctl(25)  Median  Pctl(75)     Max
-------------------------------------------------------------------------------------
store_number   2,858 1,635.6   2,668.4     2      2        6       5,998      5,998
year           2,858 2,011.6     0.9     2,010  2,011    2,012     2,012      2,013
month_index    2,858   31.0     10.2      13     22       31        40         48
month_dummy    2,858    6.5      3.5       1      3        6        10         12
bops_in_effect 2,858    0.6      0.5       0      0        1         1          1
day            2,858  564.6    311.0       1    298.2     564       837       1,096
salesvalue     2,858 99,619.0 180,986.6  0.0  8,157.6 18,157.7 152,580.9 1,685,114.0
returnvalue    2,858 15,346.2 26,160.0   0.0   799.0   2,646.6  24,473.5   200,464.5
salesquantity  2,858  584.9   1,146.3      1     43      106      784.2     11,630
returnquantity 2,858   59.1    106.5       0      4       10       89.8       895
avg_female     2,265    0.5      0.2      0.0    0.5      0.5       0.6        1.0
avg_age        2,345    4.9      1.7      0.0    4.5      5.0       5.5       13.0
avg_income     2,334    5.3      0.9      1.0    5.1      5.4       5.6        9.0
avg_homeowner  2,334    0.7      0.2      0.0    0.6      0.7       0.7        1.0
avg_residency  2,334    7.0      2.1      0.0    6.7      7.1       7.5       15.0
avg_childowner 2,334    0.4      0.2      0.0    0.3      0.4       0.4        1.0
-------------------------------------------------------------------------------------
```
```

#Filtering data for days before sept.27,2012, when BOPS was implemented only for stores 2&6.
sales1 <- ODS_Overall[(ODS_Overall$day<786),]

BA Project_BOPS_Final Submitted.Rmd × | sales1 × | cust_Level × | ODS_Overall × | ODS_prodcat × | trans_level ×

Filter

	store_number	year	month	month_index	month_dummy	bops_in_effect	day	salesvalue	returnvalue	salesquantity	retu
1	2	2012	MAR	32	3	1	583	327516.28	54313.56	1477	
2	2	2010	AUG	13	8	0	1	1290.03	48.99	11	
3	2	2010	AUG	13	8	0	2	226777.23	33090.81	1701	
4	2	2010	AUG	13	8	0	3	164627.44	25445.77	1054	
5	2	2010	AUG	13	8	0	4	202743.41	45929.06	1031	
6	2	2010	AUG	13	8	0	5	168273.48	26244.85	988	
7	2	2010	AUG	13	8	0	6	161634.70	26731.45	804	

Showing 1 to 8 of 2,005 entries

BA Project_BOPS_Final Submitted.Rmd × | sales1 × | cust_Level × | ODS_Overall × | ODS_prodcat × | trans_level ×

Filter

returnvalue	salesquantity	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner
54313.56	1477	168	0.5238829	4.363883	5.083560	0.5835598	6.724864	0.3559783
48.99	11	1	0.4444444	3.181818	4.545455	0.7272727	8.272727	0.2727273
33090.81	1701	214	0.5964794	4.982840	5.563905	0.6390532	7.076331	0.3396450
25445.77	1054	118	0.5744456	4.905033	5.462488	0.6201330	6.236467	0.3789174
45929.06	1031	155	0.5719063	5.119534	5.565598	0.6209912	7.001944	0.3624879
26244.85	988	100	0.5875706	5.084093	5.530902	0.6727457	7.467072	0.3535967
26731.45	804	106	0.5497954	5.022444	5.524938	0.6321696	7.008728	0.3678304

Showing 1 to 8 of 2,005 entries

#Add a BIE timeline variable dividing the data into two groups, one before Aug.1, 2011 and the other after.
#Add a grouping variable for stores dividing the data into two groups, one for 2&6 and the other for 5998.
sales1$BIE_timeline <- ifelse(sales1$day<366,0,1)
sales1$group_store <- ifelse((sales1$store_number==2) | (sales1$store_number==6),1,0)

BA Project_BOPS_Final Submitted.Rmd × | sales1 × | cust_Level × | ODS_Overall × | ODS_prodcat × | trans_level ×

Filter

returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store
168	0.5238829	4.363883	5.083560	0.5835598	6.724864	0.3559783	1	1
1	0.4444444	3.181818	4.545455	0.7272727	8.272727	0.2727273	0	1
214	0.5964794	4.982840	5.563905	0.6390532	7.076331	0.3396450	0	1
118	0.5744456	4.905033	5.462488	0.6201330	6.236467	0.3789174	0	1
155	0.5719063	5.119534	5.565598	0.6209912	7.001944	0.3624879	0	1
100	0.5875706	5.084093	5.530902	0.6727457	7.467072	0.3535967	0	1
106	0.5497954	5.022444	5.524938	0.6321696	7.008728	0.3678304	0	1

Showing 1 to 8 of 2,005 entries

#Assuming mean value for all 'na' values
sales1$avg_female[is.na(sales1$avg_female)] <- mean(sales1$avg_female, na.rm = TRUE)

```
sales1$avg_age[is.na(sales1$avg_age)] <- mean(sales1$avg_age, na.rm = TRUE)
sales1$avg_income[is.na(sales1$avg_income)] <- mean(sales1$avg_income, na.rm = TRUE)
sales1$avg_childowner[is.na(sales1$avg_childowner)] <- mean(sales1$avg_childowner, na.rm = TRUE)
sales1$avg_homeowner[is.na(sales1$avg_homeowner)] <- mean(sales1$avg_homeowner, na.rm = TRUE)
```

```
stargazer(sales1, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")
```

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
store_number	2,005	1,579.4	2,639.1	2	2	6	5,998	5,998
year	2,005	2,011.2	0.7	2,010	2,011	2,011	2,012	2,012
month_index	2,005	25.8	7.3	13	20	26	32	38
month_dummy	2,005	6.6	3.4	1	4	7	9	12
bops_in_effect	2,005	0.4	0.5	0	0	0	1	1
day	2,005	404.8	221.0	1	215	411	593	785
salesvalue	2,005	91,493.4	164,109.9	0.0	7,248.1	16,249.7	148,777.0	1,413,919.0
returnvalue	2,005	14,406.6	24,454.2	0.0	681.4	2,387.7	24,014.7	192,876.8
salesquantity	2,005	523.0	989.2	1	38	98	760	8,933
returnquantity	2,005	55.2	97.8	0	3	10	88	876
avg_female	2,005	0.5	0.2	0.0	0.5	0.5	0.6	1.0
avg_age	2,005	5.1	1.5	0.0	4.7	5.1	5.5	13.0
avg_income	2,005	5.4	0.8	1.0	5.2	5.4	5.5	9.0
avg_homeowner	2,005	0.7	0.2	0.0	0.6	0.7	0.7	1.0
avg_residency	1,633	7.1	2.1	0.0	6.7	7.0	7.5	15.0
avg_childowner	2,005	0.4	0.1	0.0	0.3	0.4	0.4	1.0
BIE_timeline	2,005	0.6	0.5	0	0	1	1	1
group_store	2,005	0.7	0.4	0	0	1	1	1

#check normalization

```
qqnorm(sales1$salesquantity)
```




```
qqline(sales1$salesquantity, col=2)
```



ggplot(sales1, aes(x=salesquantity)) + geom_histogram(colour="green", bins= 10)



#Analysing trend in raw dataset with boxplot for salesquantity.

df11 <- data.frame(salesquantity=sales1$salesquantity, BIE_timeline=as.factor(sales1$BIE_timeline))

BA Project_BOPS_Final Submitted.Rmd* | df11 | sales1

	salesquantity	BIE_timeline
1	1477	1
2	11	0
3	1701	0
4	1054	0
5	1031	0
6	988	0
7	804	0
8	77	0

Showing 1 to 8 of 2,005 entries

```
ggplot(df11, aes(x=BIE_timeline, y=salesquantity, fill=BIE_timeline)) + geom_boxplot() +
 xlab("BIE_timeline") + ylab("Sales Quantity")
```



Boxplot indicates that online purchase quantity decreases after BOPS implementation for stores 2&6.However, we can't come to a conclusion basis this interpretation.

Checking for multi-collinearity
sales value is 0 for 30 rows is while sales quantity exists (complementory products or service)

```
df12=sales1[c("BIE_timeline",
"group_store","month_dummy","avg_female","avg_age","avg_income","avg_homeowner",
"avg_childowner")]
```

BA Project_BOPS_Final Submitted.Rmd* × df12 × df11 × sales1 × cust_Level × ODS_Overall × ODS_prodcat × trans_l

Filter

	BIE_timeline	group_store	month_dummy	avg_female	avg_age	avg_income	avg_homeowner	avg_childowner
1	1	1	3	0.5238829	4.363883	5.083560	0.5835598	0.3559783
2	0	1	8	0.4444444	3.181818	4.545455	0.7272727	0.2727273
3	0	1	8	0.5964794	4.982840	5.563905	0.6390532	0.3396450
4	0	1	8	0.5744456	4.905033	5.462488	0.6201330	0.3789174
5	0	1	8	0.5719063	5.119534	5.565598	0.6209912	0.3624879
6	0	1	8	0.5875706	5.084093	5.530902	0.6727457	0.3535967
7	0	1	8	0.5497954	5.022444	5.524938	0.6321696	0.3678304
8	0	1	8	0.5652174	4.578948	5.644737	0.5263158	0.4210526

Showing 1 to 8 of 2,005 entries

round(cor(df12),3)

```
> round(cor(df12),3)
               BIE_timeline group_store month_dummy avg_female avg_age avg_income avg_homeowner avg_childowner
BIE_timeline          1.000      -0.084       0.059     -0.047  -0.103     -0.036         0.005          0.013
group_store          -0.084       1.000       0.035      0.126   0.106     -0.083         0.091          0.155
month_dummy           0.059       0.035       1.000     -0.031   0.020      0.069         0.079          0.027
avg_female           -0.047       0.126      -0.031      1.000   0.042     -0.073         0.039          0.031
avg_age              -0.103       0.106       0.020      0.042   1.000      0.114         0.214          0.140
avg_income           -0.036      -0.083       0.069     -0.073   0.114      1.000         0.308          0.075
avg_homeowner         0.005       0.091       0.079      0.039   0.214      0.308         1.000          0.300
avg_childowner        0.013       0.155       0.027      0.031   0.140      0.075         0.300          1.000
```

vifcor(df12)

```
> vifcor(df12) # There is no multi-collinearity in the data for the chosen variables.
No variable from the 8 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( avg_homeowner ~ BIE_timeline ):  0.004773467
max correlation ( avg_homeowner ~ avg_income ):  0.3077424

---------- VIFs of the remained variables --------
       Variables      VIF
1   BIE_timeline 1.026005
2    group_store 1.070168
3    month_dummy 1.014866
4     avg_female 1.026377
5        avg_age 1.077031
6     avg_income 1.136619
7  avg_homeowner 1.251431
8 avg_childowner 1.125705
>
```

There is no multi-collinearity in the data for the chosen variables.

##====Notes on control variables for the sales value and sales quantity mode===##

#We are not taking avg_residency as a control behavior, as conceptually, we don't see a relation between buying jewellery online and number of years a person has lived in his/her current residence.
#We are considering month_dummy to account for the effect of seasonality.

####---------------------IMPACT ON SALES QUANTITY------------------#####

#Since sales quantity is a non-negative integer value, we are using count data models to understand how it is impacted by BOPS implementation.

```
poisson11 <-
glm(salesquantity~BIE_timeline*group_store+as.factor(month_dummy)+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner, family="poisson", data=sales1)
stargazer(poisson11,
        title="Poisson Results", type="text",
        column.labels=c("Model-1"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

Poisson Results

	Dependent variable: salesquantity Model-1
BIE_timeline	0.22*** (0.01)
group_store	3.09*** (0.01)
as.factor(month_dummy)2	0.40*** (0.005)
as.factor(month_dummy)3	-0.23*** (0.01)
as.factor(month_dummy)4	-0.26*** (0.01)
as.factor(month_dummy)5	0.17*** (0.01)
as.factor(month_dummy)6	-0.32*** (0.01)
as.factor(month_dummy)7	-0.44*** (0.01)
as.factor(month_dummy)8	-0.23*** (0.01)
as.factor(month_dummy)9	-0.30*** (0.01)
as.factor(month_dummy)10	-0.15*** (0.01)
as.factor(month_dummy)11	0.41*** (0.005)
as.factor(month_dummy)12	1.23*** (0.004)
avg_female	-0.88*** (0.01)
avg_age	-0.19*** (0.001)
avg_income	0.29*** (0.002)
avg_homeowner	-0.62*** (0.01)
avg_childowner	0.42*** (0.01)
BIE_timeline:group_store	-0.21*** (0.01)
Constant	3.42*** (0.02)
Observations	2,005
Log Likelihood	-681,213.20
Akaike Inf. Crit.	1,362,466.00
Note:	*p<0.05; **p<0.01; ***p<0.001

Model fit assessment - Poisson

```
poisson11a <- glm(salesquantity~1, data=sales1, family="poisson")
```

poisson11	Large glm (30 elements, 1.8 Mb)
poisson11a	Large glm (30 elements, 1.4 Mb)

#running a comparison with null model.

```
lrtest(poisson11, poisson11a)
```

```
Likelihood ratio test

Model 1: salesquantity ~ BIE_timeline * group_store + as.factor(month_dummy) +
    avg_female + avg_age + avg_income + avg_homeowner + avg_childowner
Model 2: salesquantity ~ 1
  #Df   LogLik  Df  Chisq Pr(>Chisq)
1  20  -681213
2   1 -1108920 -19 855414  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

##Since Poisson doesn't fit the data, we will check for negative binomial model.

```
negbin11 <-
glm.nb(salesquantity~BIE_timeline*group_store+as.factor(month_dummy)+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner, data = sales1)
stargazer(negbin11,
    apply.coef = exp, t.auto=F, p.auto = F,
    title="Negative Bionomial Results", type="text",
    column.labels=c("IRRs"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

```
Negative Bionomial Results
=====================================================
                              Dependent variable:
                         ----------------------------
                                 salesquantity
                                     IRRs
-----------------------------------------------------
BIE_timeline                       1.4176**
                                   (0.1162)

group_store                       17.0403***
                                   (0.1046)

as.factor(month_dummy)2             1.3090
                                   (0.1425)

as.factor(month_dummy)3             0.7370*
                                   (0.1406)

as.factor(month_dummy)4             0.7827
                                   (0.1419)

as.factor(month_dummy)5             1.1738
                                   (0.1391)

as.factor(month_dummy)6             0.7368*
                                   (0.1417)

as.factor(month_dummy)7            0.6947**
                                   (0.1390)

as.factor(month_dummy)8             0.8239
                                   (0.1300)

as.factor(month_dummy)9             0.8129
                                   (0.1344)

as.factor(month_dummy)10            0.8494
                                   (0.1440)

as.factor(month_dummy)11           1.4442**
                                   (0.1386)

as.factor(month_dummy)12           3.2493***
                                   (0.1389)

avg_female                         0.3042***
                                   (0.1778)

avg_age                            0.8380***
                                   (0.0196)

avg_income                         1.3227***
                                   (0.0372)

avg_homeowner                      0.4862***
                                   (0.1885)

avg_childowner                      0.8962
                                   (0.2007)

BIE_timeline:group_store           0.6666**
                                   (0.1327)

Constant                          59.3098***
                                   (0.2675)

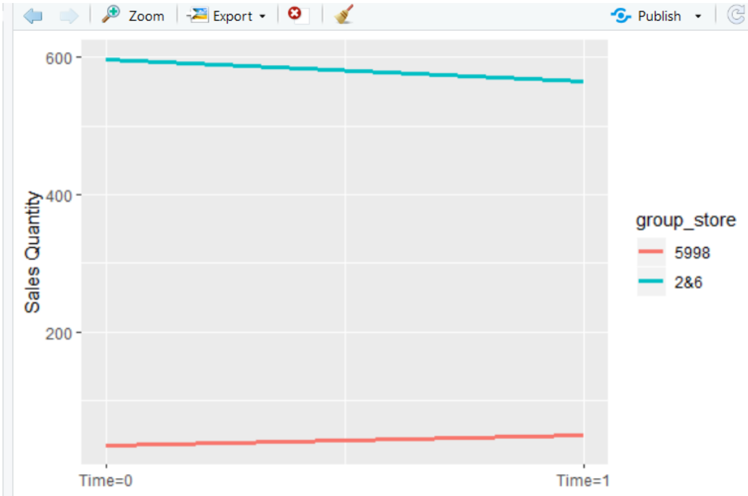
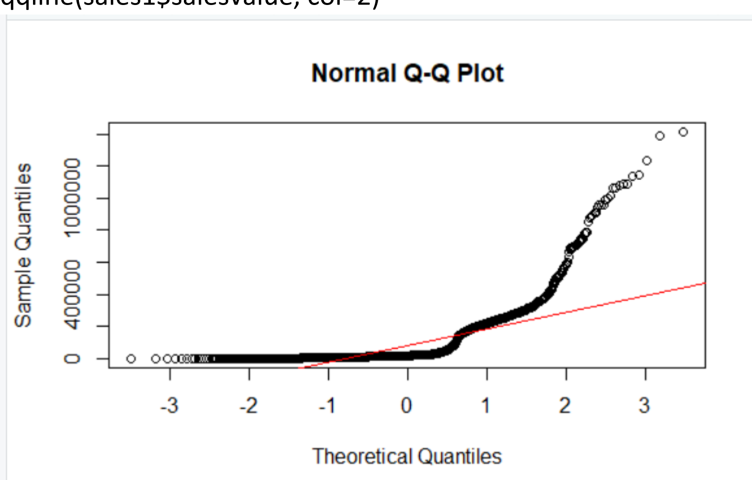
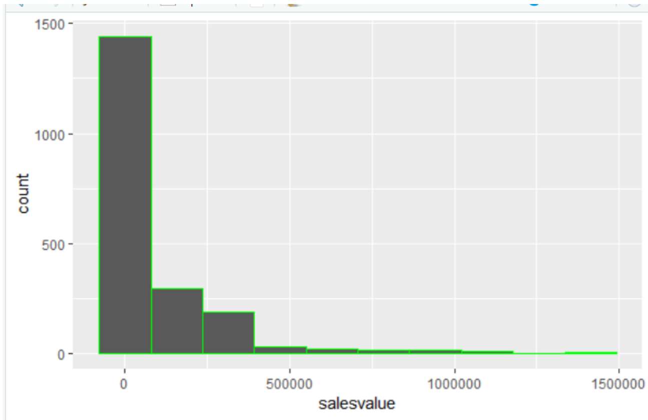
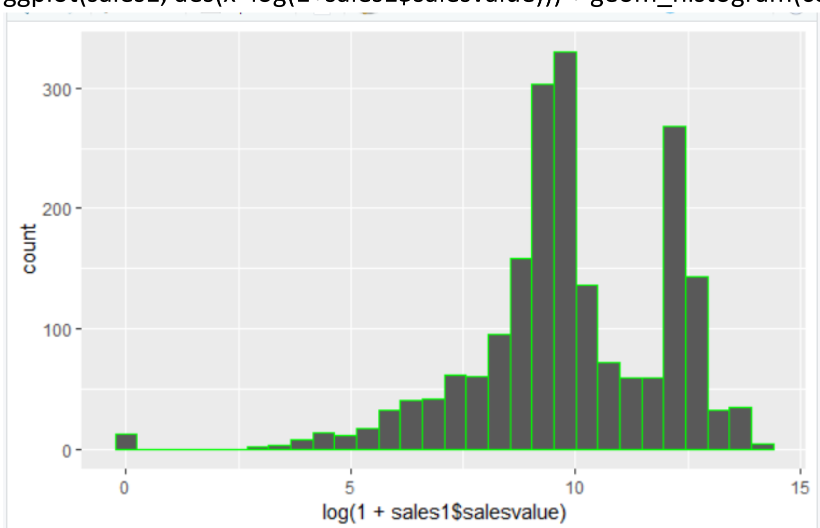
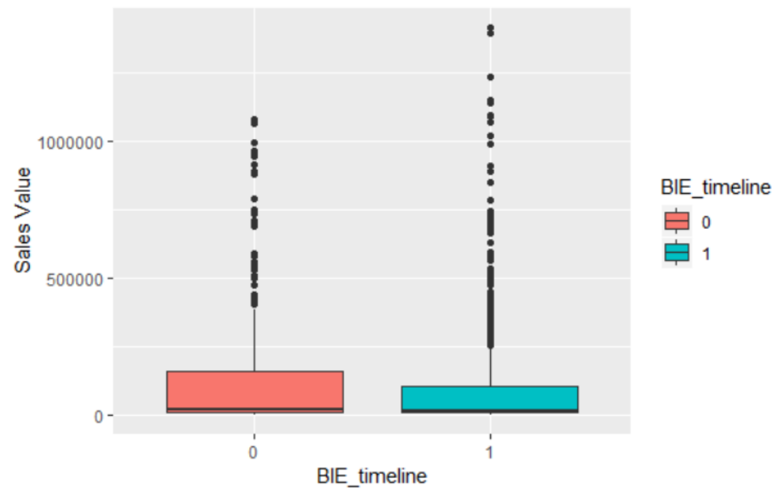
-----------------------------------------------------
Observations                         2,005
Log Likelihood                   -13,277.7000
theta                        0.6365*** (0.0172)
Akaike Inf. Crit.                 26,595.4100
=====================================================
Note:                    *p<0.05; **p<0.01; ***p<0.001
>
```

Model fit assessment - Negative Binomial
negbin11a <- glm.nb(salesquantity ~ 1, data = sales1)

negbin11	Large negbin (30 elements, 1.5 Mb)
negbin11a	Large negbin (28 elements, 1.1 Mb)

lrtest(negbin11, negbin11a)

```
Likelihood ratio test

Model 1: salesquantity ~ BIE_timeline * group_store + as.factor(month_dummy) +
    avg_female + avg_age + avg_income + avg_homeowner + avg_childowner
Model 2: salesquantity ~ 1
  #Df LogLik  Df  Chisq Pr(>Chisq)
1  21 -13277
2   2 -13856 -19 1158.9  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test results signifies that the model fits the data.

Check for heteroskedasticity
gqtest(negbin11) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(negbin11) # Breusch-Pagan test indicates heteroskedasticity
Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.

```
        Goldfeld-Quandt test

data:  negbin11
GQ = 0.0040545, df1 = 983, df2 = 982, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin11) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  negbin11
BP = 350.98, df = 19, p-value < 2.2e-16
```

HWrobstder <- sqrt(diag(vcovHC(negbin11, type="HC1"))) # produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:20] 0.256 0.0962 0.0998 0.1337 0.1...

```
stargazer(negbin11, negbin11,
    apply.coef = exp, t.auto=F, p.auto = F,
    se=list(NULL, HWrobstder),
    title="Negative Binomial Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE"),
    df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable: salesquantity	
	Normal SE (1)	HW-Robust SE (2)
BIE_timeline	1.418** (0.116)	1.418** (0.096)
group_store	17.040*** (0.105)	17.040*** (0.100)
as.factor(month_dummy)2	1.309 (0.142)	1.309 (0.134)
as.factor(month_dummy)3	0.737* (0.141)	0.737* (0.124)
as.factor(month_dummy)4	0.783 (0.142)	0.783 (0.125)
as.factor(month_dummy)5	1.174 (0.139)	1.174 (0.129)
as.factor(month_dummy)6	0.737* (0.142)	0.737* (0.132)
as.factor(month_dummy)7	0.695** (0.139)	0.695** (0.126)
as.factor(month_dummy)8	0.824 (0.130)	0.824 (0.123)
as.factor(month_dummy)9	0.813 (0.134)	0.813 (0.123)
as.factor(month_dummy)10	0.849 (0.144)	0.849 (0.135)
as.factor(month_dummy)11	1.444** (0.139)	1.444** (0.132)
as.factor(month_dummy)12	3.249*** (0.139)	3.249*** (0.126)
avg_female	0.304*** (0.178)	0.304*** (0.182)
avg_age	0.838*** (0.020)	0.838*** (0.021)
avg_income	1.323*** (0.037)	1.323*** (0.037)
avg_homeowner	0.486*** (0.188)	0.486*** (0.184)
avg_childowner	0.896 (0.201)	0.896 (0.228)
BIE_timeline:group_store	0.667** (0.133)	0.667** (0.114)
Constant	59.310*** (0.268)	59.310*** (0.256)
Observations	2,005	2,005
Log Likelihood	-13,277.700	-13,277.700
theta	0.636*** (0.017)	0.636*** (0.017)
Akaike Inf. Crit.	26,595.410	26,595.410
Note:		*p<0.05; **p<0.01; ***p<0.001

Visualize the output

meffects11 <- ggpredict(negbin11, terms=c("BIE_timeline", "group_store"))

BA Project_BOPS_Final Submitted.Rmd* | meffects11 | df12

Filter

	x	predicted	conf.low	conf.high	group
1	0	35.04624	27.16446	45.21494	0
2	0	597.19679	485.06072	735.25640	1
3	1	49.68168	39.40601	62.63688	0
4	1	564.31031	459.53045	692.98155	1

generates a tidy data frame

```
ggplot(meffects11,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Sales Quantity") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
  theme(axis.title.x=element_blank())
```



Interpretation- The interaction coefficient for BIE_timeline:group_store is significant with value 0.667. This means that the BOPS implementation is associated with 33.3% decrease in the sales quantity. This doesn't seem intuitive at first. However, a likely reason for this decrease is the knowledge of inventory that is available to a customer after BOPS implementation. When the inventory information for various stores is available while shopping online, a user might be more inclined to actually visit store and check out the product before making a purchase. So, even though the purchase was initiated online, it will actually get closed at the store and would be booked as a brick and mortar sales.

####=============== IMPACT ON SALES VALUE=====================================####

We are using OLS model to esatimate the imapct of BOPS implementation on sales value

#checking normalization
qqnorm(sales1$salesvalue)
qqline(sales1$salesvalue, col=2)



ggplot(sales1, aes(x=salesvalue)) + geom_histogram(colour="green", bins = 10)



ggplot(sales1, aes(x=log(1+sales1$salesvalue))) + geom_histogram(colour="green", bins = 30)



#Analysing trend in raw dataset with boxplot for salesvalue

df13 <- data.frame(salesvalue=sales1$salesvalue, BIE_timeline=as.factor(sales1$BIE_timeline))

BA Project_BOPS_Final Submitted.Rmd*

Filter

	salesvalue	BIE_timeline
1	327516.28	1
2	1290.03	0
3	226777.23	0
4	164627.44	0
5	202743.41	0
6	168273.48	0
7	161634.70	0
8	11058.96	0

Showing 1 to 8 of 2,005 entries

ggplot(df13, aes(x=BIE_timeline, y=salesvalue, fill=BIE_timeline)) + geom_boxplot() +
 xlab("BIE_timeline") + ylab("Sales Value")



Boxplot indicates that online purchase quantity decreases after BOPS implementation for stores 2&6.However, we can't come to a conclusion basis this interpretation.

#Since sales value is a dollar value, we will use a linear interaction model with log transformed dependent variable.

sales1$log_salesvalue <- log(1+sales1$salesvalue)

BA Project_BOPS_Final Submitted.Rmd* | df13 | meffects11 | df12 | df11 | sales1 | cust_Level | ODS_Overa

Filter

_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue
0.5238829	4.363883	5.083560	0.5835598	6.724864	0.3559783	1	1	12.699296
0.4444444	3.181818	4.545455	0.7272727	8.272727	0.2727273	0	1	7.163196
0.5964794	4.982840	5.563905	0.6390532	7.076331	0.3396450	0	1	12.331728
0.5744456	4.905033	5.462488	0.6201330	6.236467	0.3789174	0	1	12.011446
0.5719063	5.119534	5.565598	0.6209912	7.001944	0.3624879	0	1	12.219701
0.5875706	5.084093	5.530902	0.6727457	7.467072	0.3535967	0	1	12.033352
0.5497954	5.022444	5.524938	0.6321696	7.008728	0.3678304	0	1	11.993100

Showing 1 to 8 of 2,005 entries

```
ols11 =
lm(log_salesvalue~BIE_timeline*group_store+as.factor(month_dummy)+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner, data=sales1)
stargazer(ols11,
        title="Regression Results", type="text",
        column.labels=c("Model-1"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
=========================================================
                              Dependent variable:
                         --------------------------------
                                log_salesvalue
                                   Model-1
---------------------------------------------------------
BIE_timeline                         0.37*
                                    (0.18)

group_store                         1.95***
                                    (0.16)

as.factor(month_dummy)2              0.55*
                                    (0.22)

as.factor(month_dummy)3              0.17
                                    (0.22)

as.factor(month_dummy)4              0.02
                                    (0.22)

as.factor(month_dummy)5              0.15
                                    (0.22)

as.factor(month_dummy)6             -0.14
                                    (0.22)

as.factor(month_dummy)7             -0.29
                                    (0.22)

as.factor(month_dummy)8             -0.04
                                    (0.20)

as.factor(month_dummy)9             -0.19
                                    (0.21)

as.factor(month_dummy)10            -0.09
                                    (0.23)

as.factor(month_dummy)11             0.45*
                                    (0.22)

as.factor(month_dummy)12            1.45***
                                    (0.22)

avg_female                         -1.13***
                                    (0.28)

avg_age                            -0.19***
                                    (0.03)

avg_income                          0.28***
                                    (0.06)

avg_homeowner                       -0.54
                                    (0.29)

avg_childowner                       0.76*
                                    (0.31)

BIE_timeline:group_store            -0.49*
                                    (0.21)

Constant                            8.37***
                                    (0.42)

---------------------------------------------------------
Observations                         2,005
R2                                   0.18
Adjusted R2                          0.17
Residual Std. Error                  1.97
F Statistic                         22.58***
=========================================================
Note:                      *p<0.05; **p<0.01; ***p<0.001
```

Check for heteroskedasticity
gqtest(ols11) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols11) # Breusch-Pagan test indicates heteroskedasticity
Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.

```
        Goldfeld-Quandt test

data:  ols11
GQ = 0.61023, df1 = 983, df2 = 982, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols11) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols11
BP = 72.346, df = 19, p-value = 0.00000003732
```

HWrobstder <- sqrt(diag(vcovHC(ols11, type="HC1"))) # produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:20] 0.664 0.128 0.13 0.227 0.224 ...

stargazer(ols11, ols11,

```
se=list(NULL, HWrobstder),
title="OLS Results", type="text",
column.labels=c("Normal SE", "HW-Robust SE"),
df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
OLS Results
=====================================================
                            Dependent variable:
                       ------------------------------
                                log_salesvalue
                          Normal SE      HW-Robust
                             (1)            (2)
-----------------------------------------------------
BIE_timeline               0.372*        0.372**
                          (0.181)        (0.128)

group_store               1.951***       1.951***
                          (0.163)        (0.130)

as.factor(month_dummy)2    0.552*         0.552*
                          (0.223)        (0.227)

as.factor(month_dummy)3    0.168          0.168
                          (0.220)        (0.224)

as.factor(month_dummy)4    0.017          0.017
                          (0.222)        (0.236)

as.factor(month_dummy)5    0.150          0.150
                          (0.218)        (0.239)

as.factor(month_dummy)6   -0.139         -0.139
                          (0.222)        (0.235)

as.factor(month_dummy)7   -0.294         -0.294
                          (0.218)        (0.246)

as.factor(month_dummy)8   -0.043         -0.043
                          (0.204)        (0.216)

as.factor(month_dummy)9   -0.186         -0.186
                          (0.211)        (0.233)

as.factor(month_dummy)10   -0.091         -0.091
                          (0.226)        (0.239)

as.factor(month_dummy)11   0.453*         0.453
                          (0.217)        (0.239)

as.factor(month_dummy)12  1.451***       1.451***
                          (0.218)        (0.225)

avg_female               -1.128***       -1.128*
                          (0.278)        (0.478)

avg_age                  -0.192***      -0.192***
                          (0.031)        (0.047)

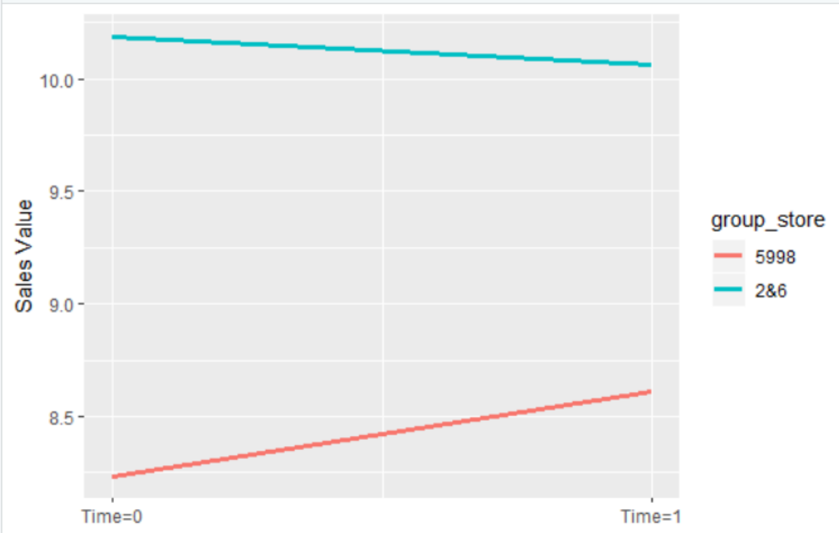
avg_income                0.283***       0.283**
                          (0.058)        (0.100)

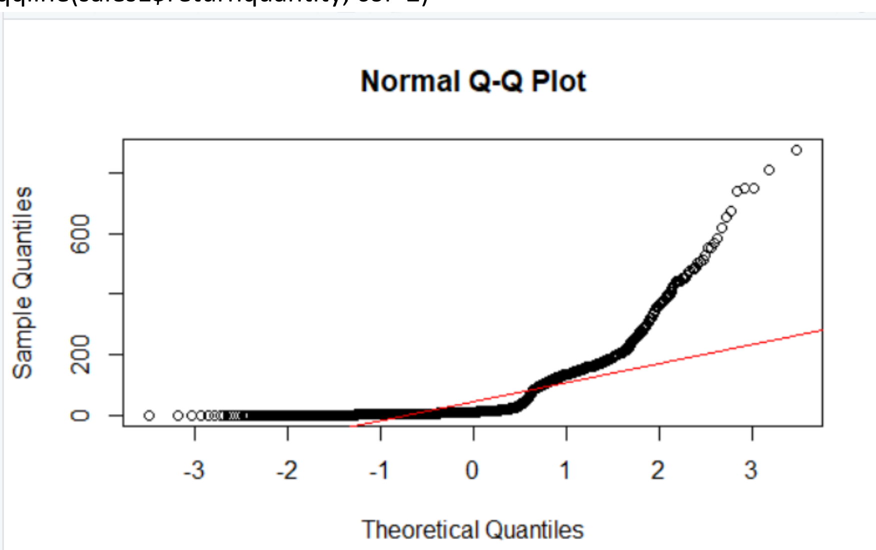
avg_homeowner             -0.544         -0.544
                          (0.295)        (0.490)

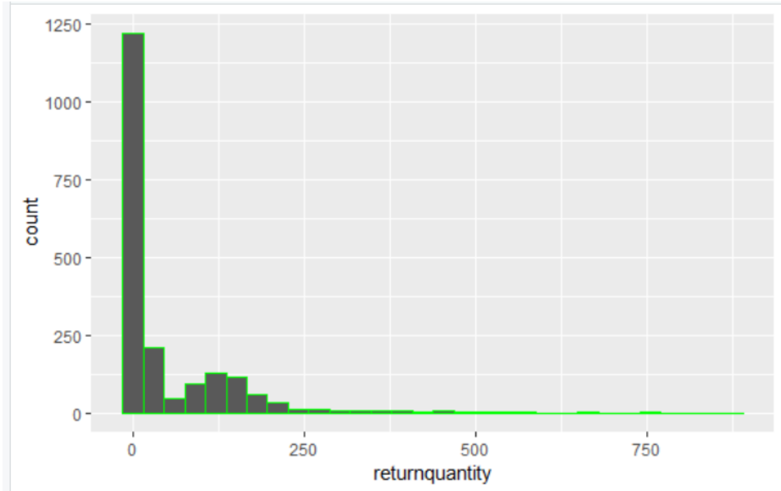
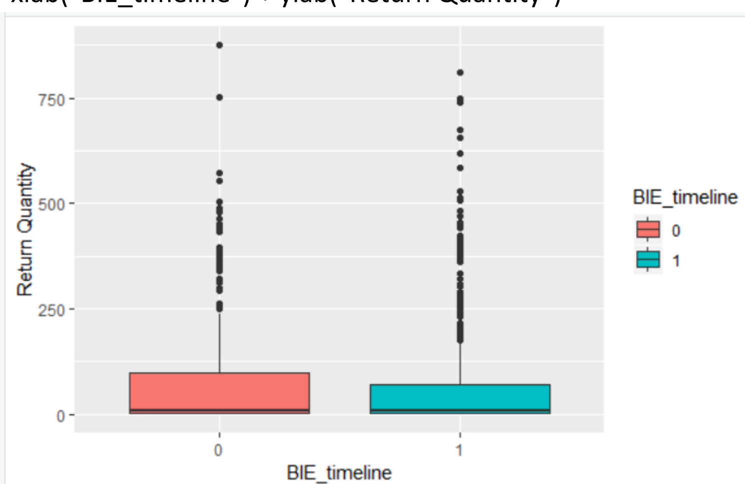
avg_childowner             0.764*         0.764
                          (0.314)        (0.605)

BIE_timeline:group_store  -0.494*        -0.494**
                          (0.207)        (0.168)

Constant                  8.366***       8.366***
                          (0.418)        (0.664)

-----------------------------------------------------
Observations               2,005          2,005
R2                         0.178          0.178
Adjusted R2                0.170          0.170
Residual Std. Error        1.969          1.969
F Statistic              22.580***      22.580***
=====================================================
Note:                 *p<0.05; **p<0.01; ***p<0.001
```

```
# Visualize the output
meffects12 <- ggpredict(ols11, terms=c("BIE_timeline", "group_store"))
# generates a tidy data frame
```

BA Project_BOPS_Final Submitted.Rmd* | meffects12 | df13

Filter

	x	predicted	conf.low	conf.high	group
1	0	8.235554	7.837718	8.633390	0
2	0	10.186837	9.860910	10.512763	1
3	1	8.607895	8.245467	8.970323	0
4	1	10.064837	9.742945	10.386728	1

```
ggplot(meffects12,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Sales Value") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
  theme(axis.title.x=element_blank())
```



Interpretation -The interaction coefficient for BIE_timeline:group_store is significant with value -0.494. This means that the BOPS implementation is associated with 49.4% decrease in the sales value. Like quantity, sales value is also decreasing after BOPS implementation. The percentage reduction in sales value is more than the reduction in sales qunatity. This implies that sales in general is reducing for relatively high priced items.

```
#=====================================================================================#
## Question2: What is the impact of implementing BOPS strategy on online channel returns?
#=====================================================================================#
```{r}
#check normalization
qqnorm(sales1$returnquantity)
qqline(sales1$returnquantity, col=2)
```




```
ggplot(sales1, aes(x=returnquantity)) + geom_histogram(colour="green", bins= 30)
```
```



#Analysing trend in raw dataset with boxplot for returnquantity
df21 <- data.frame(returnquantity=sales1$returnquantity, BIE_timeline=as.factor(sales1$BIE_timeline))

BA Project_BOPS_Final Submitted.Rmd* | df21

Filter

	returnquantity	BIE_timeline
1	168	1
2	1	0
3	214	0
4	118	0
5	155	0

Showing 1 to 5 of 2,005 entries

ggplot(df21, aes(x=BIE_timeline, y=returnquantity, fill=BIE_timeline)) + geom_boxplot() +
 xlab("BIE_timeline") + ylab("Return Quantity")



Boxplot indicates that return quantity should decrease after BOPS implementation for stores 2&6. However, we can't come to a conclusion basis this interpretation.

Checking for multi-collinearity

sales value is 0 for 30 rows is while sales quantity exists (complementory products or service)

df22=sales1[c("BIE_timeline", "group_store","salesquantity", "month_dummy","avg_female","avg_age","avg_income", "avg_childowner")]

	BIE_timeline	group_store	salesquantity	month_dummy	avg_female	avg_age	avg_income	avg_childowner
1	1	1	1477	3	0.5238829	4.363883	5.083560	0.3559783
2	0	1	11	8	0.4444444	3.181818	4.545455	0.2727273
3	0	1	1701	8	0.5964794	4.982840	5.563905	0.3396450
4	0	1	1054	8	0.5744456	4.905033	5.462488	0.3789174
5	0	1	1031	8	0.5719063	5.119534	5.565598	0.3624879
6	0	1	988	8	0.5875706	5.084093	5.530902	0.3535967

Showing 1 to 6 of 2,005 entries

round(cor(df22),3)

```
> round(cor(df22),3)
               BIE_timeline group_store salesquantity month_dummy avg_female avg_age avg_income avg_childowner
BIE_timeline          1.000      -0.084        -0.018       0.059     -0.047  -0.103     -0.036          0.013
group_store          -0.084       1.000         0.283       0.035      0.126   0.106     -0.083          0.155
salesquantity        -0.018       0.283         1.000       0.147     -0.102  -0.020      0.049          0.043
month_dummy           0.059       0.035         0.147       1.000     -0.031   0.020      0.069          0.027
avg_female           -0.047       0.126        -0.102      -0.031      1.000   0.042     -0.073          0.031
avg_age              -0.103       0.106        -0.020       0.020      0.042   1.000      0.114          0.140
avg_income           -0.036      -0.083         0.049       0.069     -0.073   0.114      1.000          0.075
avg_childowner        0.013       0.155         0.043       0.027      0.031   0.140      0.075          1.000
```

vifcor(df22)

```
> vifcor(df22)
No variable from the 8 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( avg_childowner ~ BIE_timeline ):  0.01283863
max correlation ( salesquantity ~ group_store ):  0.2830485

---------- VIFs of the remained variables --------
       Variables      VIF
1   BIE_timeline 1.024987
2    group_store 1.172569
3  salesquantity 1.140331
4    month_dummy 1.031143
5     avg_female 1.043402
6        avg_age 1.055700
7     avg_income 1.043918
8 avg_childowner 1.049147
```

##===Notes on control variables for the return value and return quantity model===##

#We are not taking AVG_RESIDENCY as a control behavior, as conceptually, we don't see a relation between returning jewellery and number of years a person has lived in his/her current residence.
#Also, we don't think that having a CHILD or being a HOMEOWNER affects jewellery returns.

####====================IMPACT ON RETURN QUANTITY ====================####

sales1$log_salesquantity <- log(sales1$salesquantity)

BA Project_BOPS_Final Submitted.Rmd* × | df22 × | df21 × | meffects12 × | df13 × | meffects11 × | df12 × | df11 × | sales1 × | cust_Level

Filter

	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue	log_salesquantity
3	0.5238829	4.363883	5.083560	0.5835598	6.724864	0.3559783	1	1	12.699296	7.297768
1	0.4444444	3.181818	4.545455	0.7272727	8.272727	0.2727273	0	1	7.163196	2.397895
4	0.5964794	4.982840	5.563905	0.6390532	7.076331	0.3396450	0	1	12.331728	7.438972
3	0.5744456	4.905033	5.462488	0.6201330	6.236467	0.3789174	0	1	12.011446	6.960348
5	0.5719063	5.119534	5.565598	0.6209912	7.001944	0.3624879	0	1	12.219701	6.938284

Showing 1 to 6 of 2,005 entries

```
poisson21 <-
glm(returnquantity~BIE_timeline*group_store+log_salesquantity+as.factor(month_dummy)+avg_age+avg_income+avg_female, family="poisson", data=sales1)
stargazer(poisson21,
      title="Poisson Results", type="text",
      column.labels=c("Model-1"),
      df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
Poisson Results
==================================================
                          Dependent variable:
                      ----------------------------
                            returnquantity
                               Model-1
--------------------------------------------------
BIE_timeline                   0.18***
                               (0.05)

group_store                    0.59***
                               (0.04)

log_salesquantity              0.95***
                               (0.004)

as.factor(month_dummy)2       -0.31***
                               (0.01)

as.factor(month_dummy)3       -0.14***
                               (0.02)

as.factor(month_dummy)4       -0.22***
                               (0.02)

as.factor(month_dummy)5       -0.34***
                               (0.02)

as.factor(month_dummy)6       -0.06***
                               (0.02)

as.factor(month_dummy)7        -0.02
                               (0.02)

as.factor(month_dummy)8       -0.24***
                               (0.01)

as.factor(month_dummy)9       -0.32***
                               (0.02)

as.factor(month_dummy)9                 -0.32***
                                        (0.02)

as.factor(month_dummy)10                -0.19***
                                        (0.02)

as.factor(month_dummy)11                -0.20***
                                        (0.01)

as.factor(month_dummy)12                -0.33***
                                        (0.01)

avg_age                                 -0.09***
                                        (0.01)

avg_income                               -0.02
                                        (0.01)

avg_female                               -0.03
                                        (0.05)

BIE_timeline:group_store                -0.19***
                                        (0.05)

Constant                                -1.63***
                                        (0.08)

------------------------------------------------------------
Observations                             2,005
Log Likelihood                          -9,646.01
Akaike Inf. Crit.                       19,330.02
============================================================
Note:                      *p<0.05; **p<0.01; ***p<0.001
```

Model fit assessment

```
poisson21a <- glm(returnquantity~1, data=sales1, family="poisson")
```

poisson21	Large glm (30 elements, 1.8 Mb)
poisson21a	Large glm (30 elements, 1.4 Mb)

running a comparison with null model.

lrtest(poisson21, poisson21a)

```
Likelihood ratio test

Model 1: returnquantity ~ BIE_timeline * group_store + log_salesquantity +
    as.factor(month_dummy) + avg_age + avg_income + avg_female
Model 2: returnquantity ~ 1
  #Df  LogLik  Df  Chisq Pr(>Chisq)
1  19   -9646
2   1 -118715 -18 218138  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

##Since Poisson doesn't fit the data, we will check for negative binomial model.

negbin21 <-
glm.nb(returnquantity~BIE_timeline*group_store+log_salesquantity+as.factor(month_dummy)+avg_age+avg_income+avg_female, data = sales1)

stargazer(negbin21,
    apply.coef = exp, t.auto=F, p.auto = F,
    title="Negative Bionomial Results", type="text",
    column.labels=c("IRRs"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))

```
Negative Bionomial Results
=====================================================
                            Dependent variable:
                        -----------------------------
                              returnquantity
                                   IRRs
-----------------------------------------------------
BIE_timeline                     1.2021**
                                 (0.0568)

group_store                      1.4453***
                                 (0.0538)

log_salesquantity                2.7196***
                                 (0.0075)

as.factor(month_dummy)2          0.7380***
                                 (0.0455)

as.factor(month_dummy)3          0.8648**
                                 (0.0468)

as.factor(month_dummy)4          0.8151***
                                 (0.0474)

as.factor(month_dummy)5          0.7114***
                                 (0.0458)

as.factor(month_dummy)6           0.8845*
                                 (0.0479)

as.factor(month_dummy)7           0.9661
                                 (0.0466)

as.factor(month_dummy)8          0.7659***
                                 (0.0427)

as.factor(month_dummy)9          0.7186***
                                 (0.0441)

as.factor(month_dummy)10         0.8076***
                                 (0.0469)

as.factor(month_dummy)11         0.7447***
                                 (0.0445)

as.factor(month_dummy)12         0.6885***
                                 (0.0432)

avg_age                          0.9698***
                                 (0.0092)

avg_income                        0.9880
                                 (0.0170)

avg_female                        0.9523
                                 (0.0843)

BIE_timeline:group_store         0.8279**
                                 (0.0599)

Constant                         0.1233***
                                 (0.1273)

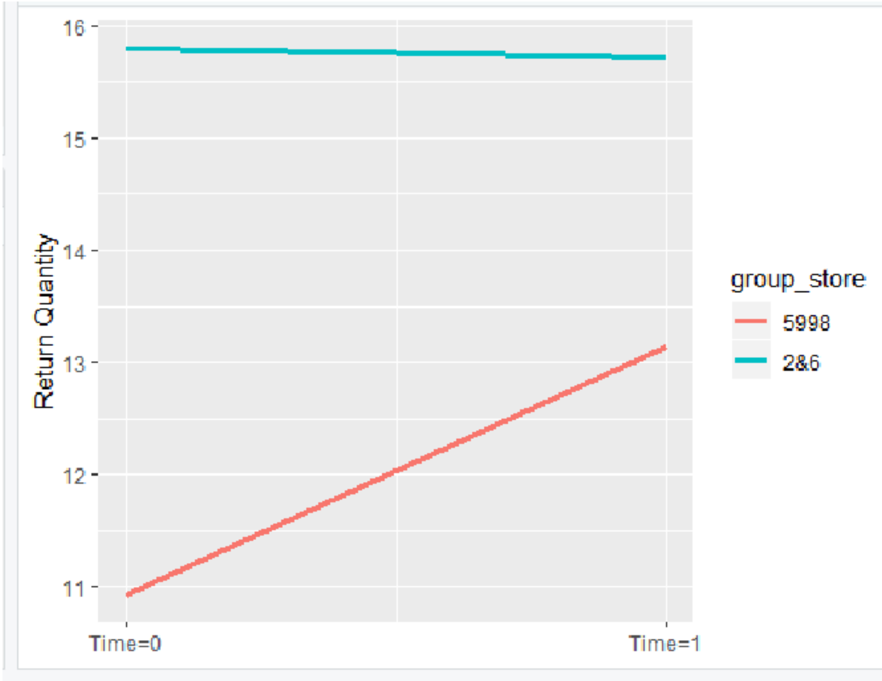
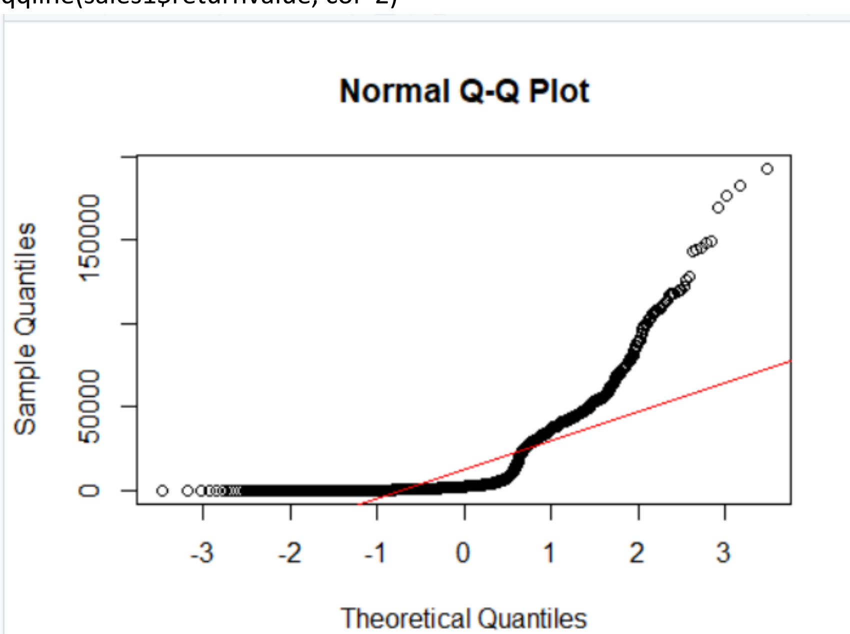
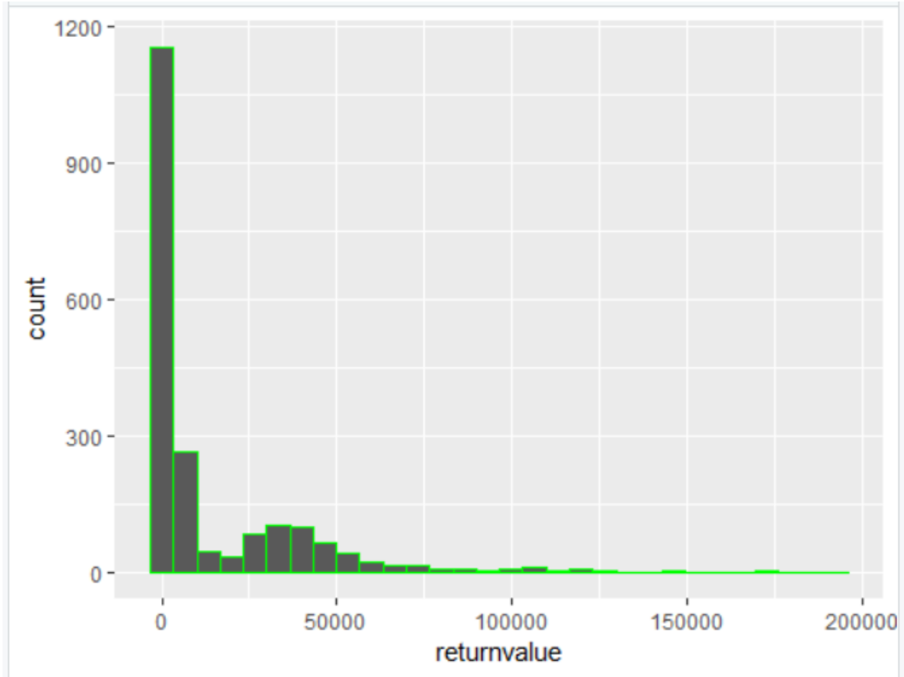
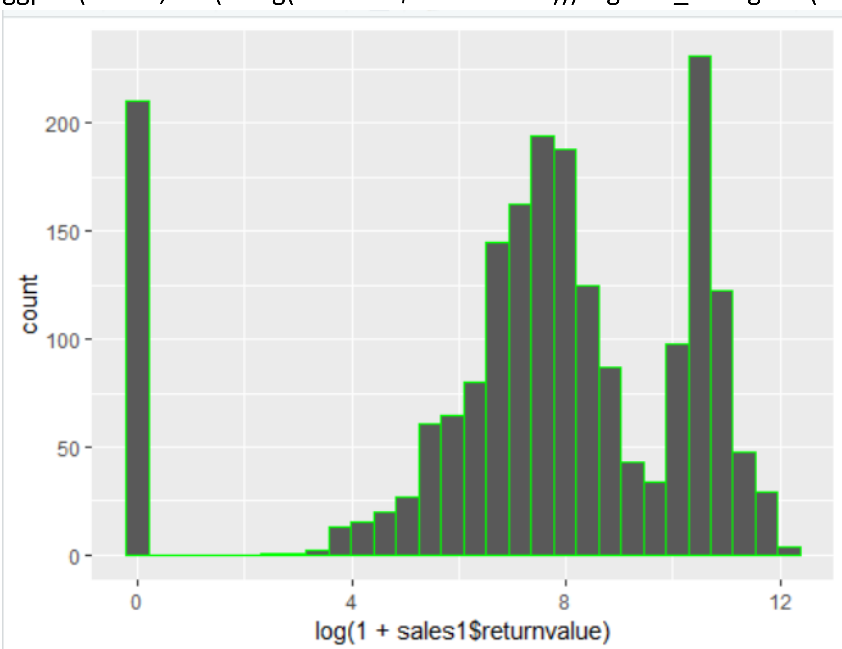
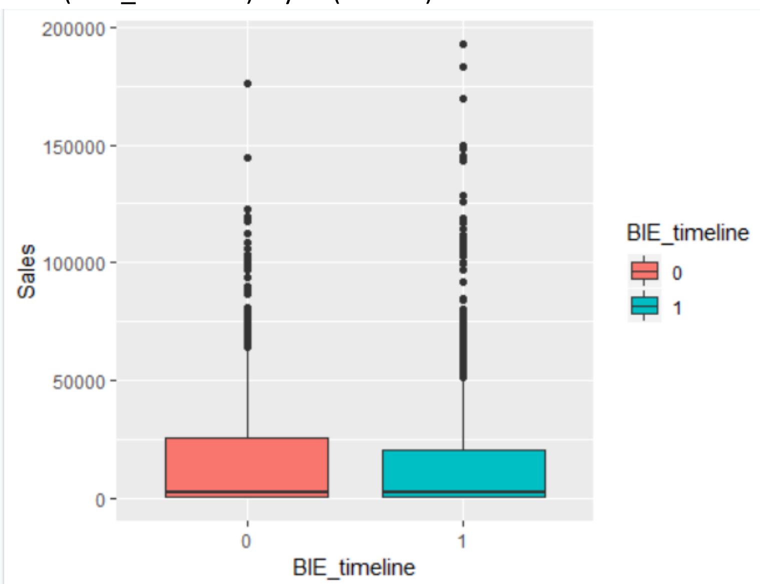
-----------------------------------------------------
Observations                       2,005
Log Likelihood                  -6,345.6270
theta                      11.7966*** (0.6434)
Akaike Inf. Crit.               12,729.2500
=====================================================
Note:                  *p<0.05; **p<0.01; ***p<0.001
```

Model fit assessment - Negative Binomial
negbin21a <- glm.nb(salesquantity ~ 1, data = sales1)

```
negbin21     Large negbin (30 elements, 1.5 Mb)
negbin21a    Large negbin (28 elements, 1.1 Mb)
```

lrtest(negbin21, negbin21a)

```
Likelihood ratio test

Model 1: returnquantity ~ BIE_timeline * group_store + log_salesquantity +
    as.factor(month_dummy) + avg_age + avg_income + avg_female
Model 2: salesquantity ~ 1
  #Df   LogLik  Df Chisq Pr(>Chisq)
1  20  -6344.6
2   2 -13856.2 -18 15023  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Check for heteroskedasticity
gqtest(negbin21) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(negbin21) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  negbin21
GQ = 0.0050659, df1 = 984, df2 = 983, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin21) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  negbin21
BP = 157.95, df = 18, p-value < 2.2e-16
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(negbin21, type="HC1")))
# produces Huber-White robust standard errors

```
Values
  HWrobstder    Named num [1:19] 0.13518 0.06045 0.05...
```

```
stargazer(negbin21, negbin21,
        apply.coef = exp, t.auto=F, p.auto = F,
        se=list(NULL, HWrobstder),
        title="Negative Binomial Results", type="text",
        column.labels=c("Normal SE", "HW-Robust SE"),
        df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable:	
	returnquantity	
	Normal SE	HW-Robust SE
	(1)	(2)
BIE_timeline	1.202**	1.202**
	(0.057)	(0.060)
group_store	1.445***	1.445***
	(0.054)	(0.059)
log_salesquantity	2.720***	2.720***
	(0.007)	(0.008)
as.factor(month_dummy)2	0.738***	0.738***
	(0.046)	(0.042)
as.factor(month_dummy)3	0.865**	0.865**
	(0.047)	(0.042)
as.factor(month_dummy)4	0.815***	0.815***
	(0.047)	(0.042)
as.factor(month_dummy)5	0.711***	0.711***
	(0.046)	(0.040)
as.factor(month_dummy)6	0.884*	0.884*
	(0.048)	(0.085)
as.factor(month_dummy)7	0.966	0.966
	(0.047)	(0.095)
as.factor(month_dummy)8	0.766***	0.766***
	(0.043)	(0.041)
as.factor(month_dummy)9	0.719***	0.719***
	(0.044)	(0.040)
as.factor(month_dummy)10	0.808***	0.808***
	(0.047)	(0.040)
as.factor(month_dummy)11	0.745***	0.745***
	(0.045)	(0.037)
as.factor(month_dummy)12	0.688***	0.688***
	(0.043)	(0.043)
avg_age	0.970***	0.970***
	(0.009)	(0.009)
avg_income	0.988	0.988
	(0.017)	(0.017)
avg_female	0.952	0.952
	(0.084)	(0.088)
BIE_timeline:group_store	0.828**	0.828**
	(0.060)	(0.066)
Constant	0.123***	0.123***
	(0.127)	(0.135)
Observations	2,005	2,005
Log Likelihood	-6,345.627	-6,345.627
theta	11.797*** (0.643)	11.797*** (0.643)
Akaike Inf. Crit.	12,729.250	12,729.250
Note:		*p<0.05; **p<0.01; ***p<0.001

# Visualize the output
meffects21 <- ggpredict(negbin21, terms=c("BIE_timeline", "group_store"))

BA Project_BOPS_Final Submitted.Rmd* | meffects21 | df22

Filter

	x	predicted	conf.low	conf.high	group
1	0	10.93292	9.787449	12.21246	0
2	0	15.80185	14.732843	16.94842	1
3	1	13.14236	12.042232	14.34298	0
4	1	15.72681	14.685611	16.84183	1

generates a tidy data frame at two different values of BIE_timeline - before and after BOPS implementation.

```
ggplot(meffects21,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Return Quantity") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
  theme(axis.title.x=element_blank())
```



Interpretation - The interaction coefficient for BIE_timeline:group_store is significant with value 0.828. This means that the BOPS implementation is associated with 17.2% decrease in the return quantity.

####===================== IMPACT ON RETURN VALUE=============================####

qqnorm(sales1$returnvalue)
qqline(sales1$returnvalue, col=2)



ggplot(sales1, aes(x=returnvalue)) + geom_histogram(colour="green", bins = 30)



ggplot(sales1, aes(x=log(1+sales1$returnvalue)) + geom_histogram(colour="green", bins = 30)



#Analysing trend in raw dataset with boxplot for returnvalue
df23 <- data.frame(returnvalue=sales1$returnvalue, BIE_timeline=as.factor(sales1$BIE_timeline))

BA Project_BOPS_Final Submitted.Rmd* | df23

Filter

	returnvalue	BIE_timeline
1	54313.56	1
2	48.99	0
3	33090.81	0
4	25445.77	0

Showing 1 to 5 of 2,005 entries

ggplot(df23, aes(x=BIE_timeline, y=returnvalue, fill=BIE_timeline)) + geom_boxplot() +
 xlab("BIE_timeline") + ylab("Sales")



Boxplot indicates that return value decreases after BOPS implementation for stores 2&6.However, we can't come to a conclusion basis this interpretation.
#Since return value is a dollar value, we will use a linear interaction model with log transformed dependent variable.
sales1$log_returnvalue <- log(1+sales1$returnvalue)

df23 | meffects21 | df22 | df21 | meffects12 | df13 | meffects11 | df12 | df11 | sales1 | cust_Leve

Filter

avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue	log_salesquantity	log_returnvalue
5.083560	0.5835598	6.724864	0.3559783	1	1	12.699296	7.297768	10.902548
4.545455	0.7272727	8.272727	0.2727273	0	1	7.163196	2.397895	3.911823
5.563905	0.6390532	7.076331	0.3396450	0	1	12.331728	7.438972	10.407041
5.462488	0.6201330	6.236467	0.3789174	0	1	12.011446	6.960348	10.144344
5.565598	0.6209912	7.001944	0.3624879	0	1	12.219701	6.938284	10.734875
5.530902	0.6727457	7.467072	0.3535967	0	1	12.033352	6.895683	10.175263
5.524938	0.6321696	7.008728	0.3678304	0	1	11.993100	6.689599	10.193633
5.644737	0.5263158	6.736842	0.4210526	0	1	9.311087	4.343805	7.367822
4.333333	0.6666667	4.000000	0.0000000	0	1	5.491785	1.098612	0.000000

Showing 1 to 11 of 2,005 entries

```
ols21 =
lm(log_returnvalue~BIE_timeline*group_store+log_salesvalue+as.factor(month_dummy)+avg_age+avg_
income+avg_female, data=sales1)
stargazer(ols21,
       title="Regression Results", type="text",
       column.labels=c("Model-1"),
       df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
=====================================================
                         Dependent variable:
                    ---------------------------------
                           log_returnvalue
                               Model-1
-----------------------------------------------------
BIE_timeline                   0.64***
                               (0.13)

group_store                    0.85***
                               (0.12)

log_salesvalue                 1.24***
                               (0.02)

as.factor(month_dummy)2       -0.47**
                               (0.16)

as.factor(month_dummy)3       -0.60***
                               (0.16)

as.factor(month_dummy)4       -0.56***
                               (0.16)

as.factor(month_dummy)5       -0.53***
                               (0.16)

as.factor(month_dummy)6       -0.51**
                               (0.16)

as.factor(month_dummy)7       -0.44**
                               (0.16)

as.factor(month_dummy)8       -0.59***
                               (0.15)

as.factor(month_dummy)9       -0.34*
                               (0.15)

as.factor(month_dummy)10      -0.52**
                               (0.16)

as.factor(month_dummy)11      -0.59***
                               (0.16)

as.factor(month_dummy)12      -0.64***
                               (0.16)

avg_age                        -0.02
                               (0.02)

avg_income                      0.04
                               (0.04)

avg_female                     0.56**
                               (0.20)

BIE_timeline:group_store      -0.66***
                               (0.15)

Constant                      -5.51***
                               (0.33)

-----------------------------------------------------
Observations                    2,005
R2                              0.79
Adjusted R2                     0.79
Residual Std. Error             1.43
F Statistic                   413.30***
=====================================================
Note:                   *p<0.05; **p<0.01; ***p<0.001
```

```
# Check for heteroskedasticity
gqtest(ols21) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols21) # Breusch-Pagan test indicates heteroskedasticity
# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
```

```
        Goldfeld-Quandt test

data:  ols21
GQ = 2.3074, df1 = 984, df2 = 983, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols21) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols21
BP = 442.85, df = 18, p-value < 2.2e-16
```

HWrobstder <- sqrt(diag(vcovHC(ols21, type="HC1"))) # produces Huber-White robust standard errors

```
Values
  HWrobstder          Named num [1:19] 0.4854 0.1892 0.183 0.027…
```

```
stargazer(ols21, ols21,
      se=list(NULL, HWrobstder),
      title="Regression Results", type="text",
      column.labels=c("Normal SE", "HW-Robust SE"),
      df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
==========================================================
                                Dependent variable:
                          --------------------------------
                                  log_returnvalue
                             Normal SE      HW-Robust SE
                                (1)             (2)
----------------------------------------------------------
BIE_timeline                 0.637***        0.637***
                             (0.131)         (0.189)

group_store                  0.849***        0.849***
                             (0.122)         (0.183)

log_salesvalue               1.243***        1.243***
                             (0.016)         (0.028)

as.factor(month_dummy)2      -0.465**        -0.465**
                             (0.162)         (0.153)

as.factor(month_dummy)3      -0.600***       -0.600***
                             (0.160)         (0.156)

as.factor(month_dummy)4      -0.563***       -0.563**
                             (0.161)         (0.177)

as.factor(month_dummy)5      -0.533***       -0.533***
                             (0.158)         (0.154)

as.factor(month_dummy)6      -0.511**        -0.511**
                             (0.161)         (0.169)

as.factor(month_dummy)7      -0.443**        -0.443*
                             (0.158)         (0.174)

as.factor(month_dummy)8      -0.589***       -0.589***
                             (0.148)         (0.145)

as.factor(month_dummy)9      -0.342*         -0.342*
                             (0.153)         (0.141)

as.factor(month_dummy)10     -0.519**        -0.519**
                             (0.164)         (0.160)

as.factor(month_dummy)11     -0.591***       -0.591***
                             (0.158)         (0.155)

as.factor(month_dummy)12     -0.644***       -0.644***
                             (0.159)         (0.138)

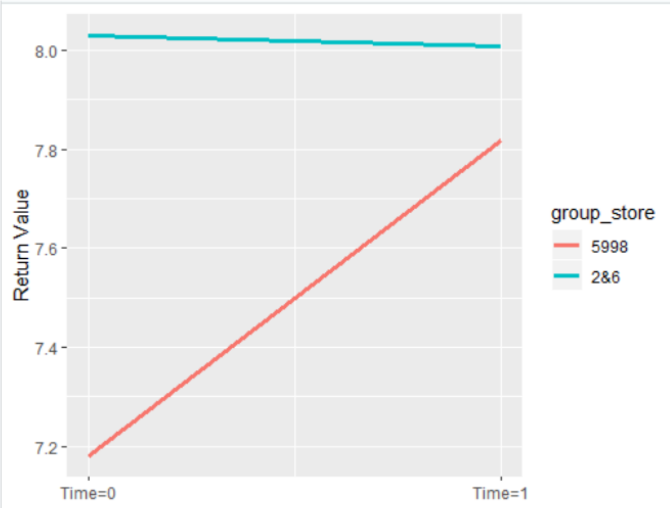
avg_age                      -0.016          -0.016
                             (0.022)         (0.030)

avg_income                   0.044           0.044
                             (0.040)         (0.061)

avg_female                   0.561**         0.561
                             (0.202)         (0.295)

BIE_timeline:group_store     -0.660***       -0.660***
                             (0.150)         (0.199)

Constant                     -5.514***       -5.514***
                             (0.329)         (0.485)

----------------------------------------------------------
Observations                 2,005           2,005
R2                           0.789           0.789
Adjusted R2                  0.787           0.787
Residual Std. Error          1.428           1.428
F Statistic                  413.297***      413.297***
==========================================================
Note:                     *p<0.05; **p<0.01; ***p<0.001
```

meffects22 <- ggpredict(ols21, terms=c("BIE_timeline", "group_store")) # generates a tidy data frame

BA Project_BOPS_Final Submitted.Rmd* | meffects22

Filter

	x	predicted	conf.low	conf.high	group
1	0	7.181039	6.888460	7.473619	0
2	0	8.029902	7.793377	8.266427	1
3	1	7.818156	7.553395	8.082917	0
4	1	8.006848	7.773493	8.240204	1

ggplot(meffects22,aes(x, predicted, colour=group)) + geom_line(size=1.3) +

```
xlab("BOPS Timeline") + ylab("Return Value") +
labs(colour="group_store") +
scale_colour_discrete(labels=c("5998","2&6")) +
scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
theme(axis.title.x=element_blank())
```




```
## Interpretation - The interaction coefficient for BIE_timeline:group_store is significant with value (-0.660). This means that the BOPS implementation is associated with ~66% decrease in the return value.
```

```
#==============================================================================#
## Question3: What is the impact of using the BOPS service on online customer purchase behavior?
#==============================================================================#
```{r}
#To answer question 3 we are using consumer data. Following variables in consumer data have missing values:
# 1.Female"-(~12%) values are missing
# 2.Age_band"
# 3.Est_income_code
# 4.Homeowner_code
# 5.Child
#Since female variable has a lot of missing entries, we used following two approaches to handle this variable in the model:
#1. Keeping Female as control variable and removing rows where female, homeonwer and child are missing.
#2. Removing Female as control variable, removing rows where homeonwer and child are missing.
#In both the approaches missing values in age_band and est_income_code are replaced with the median values of these variables.
#Creating dummy variable from factor variables ## this is done to ensure that these variables get included in descriptive stats and multi-collinearity test
cust_Level$childn = ifelse(cust_Level$child=="Y",1,ifelse(cust_Level$child=="N",0,cust_Level$child))
cust_Level$homeowner_coden =
ifelse(cust_Level$homeowner_code=="O",1,ifelse(cust_Level$homeowner_code=="R",0,cust_Level$homeowner_code))
cust_Level$childn <- as.numeric(cust_Level$childn)
```
```

cust_Level$homeowner_coden <- as.numeric(cust_Level$homeowner_coden)

BA Project_BOPS_Final Submitted.Rmd* × cust_Level ×

Filter

...th_of_residence	child	female	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user	childn	homeowner_coden
15	Y	0	0	195.30	1	1	0	1	0
15	Y	0	1	179.00	1	1	0	1	0
3	Y	NA	0	72.00	1	1	0	1	1
3	Y	NA	1	90.16	2	4	0	1	1
15	N	0	0	2898.00	2	7	0	0	1
15	N	0	1	1873.01	8	21	0	0	1
2	Y	0	0	169.15	1	1	1	1	1
2	Y	0	1	335.96	4	5	1	1	1

Showing 1 to 9 of 84,420 entries

####=======================APPROACH 1=====================================####

##--Keeping Female as control variable and removing rows where female, homeonwer and child are missing--###

#Descriptive Statistics

stargazer(cust_Level, type="text", median=TRUE, iqr=TRUE,digits=2, title="Descriptive Statistics")

```
Descriptive Statistics
=========================================================================================================
==
Statistic              N         Mean          St. Dev.        Min     Pctl(25)    Median    Pctl(75)      Max

---------------------------------------------------------------------------------------------------------
--
customer_id          84,420 60,707,965,648.00 198,215,133,553.00 103,465 18,418,146 26,079,925 27,885,126 919,600,001,7
22
store_number         84,420    208.31        1,092.20           2          2          2          2         5,998

age_band             81,454     4.88           3.93           0.00       0.00       5.00       8.00       13.00

est_income_code      81,434     5.46           2.22           1.00       4.00       6.00       7.00        9.00

length_of_residence  81,434     7.32           5.42           0.00       2.00       6.00      13.00       15.00

female               74,518     0.43           0.49           0.00       0.00       0.00       1.00        1.00

bops_in_effect       84,420     0.50           0.50             0          0         0.5          1           1

salesvalue           84,420   392.67         930.65           0.00      98.99     197.99     388.93    82,226.24

salesquantity        84,420     2.52           7.64             1          1          1          3        1,474

purchase_time_period 84,420     2.89           4.25             1          1          1          2          24

bops_user            84,420     0.19           0.39             0          0          0          0           1

childn               81,434     0.39           0.49           0.00       0.00       0.00       1.00        1.00

homeowner_coden      81,434     0.68           0.47           0.00       0.00       1.00       1.00        1.00

---------------------------------------------------------------------------------------------------------
```

#There are missing values in each of the following columns - "length_of_residence"", "female"", "est_income_code"", "age_band". Since female is a dummy variable, we are not replacing it with it's mean. Instead, we are removing the rows where "female" variable has missing values. For other variables, we will replace the NA's with median values.

#We are creating sales2 data set to remove rows where "female", "child","homeowner_code" have NA's

sales2 <- cust_Level[which((cust_Level$female==1 | cust_Level$female==0) & (cust_Level$child =="Y" | cust_Level$child =="N"),(cust_Level$homeowner_code=="R" | cust_Level$homeowner_code=="O")),] BA Project_BOPS_Final Submitted.Rmd* | sales2 | cust_Level

Filter

...th_of_residence	child	female	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user	childn	homeowner_coden
15	Y	0	0	195.30	1	1	0	1	0
15	Y	0	1	179.00	1	1	0	1	0
15	N	0	0	2898.00	2	7	0	0	1
15	N	0	1	1873.01	8	21	0	0	1
2	Y	0	0	169.15	1	1	1	1	1
2	Y	0	1	335.96	4	5	1	1	1
13	Y	1	0	558.41	1	1	0	1	1

Showing 1 to 9 of 74,504 entries

#stargazer(sales2, type="text", median=TRUE, iqr=TRUE, digits=1, title="Descriptive Statistics")

```
Descriptive Statistics
========================================================================================================================
==
Statistic                N          Mean            St. Dev.         Min     Pctl(25)      Median      Pctl(75)       Max

------------------------------------------------------------------------------------------------------------------------
--
customer_id           74,504 64,766,358,970.0 203,798,610,698.0 103,465 18,362,271 25,989,831.0 27,748,712 919,600,001,7
22
store_number          74,504       4.1             105.3          2          2           2           2          5,998

age_band              74,492       5.2              3.9          0.0        0.0         5.0         8.0          13.0

est_income_code       74,504       5.5              2.2           1          4           6           7            9

length_of_residence   74,504       7.5              5.4           0          2           6          14           15

female                74,504       0.4              0.5           0          0           0           1            1

bops_in_effect        74,504       0.5              0.5           0          0          0.5          1            1

salesvalue            74,504      382.5            756.2         0.0       98.7       194.6       384.1       47,648.4

salesquantity         74,504       2.4              4.1           1          1           1           2           369

purchase_time_period  74,504       2.9              4.3           1          1           1           2           24

bops_user             74,504       0.2              0.4           0          0           0           0            1

childn                74,504       0.4              0.5           0          0           0           1            1

homeowner_coden       74,504       0.7              0.5           0          0           1           1            1

------------------------------------------------------------------------------------------------------------------------
--
```

Replacing NA's with mean values for other variables.

sales2$age_band[is.na(sales2$age_band)] = median(sales2$age_band, na.rm = TRUE)

BA Project_BOPS_Final Submitted.Rmd* | sales2 | cust_Level

Filter

th_of_residence	child	female	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user	childn	homeowner_coden
15	Y	0	0	195.30	1	1	0	1	0
15	Y	0	1	179.00	1	1	0	1	0
15	N	0	0	2898.00	2	7	0	0	1
15	N	0	1	1873.01	8	21	0	0	1
2	Y	0	0	169.15	1	1	1	1	1
2	Y	0	1	335.96	4	5	1	1	1
13	Y	1	0	558.41	1	1	0	1	1
13	Y	1	1	143.99	1	1	0	1	1
15	N	1	0	43.61	1	1	0	0	1

Showing 1 to 10 of 74,504 entries

#Summary stats after replacing NA's with median and removing rows where "female","homeowner_code" & "child" are NA.

stargazer(sales2, type="text", median=TRUE, iqr=TRUE, digits=1, title="Descriptive Statistics")

```
Descriptive Statistics
========================================================================================================================
==
Statistic            N         Mean           St. Dev.        Min    Pctl(25)    Median     Pctl(75)       Max
------------------------------------------------------------------------------------------------------------------------
--
customer_id         74,504 64,766,358,970.0 203,798,610,698.0 103,465 18,362,271 25,989,831.0 27,748,712 919,600,001,7
22
store_number        74,504    4.1             105.3          2        2          2          2          5,998

age_band            74,504    5.2             3.9            0        0          5          8           13

est_income_code     74,504    5.5             2.2            1        4          6          7           9

length_of_residence 74,504    7.5             5.4            0        2          6          14          15

female              74,504    0.4             0.5            0        0          0          1           1

bops_in_effect      74,504    0.5             0.5            0        0          0.5        1           1

salesvalue          74,504   382.5           756.2          0.0      98.7       194.6      384.1      47,648.4

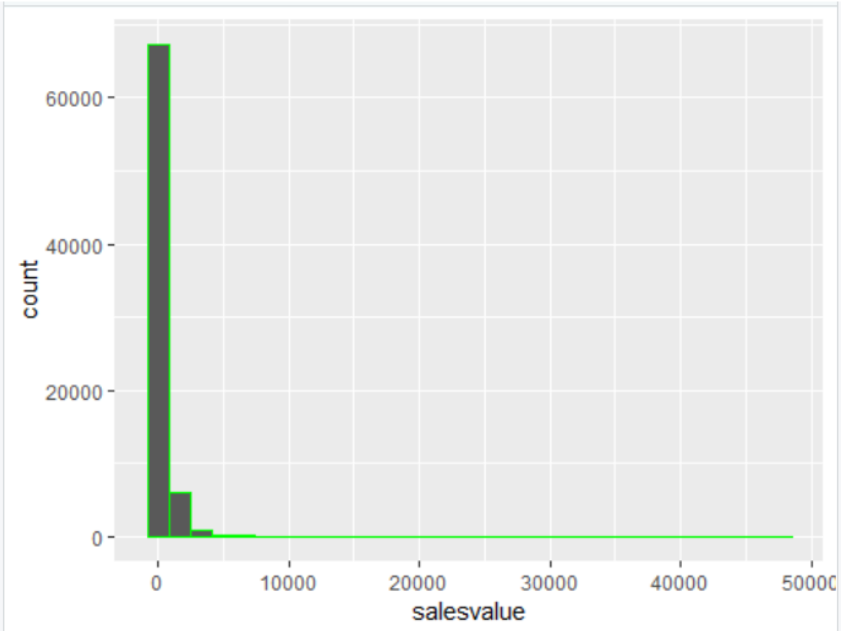
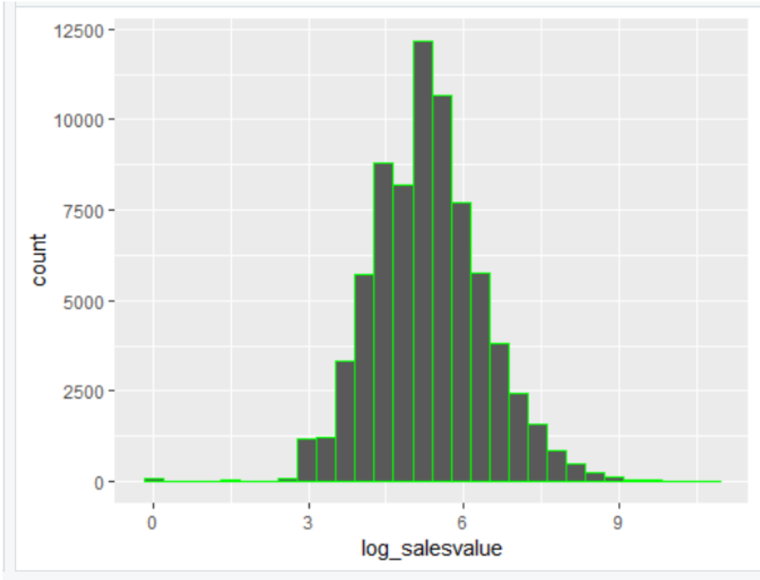
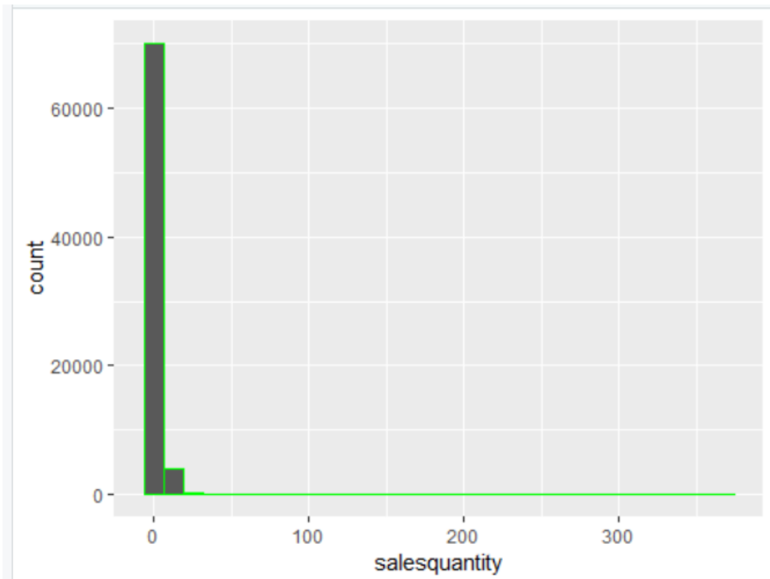
salesquantity       74,504    2.4             4.1            1        1          1          2           369

purchase_time_period 74,504   2.9             4.3            1        1          1          2           24

bops_user           74,504    0.2             0.4            0        0          0          0           1

childn              74,504    0.4             0.5            0        0          0          1           1

homeowner_coden     74,504    0.7             0.5            0        0          1          1           1

------------------------------------------------------------------------------------------------------------------------
```

#Check Normality

ggplot(sales2, aes(x=salesvalue)) + geom_histogram(colour="green", bins =30)



##salesvale doesn't seem to follow normal dist, so will use log of salesvalue as dependent variable.

sales2$log_salesvalue <- log(1+sales2$salesvalue)
#since salesvalue is 0 at some places, we are taking log of (1+salesvalue)

BA Project_BOPS_Final Submitted.Rmd* | sales2 | cust_Level

Filter

child	female	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user	childn	homeowner_coden	log_salesvalue
Y	0	0	195.30	1	1	0	1	0	5.279644
Y	0	1	179.00	1	1	0	1	0	5.192957
N	0	0	2898.00	2	7	0	0	1	7.972121
N	0	1	1873.01	8	21	0	0	1	7.535836
Y	0	0	169.15	1	1	1	1	1	5.136680
Y	0	1	335.96	4	5	1	1	1	5.819964

Showing 1 to 7 of 74,504 entries

ggplot(sales2, aes(x=log_salesvalue)) + geom_histogram(colour="green", bins =30)



ggplot(sales2, aes(x=salesquantity)) + geom_histogram(colour="green" , bins = 30)



Check Multicollinearity

df3 = sales2[c("homeowner_coden","purchase_time_period","length_of_residence","est_income_code","age_band", "bops_in_effect", "bops_user", "childn","female","store_number")]

BA Project_BOPS_Final Submitted.Rmd* | df3 | sales2 | cust_Level

Filter

	homeowner_coden	purchase_time_period	length_of_residence	est_income_code	age_band	bops_in_effect	bops_user	childn	female	store_number
1	0	1	15	7	5	0	0	1	0	2
2	0	1	15	7	5	1	0	1	0	2
5	1	7	15	1	9	0	0	0	0	2
6	1	21	15	1	9	1	0	0	0	2
7	1	1	2	6	0	0	1	1	0	2

Showing 1 to 5 of 74,504 entries

round(cor(df3),3)

```
> round(cor(df3),3)
                    homeowner_coden purchase_time_period length_of_residence est_income_code age_band bops_in_effect bops_user childn female store_number
homeowner_coden               1.000                0.029               0.278           0.315    0.176          0.000    -0.015  0.208  0.008       -0.002
purchase_time_period          0.029                1.000               0.026           0.009   -0.008          0.233     0.060  0.012  0.051       -0.006
length_of_residence           0.278                0.026               1.000           0.133    0.130          0.000    -0.025 -0.032  0.029       -0.005
est_income_code               0.315                0.009               0.133           1.000    0.112          0.000    -0.011  0.083 -0.022        0.003
age_band                      0.176               -0.008               0.130           0.112    1.000          0.000    -0.038 -0.002  0.077       -0.002
bops_in_effect                0.000                0.233               0.000           0.000    0.000          1.000     0.000  0.000  0.000       -0.004
bops_user                    -0.015                0.060              -0.025          -0.011   -0.038          0.000     1.000  0.016  0.071       -0.005
childn                        0.208                0.012              -0.032           0.083   -0.002          0.000     0.016  1.000  0.040       -0.003
female                        0.008                0.051               0.029          -0.022    0.077          0.000     0.071  0.040  1.000       -0.002
store_number                 -0.002               -0.006              -0.005           0.003   -0.002         -0.004    -0.005 -0.003 -0.002        1.000
```

vifcor(df3)

```
> vifcor(df3)
No variable from the 10 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( est_income_code ~ purchase_time_period ):  -0.00009741116
max correlation ( est_income_code ~ homeowner_coden ):  0.3239835

---------- VIFs of the remained variables --------
              Variables      VIF
1       homeowner_coden 1.276555
2  purchase_time_period 1.070553
3   length_of_residence 1.098384
4       est_income_code 1.125427
5              age_band 1.060988
6        bops_in_effect 1.063050
7             bops_user 1.006026
8                childn 1.067158
9                female 1.021652
10         store_number 1.001598
```

####========================IMPACT ON SALES QUANTITY=========================####

poisson31 <- glm(salesquantity~ bops_in_effect*bops_user + purchase_time_period + est_income_code + age_band + female + factor(store_number) + homeowner_coden +childn, family="poisson", data=sales2)

stargazer(poisson31,
    title="Poisson Results", type="text",
    column.labels=c("Model-1"),
    df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))

```
Poisson Results
=========================================================
                                 Dependent variable:
                             ----------------------------
                                   salesquantity
                                      Model-1
---------------------------------------------------------
bops_in_effect                        -0.25***
                                      (0.01)

bops_user                              0.01
                                      (0.01)

purchase_time_period                   0.10***
                                      (0.0004)

est_income_code                        0.01***
                                      (0.001)

age_band                              -0.0003
                                      (0.001)

female                                 0.31***
                                      (0.005)

factor(store_number)6                  0.01
                                      (0.01)

factor(store_number)5998              -0.44*
                                      (0.19)

homeowner_coden                        0.03***
                                      (0.01)

childn                                 0.002
                                      (0.005)

bops_in_effect:bops_user              -0.02
                                      (0.01)

Constant                               0.36***
                                      (0.01)

---------------------------------------------------------
Observations                          74,504
Log Likelihood                      -147,007.20
Akaike Inf. Crit.                   294,038.50
=========================================================
Note:                    *p<0.05; **p<0.01; ***p<0.001
```

Model fit assessment

poisson31a <- glm(salesquantity~1, data=sales2, family="poisson")

poisson31	Large glm (30 elements, 57.2 Mb)
poisson31a	Large glm (30 elements, 46.4 Mb)

##running a comparison with null model.

lrtest(poisson31, poisson31a)

```
Likelihood ratio test

Model 1: salesquantity ~ bops_in_effect * bops_user + purchase_time_period +
    est_income_code + age_band + female + factor(store_number) +
    homeowner_coden + childn
Model 2: salesquantity ~ 1
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  12 -147007
2   1 -181001 -11 67988  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

##Since Poisson doesn't fit the data, we will check for negative binomial model.

##This model has child as control variable

```
negbin31 <- glm.nb(salesquantity~ bops_in_effect*bops_user + purchase_time_period +
est_income_code + age_band + female + factor(store_number) + homeowner_coden + childn, data =
sales2)

stargazer(negbin31,
    apply.coef = exp, t.auto=F, p.auto = F,
    title="Negative Bionomial Results", type="text",
    column.labels=c("IRRs negbin31"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

```
Negative Bionomial Results
=====================================================
                              Dependent variable:
                         ----------------------------
                                salesquantity
                                IRRs negbin31
-----------------------------------------------------
bops_in_effect                    0.8274***
                                  (0.0071)

bops_user                          1.0052
                                  (0.0111)

purchase_time_period              1.1150***
                                  (0.0006)

est_income_code                   1.0077***
                                  (0.0015)

age_band                           1.0019*
                                  (0.0008)

female                            1.3330***
                                  (0.0061)

factor(store_number)6              1.0162
                                  (0.0118)

factor(store_number)5998           0.6717
                                  (0.2109)

homeowner_coden                   1.0268***
                                  (0.0072)

childn                             1.0061
                                  (0.0063)

bops_in_effect:bops_user           0.9671*
                                  (0.0154)

Constant                          1.3895***
                                  (0.0105)

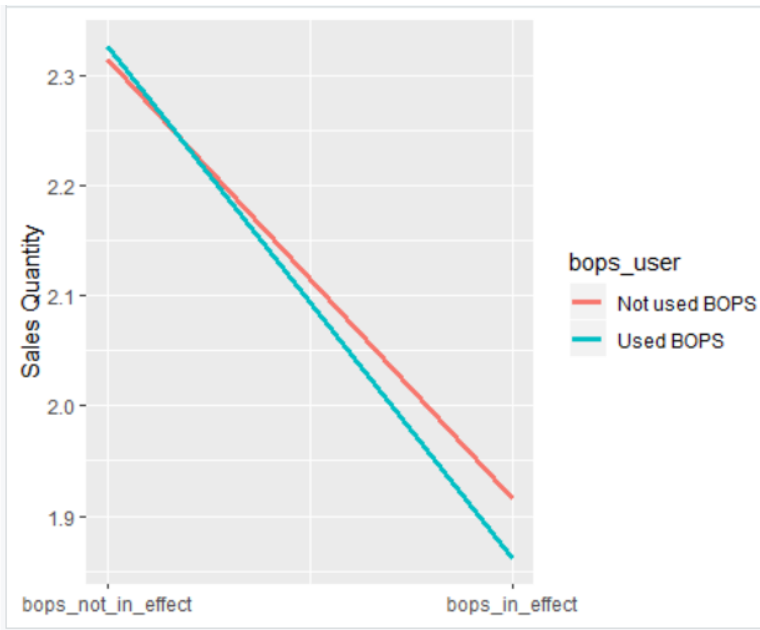
-----------------------------------------------------
Observations                        74,504
Log Likelihood                  -132,795.6000
theta                       4.5130*** (0.0504)
Akaike Inf. Crit.                265,615.2000
=====================================================
Note:                   *p<0.05; **p<0.01; ***p<0.001
```

Model fit assessment - Negative Binomial

```
negbin31a <- glm.nb(salesquantity ~ 1, data = sales2)
```

negbin31	Large negbin (30 elements, 48.7 Mb)
negbin31a	Large negbin (28 elements, 37.8 Mb)

```
lrtest(negbin31, negbin31a)
```

```
Likelihood ratio test

Model 1: salesquantity ~ bops_in_effect * bops_user + purchase_time_period +
    est_income_code + age_band + female + factor(store_number) +
    homeowner_coden + childn
Model 2: salesquantity ~ 1
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  13 -132795
2   2 -149589 -11 33588  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Check for heteroskedasticity
gqtest(negbin31) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(negbin31) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  negbin31
GQ = 0.3098, df1 = 37240, df2 = 37240, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin31) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  negbin31
BP = 226.51, df = 11, p-value < 2.2e-16
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(negbin31, type="HC1")))

Values	
HWrobstder	Named num [1:12] 0.0149 0.01043 0.02024 ...

produces Huber-White robust standard errors

```
stargazer(negbin31, negbin31,
     se=list(NULL, HWrobstder),
     apply.coef = exp, t.auto=F, p.auto = F,
     title="Negative Binomial Results", type="text",
     column.labels=c("Normal SE", "HW-Robust SE"),
     df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable: salesquantity	
	Normal SE (1)	HW-Robust SE (2)
bops_in_effect	0.827*** (0.007)	0.827*** (0.010)
bops_user	1.005 (0.011)	1.005 (0.020)
purchase_time_period	1.115*** (0.001)	1.115*** (0.001)
est_income_code	1.008*** (0.001)	1.008*** (0.002)
age_band	1.002* (0.001)	1.002* (0.001)
female	1.333*** (0.006)	1.333*** (0.009)
factor(store_number)6	1.016 (0.012)	1.016 (0.014)
factor(store_number)5998	0.672 (0.211)	0.672 (0.080)
homeowner_coden	1.027*** (0.007)	1.027*** (0.010)
childn	1.006 (0.006)	1.006 (0.009)
bops_in_effect:bops_user	0.967* (0.015)	0.967* (0.025)
Constant	1.389*** (0.010)	1.389*** (0.015)
Observations	74,504	74,504
Log Likelihood	-132,795.600	-132,795.600
theta	4.513*** (0.050)	4.513*** (0.050)
Akaike Inf. Crit.	265,615.200	265,615.200
Note:	*p<0.05; **p<0.01; ***p<0.001	

# Visualize the output- generated a data frame at two different values of bops_in_effect for two groups - who used BOPS and who didn't after BOPS implementation.
meffects31 <- ggpredict(negbin31, terms=c("bops_in_effect","bops_user"))

BA Project_BOPS_Final Submitted.Rmd* | meffects31 | df3

Filter

	x	predicted	conf.low	conf.high	group
1	0	2.314271	2.292138	2.336617	0
2	0	2.326389	2.281006	2.372674	1
3	1	1.914841	1.895670	1.934206	0
4	1	1.861633	1.825557	1.898422	1

```
ggplot(meffects31,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("bops_in_effect") + ylab("Sales Quantity") +
  labs(colour="bops_user") +
  scale_colour_discrete(labels=c("Not used BOPS","Used BOPS")) +
  scale_x_continuous(breaks=c(0,1), labels=c("bops_not_in_effect", "bops_in_effect")) +
  theme(axis.title.x=element_blank())
```



#Interpretation - The interaction coefficient for bops_in_effect:bops_user is significant with value (0.967). This means that the usage of BOPS service is associated with ~3.3% decrease in sales quantity.

####===================== IMPACT ON SALES VALUE==========================####

```
ols32 = lm(log_salesvalue~ bops_in_effect*bops_user + purchase_time_period + est_income_code +
age_band + factor(store_number) + female + homeowner_coden +childn, data=sales2)
stargazer(ols32,
        title="Regression Results", type="text",
        column.labels=c("Model-OLS32","Model-OLS31"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
=======================================================
                              Dependent variable:
                          -------------------------
                                log_salesvalue
                                  Model-OLS32
-------------------------------------------------------
bops_in_effect                    -0.15***
                                   (0.01)

bops_user                         -0.22***
                                   (0.01)

purchase_time_period               0.10***
                                   (0.001)

est_income_code                    0.01***
                                   (0.002)

age_band                           0.01***
                                   (0.001)

factor(store_number)6             -0.08***
                                   (0.01)

factor(store_number)5998           0.20
                                   (0.20)

female                            -0.15***
                                   (0.01)

homeowner_coden                   -0.001
                                   (0.01)

childn                            -0.04***
                                   (0.01)
```

```
childn                                      -0.04***
                                             (0.01)

bops_in_effect:bops_user                     -0.04
                                             (0.02)

Constant                                     5.15***
                                             (0.01)

-------------------------------------------------------
Observations                                 74,504
R2                                            0.17
Adjusted R2                                   0.17
Residual Std. Error                           0.97
F Statistic                                1,418.85***
=======================================================
Note:                       *p<0.05; **p<0.01; ***p<0.001
>
```

Check for heteroskedasticity
gqtest(ols32) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols32) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  ols32
GQ = 0.92642, df1 = 37240, df2 = 37240, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols32) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols32
BP = 191.48, df = 11, p-value < 2.2e-16
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols32, type="HC1"))) # produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:12] 0.01192 0.00798 0.01321...

stargazer(ols32, ols32,
se=list(NULL, HWrobstder),
title="Regression Results", type="text",

column.labels=c("Normal SE", "HW-Robust SE"),
df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))

```
Regression Results
=========================================================
                              Dependent variable:
                         --------------------------------
                                 log_salesvalue
                           Normal SE       HW-Robust SE
                              (1)              (2)
---------------------------------------------------------
bops_in_effect             -0.146***        -0.146***
                            (0.008)          (0.008)

bops_user                  -0.216***        -0.216***
                            (0.013)          (0.013)

purchase_time_period        0.105***         0.105***
                            (0.001)          (0.001)

est_income_code             0.006***         0.006***
                            (0.002)          (0.002)

age_band                    0.007***         0.007***
                            (0.001)          (0.001)

factor(store_number)6      -0.079***        -0.079***
                            (0.014)          (0.014)

factor(store_number)5998    0.201            0.201
                            (0.202)          (0.146)

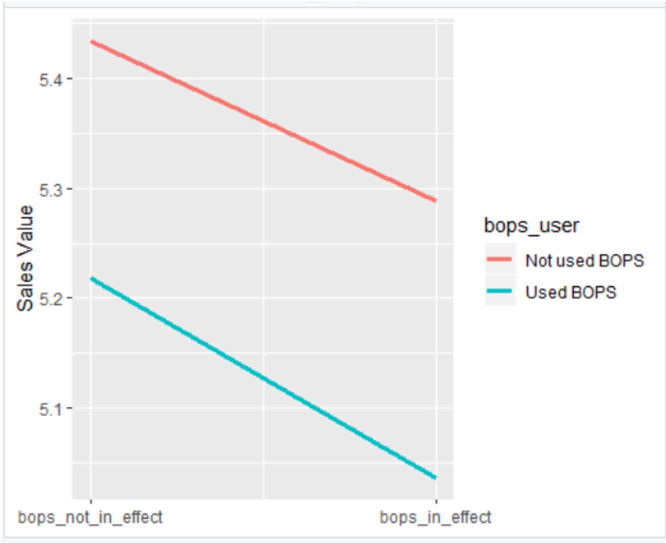
female                     -0.153***        -0.153***
                            (0.007)          (0.007)

homeowner_coden            -0.001           -0.001
                            (0.008)          (0.008)

childn                     -0.042***        -0.042***
                            (0.007)          (0.007)

bops_in_effect:bops_user   -0.035           -0.035
                            (0.018)          (0.019)

Constant                    5.145***         5.145***
                            (0.012)          (0.012)

---------------------------------------------------------
Observations                74,504           74,504
R2                          0.173            0.173
Adjusted R2                 0.173            0.173
Residual Std. Error         0.969            0.969
F Statistic              1,418.846***     1,418.846***
=========================================================
Note:                    *p<0.05; **p<0.01; ***p<0.001
```

Visualize the output

meffects32 <- ggpredict(ols32, terms=c("bops_in_effect","bops_user"))

	x	predicted	conf.low	conf.high	group
1	0	5.434436	5.423210	5.445663	0
2	0	5.218199	5.195117	5.241281	1
3	1	5.288045	5.276878	5.299211	0
4	1	5.036537	5.013288	5.059787	1

#generates a tidy data frame at two different values of bops_in_effect for two groups - who used BOPS and who didn't after BOPS implementation.

```
ggplot(meffects32,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("bops_in_effect") + ylab("Sales Value") +
  labs(colour="bops_user") +
  scale_colour_discrete(labels=c("Not used BOPS","Used BOPS")) +
  scale_x_continuous(breaks=c(0,1), labels=c("bops_not_in_effect", "bops_in_effect")) +
  theme(axis.title.x=element_blank())
```



#Interpretation - The interaction coefficient for bops_in_effect:bops_user is insignificant. Thus, we cannot comment about the impact of BOPS usage on sales quantity.

####==========================APPROACH 2=======================================####
##Removing Female as control variable, removing rows where homeonwer and child are missing.
#Here we are removing data where "child" and "homeowner_code" have NA's.
sales21 <- cust_Level[which((cust_Level$child =="Y" | cust_Level$child =="N"),(cust_Level$homeowner_code=="R" | cust_Level$homeowner_code=="O")),]

#Droppping female as control behavior
sales21$female <- NULL

stargazer(sales21, type="text", median=TRUE, iqr=TRUE, digits=1, title="Descriptive Statistics")

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	81,434	62,707,033,354.0	201,169,744,468.0	103,465	18,356,040	25,988,667	27,748,316	919,600,001,722
store_number	81,434	4.4	113.1	2	2	2	2	5,998
age_band	81,418	4.9	3.9	0.0	0.0	5.0	8.0	13.0
est_income_code	81,434	5.5	2.2	1	4	6	7	9
length_of_residence	81,434	7.3	5.4	0	2	6	13	15
bops_in_effect	81,434	0.5	0.5	0	0	0.5	1	1
salesvalue	81,434	391.3	940.1	0.0	98.7	194.6	386.5	82,226.2
salesquantity	81,434	2.5	7.8	1	1	1	3	1,474
purchase_time_period	81,434	2.9	4.3	1	1	1	2	24
bops_user	81,434	0.2	0.4	0	0	0	0	1
childn	81,434	0.4	0.5	0	0	0	1	1
homeowner_coden	81,434	0.7	0.5	0	0	1	1	1

Replacing NA's with mean values for other variables.

sales21$age_band[is.na(sales21$age_band)] = median(sales21$age_band, na.rm = TRUE)
sales21$est_income_code[is.na(sales21$est_income_code)] <- median(sales21$est_income_code, na.rm = TRUE)
sales21$length_of_residence[is.na(sales21$length_of_residence)] <- median(sales21$length_of_residence, na.rm = TRUE)

stargazer(sales21, type="text", median=TRUE, iqr=TRUE, digits=1, title="Descriptive Statistics")

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	81,434	62,707,033,354.0	201,169,744,468.0	103,465	18,356,040	25,988,667	27,748,316	919,600,001,722
store_number	81,434	4.4	113.1	2	2	2	2	5,998
age_band	81,434	4.9	3.9	0	0	5	8	13
est_income_code	81,434	5.5	2.2	1	4	6	7	9
length_of_residence	81,434	7.3	5.4	0	2	6	13	15
bops_in_effect	81,434	0.5	0.5	0	0	0.5	1	1
salesvalue	81,434	391.3	940.1	0.0	98.7	194.6	386.5	82,226.2
salesquantity	81,434	2.5	7.8	1	1	1	3	1,474
purchase_time_period	81,434	2.9	4.3	1	1	1	2	24
bops_user	81,434	0.2	0.4	0	0	0	0	1
childn	81,434	0.4	0.5	0	0	0	1	1
homeowner_coden	81,434	0.7	0.5	0	0	1	1	1

####========================IMPACT ON SALES QUANTITY=========================####

poisson32 <- glm(salesquantity~ bops_in_effect*bops_user + purchase_time_period + est_income_code + age_band + factor(store_number) + homeowner_coden +childn, family="poisson", data=sales21)

stargazer(poisson32,
title="Poisson Results", type="text",
column.labels=c("Model-1"),
df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))

Poisson Results

	Dependent variable:
	salesquantity Model-1
bops_in_effect	-0.30*** (0.01)
bops_user	0.04*** (0.01)
purchase_time_period	0.11*** (0.0004)
est_income_code	0.02*** (0.001)
age_band	-0.01*** (0.001)
factor(store_number)6	0.01 (0.01)
factor(store_number)5998	-0.25 (0.14)
homeowner_coden	-0.01* (0.01)
childn	-0.01* (0.005)
bops_in_effect:bops_user	-0.02* (0.01)
Constant	0.54*** (0.01)
Observations	81,434
Log Likelihood	-181,090.90
Akaike Inf. Crit.	362,203.80
Note:	*p<0.05; **p<0.01; ***p<0.001

Model fit assessment

poisson32a <- glm(salesquantity~1, data=sales21, family="poisson")

poisson32	Large glm (30 elements, 61 Mb)
poisson32a	Large glm (30 elements, 50.1 Mb)

##running a comparison with null model.

lrtest(poisson32, poisson32a)

```
Likelihood ratio test

Model 1: salesquantity ~ bops_in_effect * bops_user + purchase_time_period +
    est_income_code + age_band + factor(store_number) + homeowner_coden +
    childn
Model 2: salesquantity ~ 1
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  11 -181091
2   1 -222022 -10 81863  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

##Since Poisson doesn't fit the data, we will check for negative binomial model.

negbin32 <- glm.nb(salesquantity~ bops_in_effect*bops_user + purchase_time_period + est_income_code + age_band + factor(store_number) + homeowner_coden + childn, data = sales21)

stargazer(negbin32,
    apply.coef = exp, t.auto=F, p.auto = F,
    title="Negative Bionomial Results", type="text",
    column.labels=c("IRRs negbin31"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))

```
Negative Bionomial Results
=====================================================
                          Dependent variable:
                     --------------------------------
                            salesquantity
                            IRRs negbin31
-----------------------------------------------------
bops_in_effect                0.8115***
                             (0.0070)

bops_user                     1.0308**
                             (0.0109)

purchase_time_period          1.1250***
                             (0.0006)

est_income_code               1.0128***
                             (0.0014)

age_band                      0.9990
                             (0.0008)

factor(store_number)6         1.0238*
                             (0.0119)

factor(store_number)5998      0.7535
                             (0.1800)

homeowner_coden               0.9964
                             (0.0071)

childn                        1.0023
                             (0.0063)

bops_in_effect:bops_user      0.9586**
                             (0.0153)

Constant                      1.5909***
                             (0.0098)

-----------------------------------------------------
Observations                  81,434
Log Likelihood              -149,141.0000
theta                       3.5620*** (0.0328)
Akaike Inf. Crit.           298,304.0000
=====================================================
Note:               *p<0.05; **p<0.01; ***p<0.001
```

Model fit assessment - Negative Binomial

negbin32a <- glm.nb(salesquantity ~ 1, data = sales21)

negbin32	Large negbin (30 elements, 52.3 Mb)
negbin32a	Large negbin (28 elements, 41.4 Mb)

lrtest(negbin32, negbin32a)

```
Likelihood ratio test

Model 1: salesquantity ~ bops_in_effect * bops_user + purchase_time_period +
    est_income_code + age_band + factor(store_number) + homeowner_coden +
    childn
Model 2: salesquantity ~ 1
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  12 -149140
2   2 -167228 -10 36177  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Check for heteroskedasticity
gqtest(negbin32) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(negbin32) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  negbin32
GQ = 0.13122, df1 = 40706, df2 = 40706, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin32) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  negbin32
BP = 34.768, df = 10, p-value = 0.0001368
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(negbin32, type="HC1")))
# produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:11] 0.01883 0.02196 0.02345...

```
stargazer(negbin32, negbin32,
        se=list(NULL, HWrobstder),
        apply.coef = exp, t.auto=F, p.auto = F,
        title="Negative Binomial Results", type="text",
        column.labels=c("Normal SE", "HW-Robust SE"),
        df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
Negative Binomial Results
=============================================================
                                  Dependent variable:
                              -------------------------------
                                      salesquantity
                                  Normal SE      HW-Robust SE
                                     (1)              (2)
-------------------------------------------------------------
bops_in_effect                    0.811***          0.811***
                                  (0.007)           (0.022)

bops_user                         1.031**           1.031**
                                  (0.011)           (0.023)

purchase_time_period              1.125***          1.125***
                                  (0.001)           (0.002)

est_income_code                   1.013***          1.013***
                                  (0.001)           (0.004)

age_band                           0.999             0.999
                                  (0.001)           (0.002)

factor(store_number)6              1.024*            1.024*
                                  (0.012)           (0.014)

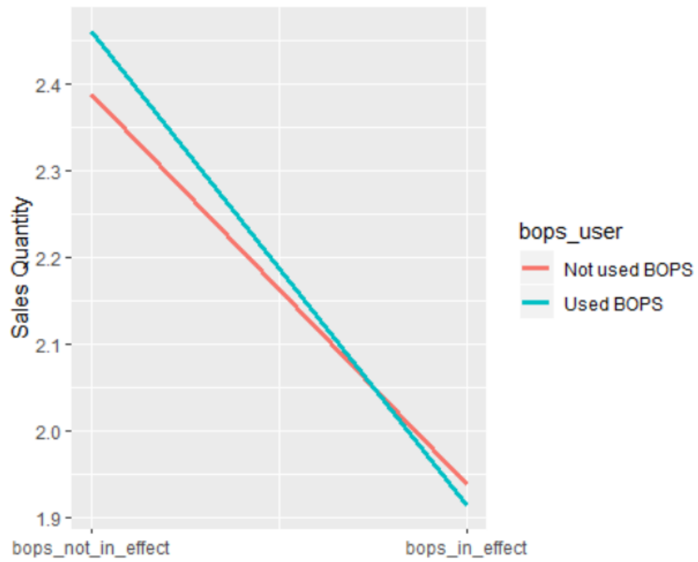
factor(store_number)5998           0.753             0.753
                                  (0.180)           (0.122)

homeowner_coden                    0.996             0.996
                                  (0.007)           (0.022)
```

```
homeowner_coden                    0.996             0.996
                                  (0.007)           (0.022)

childn                             1.002             1.002
                                  (0.006)           (0.012)

bops_in_effect:bops_user          0.959**           0.959**
                                  (0.015)           (0.031)

Constant                          1.591***          1.591***
                                  (0.010)           (0.019)

-------------------------------------------------------------
Observations                       81,434            81,434
Log Likelihood                  -149,141.000      -149,141.000
theta                        3.562*** (0.033) 3.562*** (0.033)
Akaike Inf. Crit.                298,304.000       298,304.000
=============================================================
Note:                          *p<0.05; **p<0.01; ***p<0.001
```

Visualize the output- generated a data frame at two different values of bops_in_effect for two groups - who used BOPS and who didn't after BOPS implementation.

meffects33 <- ggpredict(negbin32, terms=c("bops_in_effect","bops_user"))

Filter

	x	predicted	conf.low	conf.high	group
1	0	2.386680	2.364058	2.409517	0
2	0	2.460138	2.413039	2.508157	1
3	1	1.936744	1.917495	1.956187	0
4	1	1.913799	1.877051	1.951265	1

```
ggplot(meffects33,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("bops_in_effect") + ylab("Sales Quantity") +
  labs(colour="bops_user") +
  scale_colour_discrete(labels=c("Not used BOPS","Used BOPS")) +
  scale_x_continuous(breaks=c(0,1), labels=c("bops_not_in_effect", "bops_in_effect")) +
  theme(axis.title.x=element_blank())
```



#Interpretation of approach 2 - The interaction coefficient for bops_in_effect:bops_user is significant with value (0.959). This means that the usage of BOPS service is associated with ~4.1% decrease in sales quantity.

####=====================IMPACT ON SALES VALUE=====================####

sales21$log_salesvalue <- log(1+sales21$salesvalue)


BA Project_BOPS_Final Submitted.Rmd × | sales21 ×

Filter

_of_residence	child	bops_in_effect	salesvalue	salesquantity	purchase_time_period	bops_user	childn	homeowner_coden	log_salesvalue
15	Y	0	195.30	1	1	0	1	0	5.279644
15	Y	1	179.00	1	1	0	1	0	5.192957
3	Y	0	72.00	1	1	0	1	1	4.290459
3	Y	1	90.16	2	4	0	1	1	4.512616
15	N	0	2898.00	2	7	0	0	1	7.972121
15	N	1	1873.01	8	21	0	0	1	7.535836

Showing 1 to 7 of 81,434 entries

#since salesvalue is 0 at some places, we are taking log of (1+salesvalue)

```
ols33 = lm(log_salesvalue~ bops_in_effect*bops_user + purchase_time_period + est_income_code +
age_band + factor(store_number) + homeowner_coden +childn, data=sales21)
stargazer(ols33,
        title="Regression Results", type="text",
        column.labels=c("Model-OLS33","Model-OLS31"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
======================================================
                              Dependent variable:
                          ---------------------------
                               log_salesvalue
                                 Model-OLS33
------------------------------------------------------
bops_in_effect                    -0.14***
                                   (0.01)

bops_user                         -0.22***
                                   (0.01)

purchase_time_period               0.11***
                                   (0.001)

est_income_code                    0.01***
                                   (0.002)

age_band                           0.01***
                                   (0.001)

factor(store_number)6             -0.10***
                                   (0.01)

factor(store_number)5998           0.34
                                   (0.18)

homeowner_coden                   -0.004
                                   (0.01)

childn                                              -0.05***
                                                     (0.01)

bops_in_effect:bops_user                            -0.04*
                                                     (0.02)

Constant                                             5.08***
                                                     (0.01)

------------------------------------------------------
Observations                                         81,434
R2                                                    0.17
Adjusted R2                                           0.17
Residual Std. Error                                   0.98
F Statistic                                       1,689.92***
======================================================
Note:                         *p<0.05; **p<0.01; ***p<0.001
>
```

Check for heteroskedasticity
gqtest(ols33) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols33) # Breusch-Pagan test indicates heteroskedasticity

```
> gqtest(ols33) # Goldfeld-Quandt test indicates no heteroskedasticity

        Goldfeld-Quandt test

data:  ols33
GQ = 0.92765, df1 = 40706, df2 = 40706, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols33) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols33
BP = 54.836, df = 10, p-value = 0.00000003389
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols33, type="HC1"))) # produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:11] 0.010905 0.00769 0.0127…

```
stargazer(ols33, ols33,
     se=list(NULL, HWrobstder),
     title="Regression Results", type="text",
     column.labels=c("Normal SE", "HW-Robust SE"),
     df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
=========================================================
                              Dependent variable:
                         ----------------------------
                                log_salesvalue
                          Normal SE      HW-Robust SE
                             (1)             (2)
---------------------------------------------------------
bops_in_effect            -0.144***       -0.144***
                           (0.008)         (0.008)

bops_user                 -0.222***       -0.222***
                           (0.012)         (0.013)

purchase_time_period       0.106***        0.106***
                           (0.001)         (0.001)

est_income_code            0.008***        0.008***
                           (0.002)         (0.002)

age_band                   0.005***        0.005***
                           (0.001)         (0.001)

factor(store_number)6     -0.098***       -0.098***
                           (0.014)         (0.013)

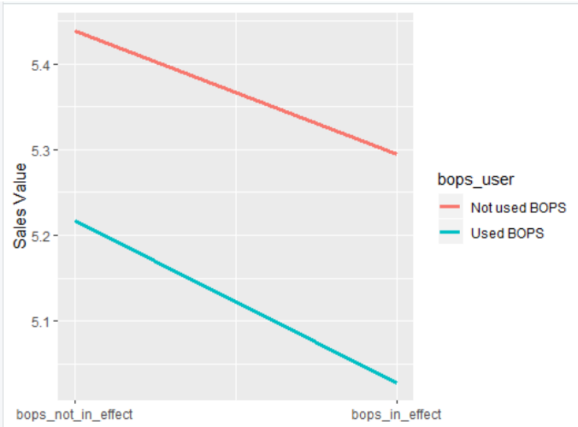
factor(store_number)5998    0.341           0.341**
                           (0.181)         (0.128)

homeowner_coden            -0.004          -0.004
                           (0.008)         (0.008)

childn                    -0.045***       -0.045***
                           (0.007)         (0.007)

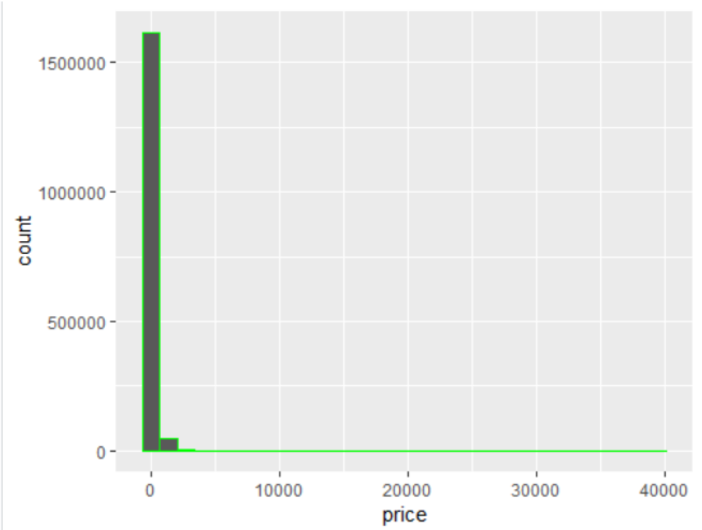
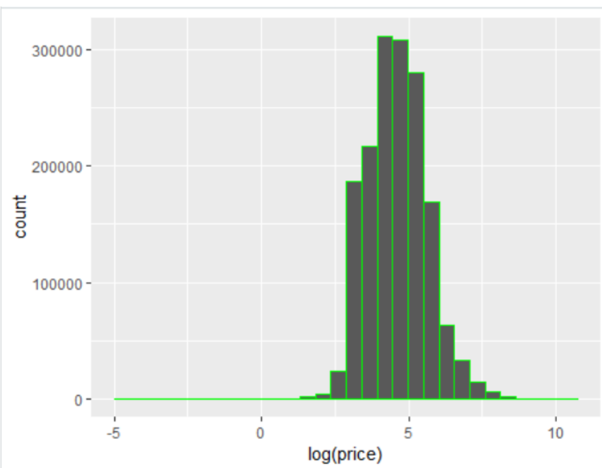
bops_in_effect:bops_user  -0.044*         -0.044*
                           (0.018)         (0.018)

Constant                   5.083***        5.083***
                           (0.011)         (0.011)

---------------------------------------------------------
Observations               81,434          81,434
R2                          0.172           0.172
Adjusted R2                 0.172           0.172
Residual Std. Error         0.977           0.977
F Statistic              1,689.924***    1,689.924***
=========================================================
Note:                   *p<0.05; **p<0.01; ***p<0.001
```

# Visualize the output
meffects34 <- ggpredict(ols33, terms=c("bops_in_effect","bops_user"))

BA Project_BOPS_Final Submitted.Rmd | meffects34 | sales21

Filter

	x	predicted	conf.low	conf.high	group
1	0	5.438957	5.428127	5.449787	0
2	0	5.216780	5.194715	5.238845	1
3	1	5.294682	5.283919	5.305445	0
4	1	5.028070	5.005819	5.050321	1

#generates a tidy data frame at two different values of bops_in_effect for two groups - who used BOPS and who didn't after BOPS implementation.

```
ggplot(meffects34,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("bops_in_effect") + ylab("Sales Value") +
  labs(colour="bops_user") +
  scale_colour_discrete(labels=c("Not used BOPS","Used BOPS")) +
  scale_x_continuous(breaks=c(0,1), labels=c("bops_not_in_effect", "bops_in_effect")) +
  theme(axis.title.x=element_blank())
```



#Interpretation of approach 2 - The interaction coefficient for bops_in_effect:bops_user is significant with value (-0.044). This means that the usage of BOPS service is associated with ~4.4% decrease in sales value.

#We got different results with Approach-1 (removing rows where female is NA) and Approach-2 (dropping female as control variable). Impact on sales value of BOPS usage was insignificant with Approach-1. However, it came out significant with Approach-2.

```
#======================================================================================#
## Question4: What is the impact of using the BOPS service on online customer return behavior?
#======================================================================================#
```{r}

#Descriptive Statistics
stargazer(trans_level, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")
```

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	1,671,502	29,441,376,672.0	141,529,202,622.0	100,348	25,694,206	30,152,198	34,992,529	919,650,001,519
transaction_id	1,671,502	3,338,930.0	817,965.5	7,198	2,649,367	3,355,444	4,030,256.0	4,702,552
store_number	1,671,502	229.5	1,145.0	2	2	2	2	5,998
price	1,671,502	170.3	319.5	0.0	45.1	90.0	189.0	39,422.0
sku	1,671,502	18,300,993.0	1,587,131.0	10,033,405	17,859,869	18,126,698	18,584,417	80,006,100
return	1,671,502	0.1	0.3	0	0	0	0	1
age_band	1,599,464	5.0	3.9	0.0	0.0	5.0	8.0	13.0
est_income_code	1,602,838	5.4	2.3	1.0	4.0	6.0	7.0	9.0
length_of_residence	1,602,838	7.1	5.4	0.0	2.0	6.0	13.0	15.0
year	1,671,502	2,011.6	1.0	2,010	2,011	2,012	2,012	2,013
month_index	1,671,502	31.5	10.4	13	22	31	41	48
product_category	1,671,498	10.1	6.5	1.0	5.0	9.0	12.0	21.0
month_dummy	1,671,502	7.4	4.0	1	4	8	12	12
week_index	1,671,502	83.0	45.0	1	41	81	124	157
bops	1,170,568	0.2	0.4	0.0	0.0	0.0	0.0	1.0
female	1,413,233	0.5	0.5	0.0	0.0	0.0	1.0	1.0

ggplot(trans_level, aes(x=price)) + geom_histogram(colour="green", bins = 30)
```



ggplot(trans_level, aes(x=log(price))) + geom_histogram(colour="green", bins = 30)



#Histogram plot shows that log-transformed price look more like a normal. Therefore, we will use log-transformed price, instead of the raw price.

trans_level$logprice <- log(1+trans_level$price)

BA Project_BOPS_Final Submitted.Rmd | trans_level | meffects34 | sales21

Filter

length_of_residence	child	year	month	month_index	product_category	month_dummy	week_index	bops	female	logprice
2	Y	2012	MAR	32	5	3	84	0	0	5.959845
2	Y	2012	MAR	32	5	3	84	0	0	7.120493
4	N	2010	AUG	13	4	8	1	NA	1	5.309158
4	Y	2010	AUG	13	5	8	1	NA	0	5.298317
12	N	2010	AUG	13	5	8	1	NA	1	3.400530
14	Y	2010	AUG	13	5	8	1	NA	1	5.385183
5	N	2010	AUG	13	5	8	1	NA	0	4.469007
5	N	2010	AUG	13	5	8	1	NA	0	3.875566

Showing 1 to 9 of 1,671,502 entries

To take care about the negative values, scaling doesn't have any impact

Creating a subset to analyze the impact of using BOPS service when BOPS was in effect

trans_level1 <- subset(trans_level, trans_level$bops == 1 | trans_level$bops == 0)

Converting child variable to a dummy variable.
trans_level1$child = ifelse(trans_level1$child== "Y",1,0)
Converting homeowner_code to a dummy variable.
trans_level1$homeowner_code = ifelse(trans_level1$homeowner_code== "O",1,0)

# 183466 records (~15.6% of trans_level1) of female variable have NA values
table(trans_level1$female == 'NA')

```
 FALSE
987102
```

To replace NA values of est_income_code, age band and length of residence with their median values
trans_level1$est_income_code[is.na(trans_level1$est_income_code)] <- median(trans_level1$est_income_code, na.rm = TRUE)
trans_level1$age_band[is.na(trans_level1$age_band)] <- median(trans_level1$age_band, na.rm = TRUE)
trans_level1$length_of_residence[is.na(trans_level1$length_of_residence)] <- median(trans_level1$length_of_residence, na.rm = TRUE)

stargazer(trans_level1, type="text", median=TRUE, iqr=TRUE,digits=2, title="Descriptive Statistics")

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	1,170,568	23,247,802,315.00	125,423,860,663.00	100,348	28,454,419	31,964,480	35,423,835	919,650,001,519
transaction_id	1,170,568	3,751,126.00	561,621.20	50,002	3,262,163.0	3,787,321	4,236,822.0	4,702,552
store_number	1,170,568	180.36	1,017.89	2	2	2	2	5,998
price	1,170,568	170.16	323.39	0.00	44.24	89.99	186.99	39,422.00
sku	1,170,568	18,424,852.00	1,380,845.00	11,373,024	17,951,286	18,214,460	18,697,427	80,006,100
return	1,170,568	0.10	0.30	0	0	0	0	1
age_band	1,170,568	4.84	3.89	0	0	5	8	13
est_income_code	1,170,568	5.41	2.23	1	4	6	7	9
homeowner_code	1,170,568	0.64	0.48	0	0	1	1	1
length_of_residence	1,170,568	7.14	5.34	0	2	6	13	15
child	1,170,568	0.39	0.49	0	0	0	1	1
year	1,170,568	2,012.05	0.71	2,011	2,012	2,012	2,013	2,013
month_index	1,170,568	37.03	6.62	25	30	39	42	48
product_category	1,170,564	10.10	6.47	1.00	5.00	9.00	12.00	21.00
month_dummy	1,170,568	7.39	4.04	1	4	8	12	12
week_index	1,170,568	107.13	28.66	53	78	117	129	157
bops	1,170,568	0.24	0.43	0	0	0	0	1
female	987,102	0.49	0.50	0.00	0.00	0.00	1.00	1.00
logprice	1,170,568	4.47	1.25	0.00	3.81	4.51	5.24	10.58

# Checking for multi-collinearity
df41=trans_level1[c("bops","month_dummy","product_category","female","logprice","year", "age_band", "est_income_code")]

BA Project_BOPS_Final Submitted.Rmd | df41 | trans_level1 | trans_level | meffects34 | sales21

Filter

	bops	month_dummy	product_category	female	logprice	year	age_band	est_income_code
434241	0	8	4	0	7.067448	2011	11	6
434242	0	8	12	1	3.044046	2011	8	9
434243	0	8	12	0	3.753027	2011	0	5
434244	0	8	6	NA	5.525413	2011	0	5
434245	0	8	21	1	4.930004	2011	10	6
434246	0	8	4	0	5.268116	2011	4	2

Showing 56 to 62 of 1,170,568 entries

round(cor(df41),3)

```
> round(cor(df41),3)
                   bops month_dummy product_category female logprice   year age_band est_income_code
bops              1.000      -0.006               NA     NA   -0.021  0.086   -0.027          -0.013
month_dummy      -0.006       1.000               NA     NA    0.026 -0.656    0.057           0.051
product_category     NA          NA                1     NA       NA     NA       NA              NA
female               NA          NA               NA      1       NA     NA       NA              NA
logprice         -0.021       0.026               NA     NA    1.000 -0.065   -0.009          -0.003
year              0.086      -0.656               NA     NA   -0.065  1.000   -0.046          -0.038
age_band         -0.027       0.057               NA     NA   -0.009 -0.046    1.000           0.169
est_income_code  -0.013       0.051               NA     NA   -0.003 -0.038    0.169           1.000
```

vifcor(df41)

```
> vifcor(df41)
No variable from the 8 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( age_band ~ product_category ):  0.0007822897
max correlation ( year ~ month_dummy ):  -0.6539523

---------- VIFs of the remained variables --------
         Variables      VIF
1             bops 1.017853
2      month_dummy 1.765473
3 product_category 1.009632
4           female 1.029113
5         logprice 1.033008
6             year 1.778289
7         age_band 1.017496
8  est_income_code 1.015664
```

#To remove NA value from female

trans_level2 <- subset(trans_level1, trans_level1$female!= 'NA')

stargazer(trans_level2, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	987,102	25,917,443,507.0	132,002,375,089.0	100,348	27,948,575.0	31,773,151.0	35,404,273.0	919,650,001,519
transaction_id	987,102	3,721,998.0	560,640.4	50,002	3,223,963.0	3,746,786	4,200,095.0	4,702,383
store_number	987,102	3.4	82.1	2	2	2	2	5,998
price	987,102	170.1	327.4	0.0	44.1	90.0	186.1	39,422.0
sku	987,102	18,417,145.0	1,402,656.0	11,373,024	17,945,767	18,209,999	18,688,697	80,006,100
return	987,102	0.1	0.3	0	0	0	0	1
age_band	987,102	5.4	3.8	0	3	5	8	13
est_income_code	987,102	5.5	2.2	1	4	6	7	9
homeowner_code	987,102	0.7	0.5	0	0	1	1	1
length_of_residence	987,102	7.5	5.4	0	2	6	14	15
child	987,102	0.4	0.5	0	0	0	1	1
year	987,102	2,012.0	0.7	2,011	2,012	2,012	2,013	2,013
month_index	987,102	36.7	6.6	25	30	39	42	48
product_category	987,099	10.1	6.5	1.0	5.0	9.0	12.0	21.0
month_dummy	987,102	7.5	4.0	1	4	8	12	12
week_index	987,102	105.6	28.7	53	75	114	128	157
bops	987,102	0.2	0.4	0	0	0	0	1
female	987,102	0.5	0.5	0	0	0	1	1
logprice	987,102	4.5	1.2	0.0	3.8	4.5	5.2	10.6

To remove 3 blank records of product category as the number of records are less and we cannot replace the missing values by its mean or median

trans_level2 <- na.omit(trans_level2)

stargazer(trans_level2, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
customer_id	987,099	25,917,522,176.0	132,002,567,967.0	100,348	27,948,555.0	31,773,150	35,404,274.0	919,650,001,519
transaction_id	987,099	3,721,998.0	560,639.4	50,002	3,223,963	3,746,785	4,200,093.0	4,702,383
store_number	987,099	3.4	82.1	2	2	2	2	5,998
price	987,099	170.1	327.4	0.0	44.1	90.0	186.1	39,422.0
sku	987,099	18,417,080.0	1,401,287.0	11,373,024	17,945,767	18,209,999	18,688,697	40,005,720
return	987,099	0.1	0.3	0	0	0	0	1
age_band	987,099	5.4	3.8	0	3	5	8	13
est_income_code	987,099	5.5	2.2	1	4	6	7	9
homeowner_code	987,099	0.7	0.5	0	0	1	1	1
length_of_residence	987,099	7.5	5.4	0	2	6	14	15
child	987,099	0.4	0.5	0	0	0	1	1
year	987,099	2,012.0	0.7	2,011	2,012	2,012	2,013	2,013
month_index	987,099	36.7	6.6	25	30	39	42	48
product_category	987,099	10.1	6.5	1	5	9	12	21
month_dummy	987,099	7.5	4.0	1	4	8	12	12
week_index	987,099	105.6	28.7	53	75	114	128	157
bops	987,099	0.2	0.4	0	0	0	0	1
female	987,099	0.5	0.5	0	0	0	1	1
logprice	987,099	4.5	1.2	0.0	3.8	4.5	5.2	10.6

Total number of records are same after cleansing the data

Checking multi collinearity

df42=trans_level2[c("bops","month_dummy","product_category","female","logprice","year", "age_band", "est_income_code")]

BA Project_BOPS_Final Submitted.Rmd | df42 | trans_level2 | df41 | trans_level1 | trans_level | meffects34

Filter

	bops	month_dummy	product_category	female	logprice	year	age_band	est_income_code
1	0	3	5	0	5.959845	2012	7	8
2	0	3	5	0	7.120493	2012	7	8
434187	0	8	12	1	3.693867	2011	9	5
434188	0	8	12	1	3.753027	2011	9	5
434189	0	8	5	1	4.117898	2011	5	6
434190	0	8	17	1	6.917616	2011	4	4

Showing 1 to 7 of 987,099 entries

round(cor(df42),3)

```
> round(cor(df42),3)
                   bops month_dummy product_category female logprice   year age_band est_income_code
bops              1.000      -0.006            0.003  0.041   -0.021  0.079   -0.032          -0.019
month_dummy      -0.006       1.000            0.020 -0.040    0.024 -0.653    0.048           0.049
product_category  0.003       0.020            1.000  0.027   -0.082 -0.031    0.005           0.014
female            0.041      -0.040            0.027  1.000   -0.142  0.035    0.068          -0.024
logprice         -0.021       0.024           -0.082 -0.142    1.000 -0.071   -0.011          -0.001
year              0.079      -0.653           -0.031  0.035   -0.071  1.000   -0.022          -0.033
age_band         -0.032       0.048            0.005  0.068   -0.011 -0.022    1.000           0.137
est_income_code  -0.019       0.049            0.014 -0.024   -0.001 -0.033    0.137           1.000
```

vifcor(df42)

```
> vifcor(df42)
No variable from the 8 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( logprice ~ bops ):  -0.001422741
max correlation ( year ~ month_dummy ):  -0.6560216

---------- VIFs of the remained variables --------
         Variables      VIF
1             bops 1.016310
2      month_dummy 1.772950
3 product_category 1.008077
4           female 1.030778
5         logprice 1.025189
6             year 1.785112
7         age_band 1.037572
8  est_income_code 1.027502
```

Creating factor variables of store number, year and month_dummy if we have to use the factor variable

```
trans_level2$f_store <- as.factor(trans_level2$store_number)
trans_level2$f_year <- as.factor(trans_level2$year)
trans_level2$f_month <- as.factor(trans_level2$month_dummy)
```

BA Project_BOPS_Final Submitted.Rmd | df42 | trans_level2 | df41 | trans_level1 | trans_level | meffects34 | sales2

Filter

	month	month_index	product_category	month_dummy	week_index	bops	female	logprice	f_store	f_year	f_month
12	MAR	32	5	3	84	0	0	5.959845	2	2012	3
12	MAR	32	5	3	84	0	0	7.120493	2	2012	3
11	AUG	25	12	8	53	0	1	3.693867	2	2011	8
11	AUG	25	12	8	53	0	1	3.753027	2	2011	8
11	AUG	25	5	8	53	0	1	4.117898	2	2011	8

Showing 1 to 7 of 987,099 entries

#####====================== OLS MODEL ==####

```
lpm41<-
lm(return~bops+logprice+product_category+f_store+est_income_code+female+age_band+f_month+f_year, data=trans_level2)

stargazer(lpm41,
        title="Regression Results", type="text",
        column.labels=c("Model-1"),
        df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
=============================================
                     Dependent variable:
                  ---------------------------
                           return
                           Model-1
---------------------------------------------
bops                      0.0151***
                          (0.0007)

logprice                  0.0354***
                          (0.0002)

product_category         -0.0007***
                          (0.00005)

f_store6                 -0.0063***
                          (0.0012)

f_store5998              -0.0834***
                          (0.0219)

est_income_code           0.0014***
                          (0.0001)

female                    0.0301***
                          (0.0006)

age_band                 -0.0011***
                          (0.0001)

f_month2                 -0.0317***
                          (0.0014)

f_month3                 -0.0146***
                          (0.0017)

f_month4                 -0.0170***
                          (0.0016)

f_month5                 -0.0374***
                          (0.0015)

f_month6                 -0.0220***
                          (0.0018)

f_month7                 -0.0282***
                          (0.0018)

f_month8                 -0.0241***
                          (0.0019)

f_month9                 -0.0262***
                          (0.0019)

f_month10                -0.0218***
                          (0.0018)

f_month11                -0.0341***
                          (0.0015)

f_month12                -0.0440***
                          (0.0014)

f_year2012               -0.0086***
                          (0.0008)

f_year2013               -0.0106***
                          (0.0012)

Constant                 -0.0333***
                          (0.0021)

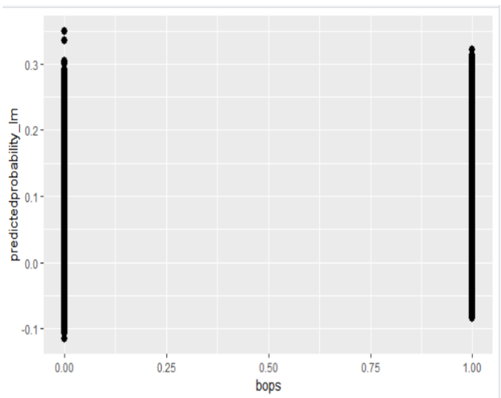
-------------------------------------------------
Observations                  987,099
R2                            0.0245
Adjusted R2                   0.0244
Residual Std. Error           0.2973
F Statistic                1,178.2720***
=================================================
Note:              *p<0.05; **p<0.01; ***p<0.001
>
```

Usage of BOPS service is associated with 1.51 percentage points increase in return probability. This means that a BOPS user is 1.51 percentage points more likely to return a jewellery item.

You can also use heteroscedastic-robust standard errors for 2SLS

gqtest(lpm41) # Significant Goldfeld-Quandt test indicates heteroskedasticity
bptest(lpm41) # Significant Breusch-Pagan test indicates heteroskedasticity

```
> gqtest(lpm41) # Significant Goldfeld-Quandt test indicates heteroskedasticity

        Goldfeld-Quandt test

data:  lpm41
GQ = 0.88408, df1 = 493530, df2 = 493530, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(lpm41) # Significant Breusch-Pagan test  indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  lpm41
BP = 24752, df = 21, p-value < 2.2e-16
```

consstder_lpm41 <- sqrt(diag(vcovHC(lpm41, type="const"))) # produces normal standard errors
HWrobstder_lpm41 <- sqrt(diag(vcovHC(lpm41, type="HC1"))) # produces Huber-White robust standard errors

Values	
consstder_lpm41	Named num [1:22] 0.0021194 0.0007191 0.0002477...
HWrobstder	Named num [1:11] 0.010905 0.00769 0.012733 0.0...

```
stargazer(lpm41, lpm41,
    se=list(consstder_lpm41, HWrobstder_lpm41),
    title="Regression Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

Regression Results

	Dependent variable: return	
	Normal SE (1)	HW-Robust SE (2)
bops	0.0151*** (0.0007)	0.0151*** (0.0007)
logprice	0.0354*** (0.0002)	0.0354*** (0.0002)
product_category	-0.0007*** (0.00005)	-0.0007*** (0.00005)
f_store6	-0.0063*** (0.0012)	-0.0063*** (0.0011)
f_store5998	-0.0834*** (0.0219)	-0.0834*** (0.0151)
est_income_code	0.0014*** (0.0001)	0.0014*** (0.0001)
female	0.0301*** (0.0006)	0.0301*** (0.0006)
age_band	-0.0011*** (0.0001)	-0.0011*** (0.0001)
f_month2	-0.0317*** (0.0014)	-0.0317*** (0.0014)
f_month3	-0.0146*** (0.0017)	-0.0146*** (0.0019)
f_month4	-0.0170*** (0.0016)	-0.0170*** (0.0016)
f_month5	-0.0374*** (0.0015)	-0.0374*** (0.0015)
f_month6	-0.0220*** (0.0018)	-0.0220*** (0.0020)
f_month7	-0.0282*** (0.0018)	-0.0282*** (0.0019)
f_month8	-0.0241*** (0.0019)	-0.0241*** (0.0020)
f_month9	-0.0262*** (0.0019)	-0.0262*** (0.0020)
f_month10	-0.0218*** (0.0018)	-0.0218*** (0.0018)
f_month11	-0.0341*** (0.0015)	-0.0341*** (0.0016)
f_month12	-0.0440*** (0.0014)	-0.0440*** (0.0014)
f_year2012	-0.0086*** (0.0008)	-0.0086*** (0.0008)
f_year2013	-0.0106*** (0.0012)	-0.0106*** (0.0012)
Constant	-0.0333*** (0.0021)	-0.0333*** (0.0020)
Observations	987,099	987,099
R2	0.0245	0.0245
Adjusted R2	0.0244	0.0244
Residual Std. Error	0.2973	0.2973
F Statistic	1,178.2720***	1,178.2720***
Note:	*p<0.05; **p<0.01; ***p<0.001	

displays normal and HW robust standard errors. # we observe that with robust standard errors.

trans_level2$predictedprobability_lm<-predict(lpm41) # let's look at the predicted probability of return for each observation in the data

BA Project_BOPS_Final Submitted.Rmd × | df42 × | trans_level2 × | df41 × | trans_level1 × | trans_level × | meffects3

Filter

egory	month_dummy	week_index	bops	female	logprice	f_store	f_year	f_month	predictedprobability_lm
5	3	84	0	0	5.959845	2	2012	3	0.155081625
5	3	84	0	0	7.120493	2	2012	3	0.196221157
12	8	53	0	1	3.693867	2	2011	8	0.092757119
12	8	53	0	1	3.753027	2	2011	8	0.094854074
5	8	53	0	1	4.117898	2	2011	8	0.118242085
17	8	53	0	1	6.917616	2	2011	8	0.207589767

Showing 1 to 7 of 987,099 entries

ggplot(trans_level2, aes(y=predictedprobability_lm, x=bops)) + geom_point(size=2.5)



range(trans_level2$predictedprobability_lm)

```
[1] -0.1153405  0.3497532
```

Range of the predicted probability tells us there are "negative" probabilities of return for some observations!!! This cannot be possible. Therefore, linear probability model is not the right model

confint(lpm41,"bops") # Generating confidence interval for variable BOPS

```
           2.5 %     97.5 %
bops 0.01369205 0.01651074
```

####====================== LOGIT MODEL =====================================####

sum(trans_level2$return==0)
sum(trans_level2$return==1)

```
> sum(trans_level2$return==0)
[1] 887623
> sum(trans_level2$return==1)
[1] 99476
```

We have 99476 observations with Return=1 and 887623 observations with Return=0. Considering that we will estimate 21 parameters, we satisfy the minimum 10:1 ratio requirement

```
logit42<-
glm(return~bops+logprice+product_category+f_store+f_month+f_year+est_income_code+female+age_band, data=trans_level2, family="binomial")
stargazer(logit42,
        apply.coef = exp, t.auto=F, p.auto = F,
        title="Regression Results", type="text",
        column.labels=c("OddsRatios"),
        df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
===============================================
                      Dependent variable:
                 ------------------------------
                            return
                          OddsRatios
-----------------------------------------------
bops                       1.2068***
                           (0.0079)

logprice                   1.5831***
                           (0.0033)

product_category            0.9987*
                           (0.0005)

f_store6                   0.9276***
                           (0.0142)

f_store5998                0.2983***
                           (0.3633)

f_month2                   0.7640***
                           (0.0153)

f_month3                   0.8900***
                           (0.0178)

f_month4                   0.7978***
                           (0.0180)

f_month5                   0.7082***
                           (0.0164)

f_month6                   0.8345***
                           (0.0192)

f_month7                   0.7847***
                           (0.0192)

f_month8                   0.8046***
                           (0.0201)

f_month9                   0.7790***
                           (0.0206)

f_month10                  0.7931***
                           (0.0194)

f_month11                  0.7305***
                           (0.0164)

f_month12                  0.6598***
                           (0.0150)

f_year2012                 0.9073***
                           (0.0096)

f_year2013                 0.8578***
                           (0.0137)

est_income_code            1.0163***
                           (0.0015)

female                     1.4522***
                           (0.0070)

age_band                   0.9873***
                           (0.0009)

Constant                   0.0146***
                           (0.0265)

-----------------------------------------------
Observations                987,099
Log Likelihood           -309,446.1000
Akaike Inf. Crit.        618,936.2000
===============================================
Note:               *p<0.05; **p<0.01; ***p<0.001
```

The odds of being returned (versus not being returned) increase by a factor of 1.2068 when a customer is a BOPS user.

Model fit assessment

logit42a <- glm(return~1, data=trans_level2, family="binomial") # This is the command to run a logit on null model

logit42	Large glm (30 elements, 836.2 Mb)
logit42a	Large glm (30 elements, 632.8 Mb)

lrtest(logit42, logit42a)

```
Likelihood ratio test

Model 1: return ~ bops + logprice + product_category + f_store + f_month +
    f_year + est_income_code + female + age_band
Model 2: return ~ 1
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  22 -309446
2   1 -322569 -21 26246  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#We compare the null model to our model to determine the model fit. The chi-square of 26246 with 21 degrees of freedom and an associated p-value of less than 0.001 tells us that our model as a whole fits significantly better than the null model.

Obtain marginal effects
a42 <-
logitmfx(formula=return~bops+logprice+product_category+f_store+f_month+f_year+est_income_code+female+age_band, data=trans_level2) # We can generate the marginal effects with this command.

marginaleffects42 <- a42$mfxest[,1]
marg.std.err42 <- a42$mfxest[,2]

marg.std.err42	Named num [1:21] 0.0007031 0.0002453 0.0…
marginaleffect…	Named num [1:21] 0.016027 0.037552 -0.00…

```
stargazer(logit42,
        omit=c("Constant"),
        coef = list(marginaleffects42), se = list(marg.std.err42),
        title="Regression Results", type="text",
        column.labels=c("Marginal Effects"),
        df=FALSE, digits=5, star.cutoffs = c(0.05,0.01,0.001))
```

```
Regression Results
==============================================
                    Dependent variable:
                ----------------------------
                          return
                     Marginal Effects
----------------------------------------------
bops                    0.01603***
                        (0.00070)

logprice                0.03755***
                        (0.00025)

product_category        -0.00011*
                        (0.00004)

f_store6                -0.00598***
                        (0.00110)

f_store5998             -0.06123***
                        (0.01009)

f_month2                -0.02021***
                        (0.00105)

f_month3                -0.00913***
                        (0.00134)

f_month4                -0.01703***
                        (0.00125)

f_month5                -0.02513***
                        (0.00106)

f_month6                -0.01382***
                        (0.00137)

f_month7                -0.01811***
                        (0.00131)

f_month8                -0.01639***
                        (0.00139)

f_month9                -0.01860***
                        (0.00139)

f_month10               -0.01743***
                        (0.00134)

f_month11               -0.02328***
                        (0.00110)

f_month12               -0.03147***
                        (0.00105)

f_year2012              -0.00795***
                        (0.00078)

f_year2013              -0.01218***
                        (0.00105)

est_income_code         0.00132***
                        (0.00013)

female                  0.03070***
                        (0.00057)

age_band                -0.00104***
                        (0.00007)

----------------------------------------------
Observations                987,099
Log Likelihood          -309,446.10000
Akaike Inf. Crit.        618,936.20000
==============================================
Note:           *p<0.05; **p<0.01; ***p<0.001
```

Usage of BOPS service increases the probability of return by 0.01603, holding other variables at their means.

b42 <- logitmfx(formula=return~bops+logprice+product_category+f_store+f_month+f_year+est_income_code +female+age_band, data=trans_level2, robust=TRUE) # We can obtain the marginal effects from a logit that uses robust standard errors. Note that marginal effects do not change, however, std. errors, and therefore, p-values change.
rob.std.err42 <- b42$mfxest[,2]

```
stargazer(logit42, logit42,
    se=list(marg.std.err42, rob.std.err42),
    omit=c("Constant"),
    coef = list(marginaleffects42,marginaleffects42),
    title="Regression Results", type="text",
    column.labels=c("Marginal Effects","Marg.Eff.w/RobStdEr" ),
    df=FALSE, digits=5, star.cutoffs = c(0.05,0.01,0.001))
```

Regression Results

	Dependent variable: return	
	Marginal Effects	Marg.Eff.w/RobStdEr
	(1)	(2)
bops	0.01603***	0.01603***
	(0.00070)	(0.00070)
logprice	0.03755***	0.03755***
	(0.00025)	(0.00024)
product_category	-0.00011*	-0.00011*
	(0.00004)	(0.00004)
f_store6	-0.00598***	-0.00598***
	(0.00110)	(0.00109)
f_store5998	-0.06123***	-0.06123***
	(0.01009)	(0.01015)
f_month2	-0.02021***	-0.02021***
	(0.00105)	(0.00105)
f_month3	-0.00913***	-0.00913***
	(0.00134)	(0.00134)
f_month4	-0.01703***	-0.01703***
	(0.00125)	(0.00125)
f_month5	-0.02513***	-0.02513***
	(0.00106)	(0.00106)
f_month6	-0.01382***	-0.01382***
	(0.00137)	(0.00138)
f_month7	-0.01811***	-0.01811***
	(0.00131)	(0.00131)
f_month8	-0.01639***	-0.01639***
	(0.00139)	(0.00139)
f_month9	-0.01860***	-0.01860***
	(0.00139)	(0.00140)
f_month10	-0.01743***	-0.01743***
	(0.00134)	(0.00134)
f_month11	-0.02328***	-0.02328***
	(0.00110)	(0.00111)
f_month12	-0.03147***	-0.03147***
	(0.00105)	(0.00106)
f_year2012	-0.00795***	-0.00795***
	(0.00078)	(0.00079)
f_year2013	-0.01218***	-0.01218***
	(0.00105)	(0.00105)
est_income_code	0.00132***	0.00132***
	(0.00013)	(0.00013)
female	0.03070***	0.03070***
	(0.00057)	(0.00058)
age_band	-0.00104***	-0.00104***
	(0.00007)	(0.00007)
Observations	987,099	987,099
Log Likelihood	-309,446.10000	-309,446.10000
Akaike Inf. Crit.	618,936.20000	618,936.20000
Note:		*p<0.05; **p<0.01; ***p<0.001

Usage of BOPS service increases the probability of return by 0.01603, holding other variables at their means.

```
pred42 = predict(logit42, data=trans_level2, type="response")
return_prediction42 <- ifelse(pred42 >= 0.5,1,0)
misClasificError42 <- mean(return_prediction42 != trans_level2$return)
```

misClasificErr…	0.10079840016047
pred42	Large numeric (987099 elements, 67.8 Mb)
return_predict…	Large numeric (987099 elements, 67.8 Mb)
rob.std.err42	Named num [1:21] 0.0007046 0.0002353 0.0…

```
print(paste('Accuracy',1-misClasificError42))
[1] "Accuracy 0.89920159983953"
```

the correct classification rate increased to 89.92%

```
table(trans_level2$return, pred42>=0.5)
```

	FALSE	TRUE
0	887590	33
1	99465	11

The coefficient of BOPS from linear probability model is 0.0151 with confidence intervals (0.01369205, 0.01651074) and the coefficient of LOGIT model is 0.01603. Since coefficent of LOGIT model lies in the confidence interval of coefficient of linear probability model, we can say that these models return similar results.

####================= TEST FOR ENDOGENEITY USING OLS========================####
Since using BOPS is a customer's decision , we suspect that key independent variable is endogenous. Child and length of residence are potential instrument variables. Proximity to store is our omitted variable.

Conceptually, these 2 variables do not explain any variation on return of jewellery items. For a customer with a child, using BOPS service would create a hassle for the customer to go back to store and thus child variable is negatively correlated with BOPS.
Length of residence is the number of years a customer has lived on his/her current residence. The higher the number of years the customer has lived in his residence, the more familiar he is with the stores around and is more likely to use BOPS

With length of residence, child as instrument variables
endo41<-
ivreg(return~bops+logprice+product_category+f_store+f_month+f_year+est_income_code+female+age_band|length_of_residence+child+logprice+product_category+f_store+f_month+f_year+est_income_code+female+age_band, data=trans_level2)
# The significant test statistics indicates that endogeneity exists and OLS results model will generate biased effects. Both relevance and exogeneity assumption has passed. Sargan statistic is not significant, meaning the instruments are exogenous. Weak instrument statistic is the F-statistic from the first stage. Since it is 121.89>10, it indicates the instruments are relevant. The significant durbin-wu-hausman test statistics indicates that endogeneity exists and OLS results model will generate biased effects.
summary(endo41,diagnostics = TRUE)

```
Call:
ivreg(formula = return ~ bops + logprice + product_category +
    f_store + f_month + f_year + est_income_code + female + age_band |
    length_of_residence + child + logprice + product_category +
        f_store + f_month + f_year + est_income_code + female +
        age_band, data = trans_level2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.62735 -0.09321 -0.02713  0.02258  1.17781

Coefficients:
                    Estimate  Std. Error t value Pr(>|t|)
(Intercept)      -0.07635747  0.00665133 -11.480  < 2e-16 ***
bops              0.36904126  0.05107061   7.226 4.97e-13 ***
logprice          0.03706413  0.00036192 102.409  < 2e-16 ***
product_category -0.00075432  0.00005311 -14.204  < 2e-16 ***
f_store6          0.02890149  0.00524564   5.510 3.60e-08 ***
f_store5998      -0.14716693  0.02607943  -5.643 1.67e-08 ***
f_month2         -0.04043300  0.00200233 -20.193  < 2e-16 ***
f_month3         -0.02814154  0.00274077 -10.268  < 2e-16 ***
f_month4         -0.02016118  0.00185282 -10.881  < 2e-16 ***
f_month5         -0.04500864  0.00196522 -22.903  < 2e-16 ***
f_month6         -0.03397232  0.00267187 -12.715  < 2e-16 ***
f_month7         -0.04315030  0.00294558 -14.649  < 2e-16 ***
f_month8         -0.02105389  0.00212501  -9.908  < 2e-16 ***
f_month9         -0.04828603  0.00383782 -12.582  < 2e-16 ***
f_month10        -0.04478738  0.00385257 -11.625  < 2e-16 ***
f_month11        -0.06218424  0.00439275 -14.156  < 2e-16 ***
f_month12        -0.07446752  0.00465185 -16.008  < 2e-16 ***
f_year2012       -0.05514420  0.00678403  -8.129 4.35e-16 ***
f_year2013       -0.06034666  0.00730686  -8.259  < 2e-16 ***
est_income_code   0.00232129  0.00020260  11.457  < 2e-16 ***
female            0.01746318  0.00194665   8.971  < 2e-16 ***
age_band          0.00020169  0.00020437   0.987    0.324
```

```
Diagnostic tests:
                    df1    df2 statistic  p-value
Weak instruments      2 987076    121.89  < 2e-16 ***
Wu-Hausman            1 987076     59.84 1.03e-14 ***
Sargan                1     NA      1.95    0.163
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3318 on 987077 degrees of freedom
Multiple R-Squared: -0.215,     Adjusted R-squared: -0.215
Wald test: 931.7 on 21 and 987077 DF,  p-value: < 2.2e-16
```

stargazer(endo41,
 title="Regression Results", type="text",
 column.labels=c("Endo442"),
 df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))

```
Regression Results
==================================================
                         Dependent variable:
                     -----------------------------
                               return
                               Endo442
--------------------------------------------------
bops                          0.369***
                              (0.051)

logprice                      0.037***
                             (0.0004)

product_category             -0.001***
                             (0.0001)

f_store6                      0.029***
                              (0.005)

f_store5998                  -0.147***
                              (0.026)

f_month2                     -0.040***
                              (0.002)

f_month3                     -0.028***
                              (0.003)

f_month4                     -0.020***
                              (0.002)

f_month5                     -0.045***
                              (0.002)

f_month6                     -0.034***
                              (0.003)

f_month7                     -0.043***
                              (0.003)

f_month8                     -0.021***
                              (0.002)

f_month9                     -0.048***
                              (0.004)

f_month10                    -0.045***
                              (0.004)

f_month11                    -0.062***
                              (0.004)

f_month12                    -0.074***
                              (0.005)

f_year2012                   -0.055***
                              (0.007)

f_year2013                   -0.060***
                              (0.007)

est_income_code               0.002***
                             (0.0002)

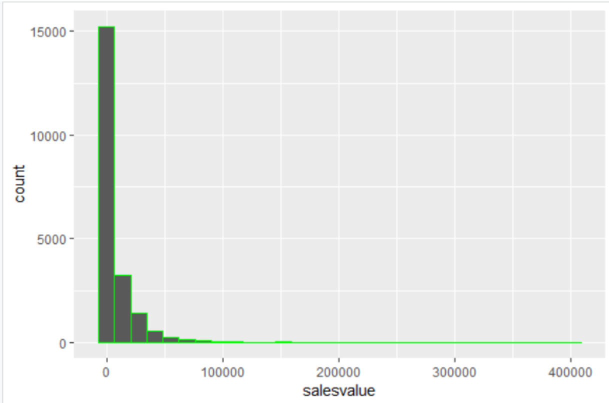
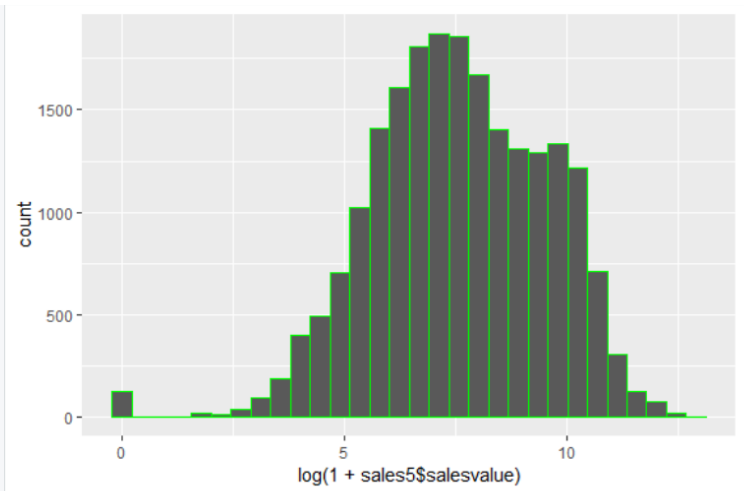
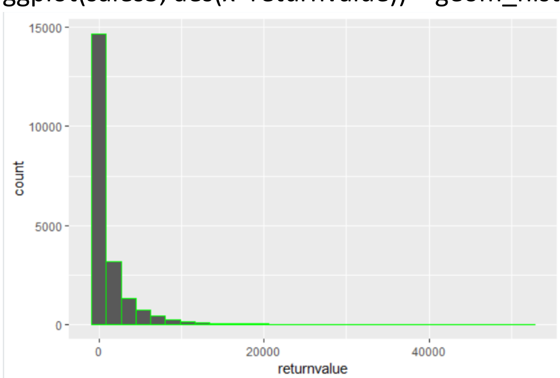
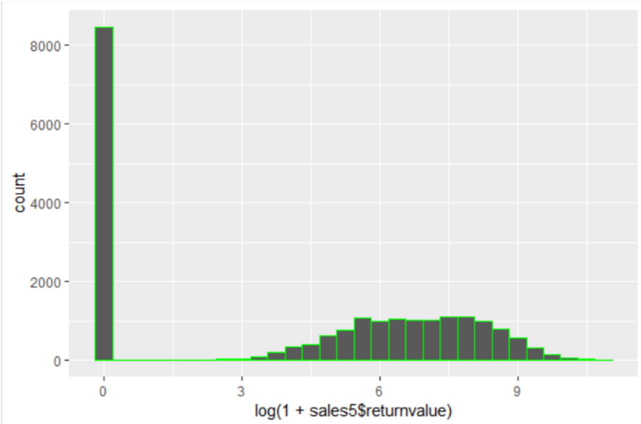
female                        0.017***
                              (0.002)

age_band                       0.0002
                             (0.0002)

Constant                     -0.076***
                              (0.007)

--------------------------------------------------
Observations                  987,099
R2                            -0.215
Adjusted R2                   -0.215
Residual Std. Error            0.332
==================================================
Note:               *p<0.05; **p<0.01; ***p<0.001
```

Interpretations: After treating endogeneity in our model, the coefficient for BOPS in 2SLS model is 0.369. This means that a BOPS user is 36.9 percentage points more likely to return a jewellery item. Since the results from our OLS and Logit models were similar, we are proceeding with 2SLS results.
```

#================================================================================#
## Question5: What is the impact of implementing BOPS strategy on product-level sales and returns?
#================================================================================#

```{r}
#Descriptive Statistics

stargazer(ODS_prodcat, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

Descriptive Statistics

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| store_number | 30,475 | 1,398.8 | 2,533.2 | 2 | 2 | 6 | 6 | 5,998 |
| year | 30,475 | 2,011.6 | 0.9 | 2,010 | 2,011 | 2,012 | 2,012 | 2,013 |
| month_index | 30,475 | 31.1 | 10.3 | 13 | 22 | 31 | 40 | 48 |
| product_category | 30,475 | 9.3 | 6.3 | 1 | 4 | 7 | 13 | 21 |
| month_dummy | 30,475 | 6.5 | 3.6 | 1 | 3 | 6 | 10 | 12 |
| bops_in_effect | 30,475 | 0.6 | 0.5 | 0 | 0 | 1 | 1 | 1 |
| day | 30,475 | 566.6 | 313.4 | 1 | 292 | 569 | 846 | 1,096 |
| salesvalue | 30,475 | 9,342.4 | 20,865.0 | 0.0 | 546.3 | 1,950.7 | 8,952.2 | 424,294.6 |
| returnvalue | 30,475 | 1,439.1 | 3,166.5 | 0.0 | 0.0 | 199.0 | 1,398.9 | 52,508.2 |
| salesquantity | 30,475 | 54.8 | 162.8 | 1 | 3 | 9 | 34 | 5,111 |
| returnquantity | 30,475 | 5.5 | 13.5 | 0 | 0 | 1 | 5 | 380 |
| avg_female | 23,277 | 0.5 | 0.3 | 0.0 | 0.3 | 0.5 | 0.7 | 1.0 |
| avg_age | 23,795 | 4.9 | 2.1 | 0.0 | 4.0 | 4.9 | 5.9 | 13.0 |
| avg_income | 23,763 | 5.3 | 1.2 | 1.0 | 4.8 | 5.3 | 5.8 | 9.0 |
| avg_homeowner | 23,763 | 0.6 | 0.3 | 0.0 | 0.5 | 0.7 | 0.8 | 1.0 |
| avg_residency | 23,763 | 6.9 | 2.8 | 0.0 | 5.6 | 6.9 | 8.0 | 15.0 |
| avg_childowner | 23,763 | 0.4 | 0.2 | 0.0 | 0.2 | 0.4 | 0.5 | 1.0 |

#Filtering data for days before sept.27,2012

sales5 <- ODS_prodcat[((ODS_prodcat$day<788) | (ODS_prodcat$day==788)),]

stargazer(sales5, type="text", median=TRUE, iqr=TRUE, digits=1, title="Descriptive Statistics")

Descriptive Statistics

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| store_number | 21,115 | 1,286.5 | 2,458.6 | 2 | 2 | 6 | 6 | 5,998 |
| year | 21,115 | 2,011.2 | 0.7 | 2,010 | 2,011 | 2,011 | 2,012 | 2,012 |
| month_index | 21,115 | 25.7 | 7.3 | 13 | 19 | 26 | 32 | 38 |
| product_category | 21,115 | 9.4 | 6.3 | 1 | 4 | 7 | 13 | 21 |
| month_dummy | 21,115 | 6.7 | 3.5 | 1 | 4 | 7 | 9 | 12 |
| bops_in_effect | 21,115 | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| day | 21,115 | 401.8 | 223.4 | 1 | 205 | 409 | 591 | 788 |
| salesvalue | 21,115 | 8,722.2 | 19,132.2 | 0.0 | 508.7 | 1,856.6 | 8,526.0 | 402,473.4 |
| returnvalue | 21,115 | 1,373.6 | 2,992.2 | 0.0 | 0.0 | 193.8 | 1,376.5 | 51,973.1 |
| salesquantity | 21,115 | 49.9 | 139.6 | 1 | 2 | 9 | 32 | 4,474 |
| returnquantity | 21,115 | 5.3 | 12.5 | 0 | 0 | 1 | 5 | 380 |
| avg_female | 16,489 | 0.5 | 0.3 | 0.0 | 0.3 | 0.5 | 0.7 | 1.0 |
| avg_age | 16,839 | 5.0 | 2.1 | 0.0 | 4.1 | 5.0 | 6.0 | 13.0 |
| avg_income | 16,827 | 5.3 | 1.2 | 1.0 | 4.8 | 5.4 | 5.8 | 9.0 |
| avg_homeowner | 16,827 | 0.6 | 0.3 | 0.0 | 0.5 | 0.7 | 0.8 | 1.0 |
| avg_residency | 16,827 | 6.9 | 2.8 | 0.0 | 5.5 | 6.8 | 7.9 | 15.0 |
| avg_childowner | 16,827 | 0.4 | 0.2 | 0.0 | 0.2 | 0.3 | 0.5 | 1.0 |

#Add a BIE timeline variable dividing the data into two groups, one before Aug.1, 2011 and the other after.

sales5$BIE_timeline_PC <- ifelse(sales5$day<366,0,1)

#Add a grouping variable for stores dividing the data into two groups, one for 2 and 6 and the other for 5998.

sales5$group_store_PC <- ifelse((sales5$store_number==2) | (sales5$store_number==6),1,0)

#Convert product category and month to factor variables

sales5$factor_PC <- as.factor(sales5$product_category)

sales5$factor_Month <- as.factor(sales5$month_dummy)

BA Project_BOPS_Final Submitted.Rmd × | sales5 ×

Filter

| e | avg_income | avg_homeowner | avg_residency | avg_childowner | BIE_timeline_PC | group_store_PC | factor_PC | factor_Month |
|---|---|---|---|---|---|---|---|---|
| 3311 | 5.612794 | 0.6296296 | 7.037037 | 0.36363637 | 1 | 1 | 5 | 3 |
| 0000 | 8.000000 | 1.0000000 | 4.000000 | 0.00000000 | 0 | 1 | 4 | 8 |
| 4286 | 4.428571 | 0.8571429 | 8.428572 | 0.28571430 | 0 | 1 | 5 | 8 |
| 0000 | 4.500000 | 0.5000000 | 10.000000 | 0.00000000 | 0 | 1 | 13 | 8 |
| 0000 | 2.000000 | 0.0000000 | 8.000000 | 1.00000000 | 0 | 1 | 2 | 8 |
| 6448 | 5.691257 | 0.6229508 | 7.628415 | 0.39071038 | 0 | 1 | 12 | 8 |

Showing 1 to 7 of 21,115 entries

#Assuming mean value for all 'na' values

sales5$avg_female[is.na(sales5$avg_female)] <- mean(sales5$avg_female, na.rm = TRUE)
sales5$avg_age[is.na(sales5$avg_age)] <- mean(sales5$avg_age, na.rm = TRUE)
sales5$avg_income[is.na(sales5$avg_income)] <- mean(sales5$avg_income, na.rm = TRUE)
sales5$avg_homeowner[is.na(sales5$avg_homeowner)] <- mean(sales5$avg_homeowner, na.rm = TRUE)
sales5$avg_residency[is.na(sales5$avg_residency)] <- mean(sales5$avg_residency, na.rm = TRUE)
sales5$avg_childowner[is.na(sales5$avg_childowner)] <- mean(sales5$avg_childowner, na.rm = TRUE)

#Descriptive Statistics after placing mean values for all 'na' values

stargazer(sales5, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

```
Descriptive Statistics
==================================================================================
Statistic          N     Mean    St. Dev.  Min  Pctl(25) Median  Pctl(75)    Max
----------------------------------------------------------------------------------
store_number     21,115 1,286.5 2,458.6    2      2        6       6       5,998
year             21,115 2,011.2   0.7    2,010  2,011    2,011   2,012     2,012
month_index      21,115  25.7     7.3     13     19       26      32        38
product_category 21,115   9.4     6.3      1      4        7      13        21
month_dummy      21,115   6.7     3.5      1      4        7       9        12
bops_in_effect   21,115   0.4     0.5      0      0        0       1         1
day              21,115  401.8   223.4     1     205      409     591       788
salesvalue       21,115 8,722.2 19,132.2  0.0   508.7   1,856.6 8,526.0  402,473.4
returnvalue      21,115 1,373.6 2,992.2   0.0    0.0     193.8  1,376.5   51,973.1
salesquantity    21,115  49.9    139.6     1      2        9      32       4,474
returnquantity   21,115   5.3     12.5     0      0        1       5        380
avg_female       21,115   0.5     0.3     0.0    0.4      0.5     0.6       1.0
avg_age          21,115   5.0     1.9     0.0    4.4      5.0     5.7      13.0
avg_income       21,115   5.3     1.1     1.0    5.0      5.3     5.7       9.0
avg_homeowner    21,115   0.6     0.2     0.0    0.6      0.6     0.7       1.0
avg_residency    21,115   6.9     2.5     0.0    6.0      6.9     7.5      15.0
avg_childowner   21,115   0.4     0.2     0.0    0.3      0.4     0.4       1.0
BIE_timeline_PC  21,115   0.6     0.5      0      0        1       1         1
group_store_PC   21,115   0.8     0.4      0      1        1       1         1
----------------------------------------------------------------------------------
```

#Generate plot for Sales Value

ggplot(sales5, aes(x=salesvalue)) + geom_histogram(colour="green", bins =30)



ggplot(sales5, aes(x=log(1+sales5$salesvalue))) + geom_histogram(colour="green", bins =30)



#Generate plot for Return Value

ggplot(sales5, aes(x=returnvalue)) + geom_histogram(colour="green", bins =30)



ggplot(sales5, aes(x=log(1+sales5$returnvalue))) + geom_histogram(colour="green", bins =30)



####=====================IMPACT ON SALES QUANTITY===============================####

#Count Data model approach for sales quantity
#Poisson Model

```
poisson5 <-
glm(salesquantity~BIE_timeline_PC*group_store_PC+factor_PC+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner+factor_Month, family="poisson", data=sales5)
stargazer(poisson5,
        title="Poisson Results", type="text",
        column.labels=c("Model-1"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

Poisson Results

	Dependent variable: salesquantity Model-1
BIE_timeline_PC	0.40*** (0.01)
group_store_PC	2.75*** (0.01)
factor_PC2	0.84*** (0.01)
factor_PC3	-0.20*** (0.01)
factor_PC4	1.78*** (0.01)
factor_PC5	2.13*** (0.01)
factor_PC6	0.23*** (0.01)
factor_PC7	1.05*** (0.01)
factor_PC8	-1.95*** (0.03)
factor_PC9	0.90*** (0.01)
factor_PC10	-2.63*** (0.07)
factor_PC11	0.38*** (0.01)
factor_PC12	2.29*** (0.01)
factor_PC13	0.23*** (0.01)
factor_PC14	0.05*** (0.01)
factor_PC15	-3.19*** (0.18)
factor_PC17	-1.30*** (0.02)
factor_PC20	-0.02* (0.01)
factor_PC21	2.12*** (0.01)
avg_female	-0.37*** (0.005)
avg_age	-0.09*** (0.001)
avg_income	0.14*** (0.001)

```
avg_homeowner                          -0.27***
                                        (0.01)

avg_childowner                          0.001
                                        (0.01)

factor_Month2                           0.44***
                                        (0.005)

factor_Month3                          -0.22***
                                        (0.01)

factor_Month4                          -0.18***
                                        (0.01)

factor_Month5                           0.21***
                                        (0.01)

factor_Month6                          -0.26***
                                        (0.01)

factor_Month7                          -0.30***
                                        (0.01)

factor_Month8                          -0.16***
                                        (0.01)

factor_Month9                          -0.12***
                                        (0.01)

factor_Month10                         -0.10***
                                        (0.01)

factor_Month11                          0.44***
                                        (0.005)

factor_Month12                          1.20***
                                        (0.004)

BIE_timeline_PC:group_store_PC         -0.38***
                                        (0.01)

Constant                                -0.02
                                        (0.01)

------------------------------------------------------------
Observations                            21,115
Log Likelihood                        -804,460.10
Akaike Inf. Crit.                     1,608,994.00
============================================================
Note:                          *p<0.05; **p<0.01; ***p<0.001
>
```

#Model fit assessment
poisson5a <- glm(salesquantity~1, data=sales5, family="poisson") #running a comparison with null model.

lrtest(poisson5, poisson5a) # We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

```
Likelihood ratio test

Model 1: salesquantity ~ BIE_timeline_PC * group_store_PC + factor_PC +
    avg_female + avg_age + avg_income + avg_homeowner + avg_childowner +
    factor_Month
Model 2: salesquantity ~ 1
  #Df   LogLik  Df   Chisq Pr(>Chisq)
1  37  -804460
2   1 -1524545 -36 1440171  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#Since Poisson doesn't fit the data, we will check for negative binomial model.
#Negative Binomial Model
negbin5 <-
glm.nb(salesquantity~BIE_timeline_PC*group_store_PC+factor_PC+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner+factor_Month, data = sales5)

stargazer(negbin5,
    apply.coef = exp, t.auto=F, p.auto = F,
    title="Negative Bionomial Results", type="text",
    column.labels=c("IRRs"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))

Negative Bionomial Results

	Dependent variable: salesquantity IRRs
BIE_timeline_PC	1.3906*** (0.0403)
group_store_PC	12.4154*** (0.0354)
factor_PC2	2.0824*** (0.0457)
factor_PC3	0.8369*** (0.0468)
factor_PC4	4.8313*** (0.0429)
factor_PC5	6.7282*** (0.0432)
factor_PC6	1.1427** (0.0460)
factor_PC7	2.9223*** (0.0482)
factor_PC8	0.1696*** (0.0808)
factor_PC9	3.5301*** (0.0621)
factor_PC10	0.0883*** (0.1301)
factor_PC11	1.3540*** (0.0460)
factor_PC12	7.4624*** (0.0435)
factor_PC13	1.1304** (0.0464)
factor_PC14	1.0746 (0.0502)
factor_PC15	0.0578*** (0.3028)
factor_PC17	0.2922*** (0.0650)
factor_PC20	1.0459 (0.0465)
factor_PC21	6.3250*** (0.0433)
avg_female	0.8112*** (0.0354)
avg_age	0.8760*** (0.0047)
avg_income	1.2317*** (0.0085)
avg_homeowner	0.6094*** (0.0413)
avg_childowner	1.0441 (0.0386)
factor_Month2	1.5271*** (0.0411)
factor_Month3	0.8808** (0.0420)
factor_Month4	0.8882** (0.0424)
factor_Month5	1.2753*** (0.0411)
factor_Month6	0.8487*** (0.0427)
factor_Month7	0.8195*** (0.0422)
factor_Month8	0.9268* (0.0386)
factor_Month9	0.9153* (0.0396)
factor_Month10	0.9519 (0.0429)
factor_Month11	1.5923*** (0.0409)
factor_Month12	2.9406*** (0.0393)
BIE_timeline_PC:group_store_PC	0.6834*** (0.0441)
Constant	1.2416** (0.0701)
Observations	21,115
Log Likelihood	-87,498.7200
theta	0.7400*** (0.0066)
Akaike Inf. Crit.	175,071.4000
Note:	*p<0.05; **p<0.01; ***p<0.001

#Model fit assessment

negbin5a <- glm.nb(salesquantity~1, data=sales5) #running a comparison with null model.

lrtest(negbin5, negbin5a) # We conclude that the model fit because the goodness-of-fit chi-squared test is statistically significant.

```
Likelihood ratio test

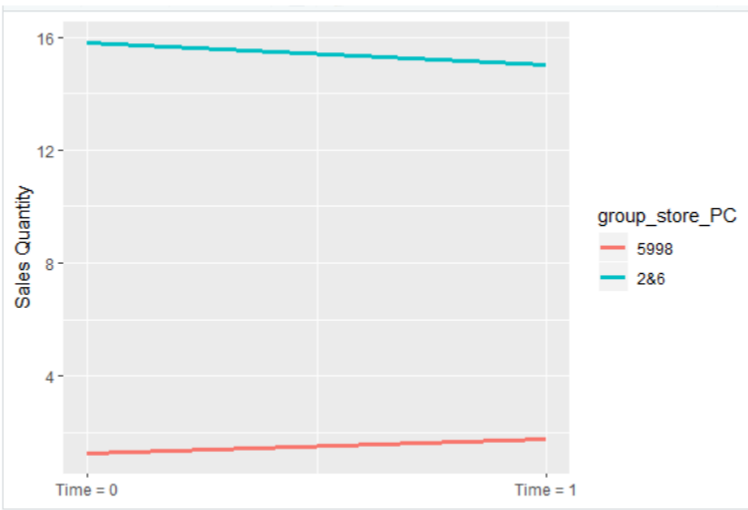
Model 1: salesquantity ~ BIE_timeline_PC * group_store_PC + factor_PC +
    avg_female + avg_age + avg_income + avg_homeowner + avg_childowner +
    factor_Month
Model 2: salesquantity ~ 1
  #Df LogLik  Df Chisq Pr(>Chisq)
1  38 -87498
2   2 -96093 -36 17190  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
# Check for heteroskedasticity
gqtest(negbin5) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(negbin5) # Breusch-Pagan test indicates heteroskedasticity
```

```
> gqtest(negbin5) # Goldfeld-Quandt test indicates no heteroskedasticity

	Goldfeld-Quandt test

data:  negbin5
GQ = 0.0083442, df1 = 10521, df2 = 10520, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin5) # Breusch-Pagan test indicates heteroskedasticity

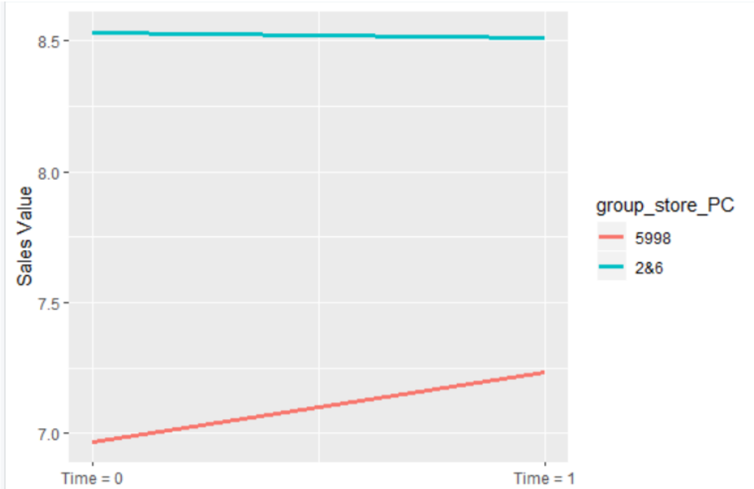
	studentized Breusch-Pagan test

data:  negbin5
BP = 650.41, df = 36, p-value < 2.2e-16
```

```
# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(negbin5, type="HC1"))) # produces Huber-White robust standard errors
```

Values	
HWrobstder	Named num [1:37] 0.0733 0.0364 0.0344 0.044 0.0392…

```
stargazer(negbin5, negbin5,
        apply.coef = exp, t.auto=F, p.auto = F,
        se=list(NULL, HWrobstder),
        title="Negative Binomial Results", type="text",
        column.labels=c("Normal SE", "HW-Robust SE"),
        df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable: salesquantity	
	Normal SE (1)	HW-Robust SE (2)
BIE_timeline_PC	1.391*** (0.040)	1.391*** (0.036)
group_store_PC	12.415*** (0.035)	12.415*** (0.034)
factor_PC2	2.082*** (0.046)	2.082*** (0.044)
factor_PC3	0.837*** (0.047)	0.837*** (0.039)
factor_PC4	4.831*** (0.043)	4.831*** (0.041)
factor_PC5	6.728*** (0.043)	6.728*** (0.043)
factor_PC6	1.143** (0.046)	1.143** (0.041)
factor_PC7	2.922*** (0.048)	2.922*** (0.103)
factor_PC8	0.170*** (0.081)	0.170*** (0.060)
factor_PC9	3.530*** (0.062)	3.530*** (0.077)
factor_PC10	0.088*** (0.130)	0.088*** (0.089)
factor_PC11	1.354*** (0.046)	1.354*** (0.048)
factor_PC12	7.462*** (0.043)	7.462*** (0.042)
factor_PC13	1.130** (0.046)	1.130** (0.044)
factor_PC14	1.075 (0.050)	1.075 (0.049)
factor_PC15	0.058*** (0.303)	0.058*** (0.173)
factor_PC17	0.292*** (0.065)	0.292*** (0.045)
factor_PC20	1.046 (0.047)	1.046 (0.040)
factor_PC21	6.325*** (0.043)	6.325*** (0.041)
avg_female	0.811*** (0.035)	0.811*** (0.034)
avg_age	0.876*** (0.005)	0.876*** (0.004)
avg_income	1.232*** (0.008)	1.232*** (0.008)
avg_homeowner	0.609*** (0.041)	0.609*** (0.033)
avg_childowner	1.044 (0.039)	1.044 (0.031)
factor_Month2	1.527*** (0.041)	1.527*** (0.049)
factor_Month3	0.881** (0.042)	0.881** (0.043)
factor_Month4	0.888** (0.042)	0.888** (0.046)
factor_Month5	1.275*** (0.041)	1.275*** (0.047)
factor_Month6	0.849*** (0.043)	0.849*** (0.042)
factor_Month7	0.820*** (0.042)	0.820*** (0.041)
factor_Month8	0.927* (0.039)	0.927* (0.048)
factor_Month9	0.915* (0.040)	0.915* (0.040)
factor_Month10	0.952 (0.043)	0.952 (0.047)
factor_Month11	1.592*** (0.041)	1.592*** (0.047)
factor_Month12	2.941*** (0.039)	2.941*** (0.048)
BIE_timeline_PC:group_store_PC	0.683*** (0.044)	0.683*** (0.043)
Constant	1.242** (0.070)	1.242** (0.073)
Observations	21,115	21,115
Log Likelihood	-87,498.720	-87,498.720
theta	0.740*** (0.007)	0.740*** (0.007)
Akaike Inf. Crit.	175,071.400	175,071.400
Note:	*p<0.05; **p<0.01; ***p<0.001	

# Visualize the output
meffects51 <- ggpredict(negbin5, terms=c("BIE_timeline_PC", "group_store_PC"))
# generates a tidy data frame at two different values of BIE_timeline - before and after BOPS implementation.

BA Project_BOPS_Final Submitted.Rmd × meffects51 × sales5 ×

Filter

	x	predicted	conf.low	conf.high	group
1	0	1.273613	1.145794	1.415691	0
2	0	15.812390	14.488098	17.257730	1
3	1	1.771093	1.607197	1.951703	0
4	1	15.026396	13.771667	16.395442	1

```
ggplot(meffects51,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline_PC") + ylab("Sales Quantity") +
  labs(colour="group_store_PC") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time = 0", "Time = 1")) +
  theme(axis.title.x=element_blank())
```

The interaction coefficient of BIE_timeline_PC:group_store_PC is significant with IRR value 0.6834. This means that BOPS implementation is associated with 31.66% decrease in product-level sales quantity.



####====================IMPACT ON SALES VALUE============================####

#Since sales value is a dollar value, we will use a linear interaction model with log transformed dependent variable.

```
sales5$log_salesvalue <- log(1+sales5$salesvalue)
```

BA Project_BOPS_Final Submitted.Rmd | meffects51 | sales5

Filter

	avg_homeowner	avg_residency	avg_childowner	BIE_timeline_PC	group_store_PC	factor_PC	factor_Month	log_salesvalue
94	0.6296296	7.037037	0.36363637	1	1	5	3	10.575784
00	1.0000000	4.000000	0.00000000	0	1	4	8	5.309158
71	0.8571429	8.428572	0.28571430	0	1	5	8	6.730517
00	0.5000000	10.000000	0.00000000	0	1	13	8	5.319492
00	0.0000000	8.000000	1.00000000	0	1	2	8	3.911823

Showing 1 to 6 of 21,115 entries

#OLS model approach for sales quantity

```
ols51 =
lm(log_salesvalue~BIE_timeline_PC*group_store_PC+factor_PC+avg_female+avg_age+avg_income+avg_homeowner+avg_childowner+factor_Month, data=sales5)
stargazer(ols51,
```

```
title="Regression Results", type="text",
column.labels=c("Model-1"),
df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

Check for heteroskedasticity

```
gqtest(ols51) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols51) # Breusch-Pagan test indicates heteroskedasticity
```

```
        Goldfeld-Quandt test

data:  ols51
GQ = 0.44738, df1 = 10521, df2 = 10520, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols51) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols51
BP = 1022.8, df = 36, p-value < 2.2e-16
```

Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.

```
HWrobstder <- sqrt(diag(vcovHC(ols51, type="HC1"))) # produces Huber-White robust standard errors
```

```
Values
  HWrobstder        Named num [1:37] 0.103 0.033 0.0321 0.0612 0.059 ...
```

```
stargazer(ols51, ols51,
    se=list(NULL, HWrobstder),
    title="Negative Binomial Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE"),
    df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable: log_salesvalue	
	Normal SE (1)	HW-Robust SE (2)
BIE_timeline_PC	0.265*** (0.051)	0.265*** (0.033)
group_store_PC	1.565*** (0.045)	1.565*** (0.032)
factor_PC2	-1.197*** (0.061)	-1.197*** (0.061)
factor_PC3	-0.231*** (0.062)	-0.231*** (0.059)
factor_PC4	0.132* (0.057)	0.132* (0.058)
factor_PC5	-0.332*** (0.058)	-0.332*** (0.059)
factor_PC6	-1.126*** (0.061)	-1.126*** (0.059)
factor_PC7	-2.832*** (0.064)	-2.832*** (0.062)
factor_PC8	-2.090*** (0.100)	-2.090*** (0.088)
factor_PC9	-2.755*** (0.084)	-2.755*** (0.098)
factor_PC10	-2.897*** (0.151)	-2.897*** (0.122)
factor_PC11	-0.829*** (0.061)	-0.829*** (0.061)
factor_PC12	-0.309*** (0.058)	-0.309*** (0.060)
factor_PC13	-1.240*** (0.061)	-1.240*** (0.058)
factor_PC14	-2.542*** (0.067)	-2.542*** (0.074)
factor_PC15	-3.576*** (0.325)	-3.576*** (0.321)
factor_PC17	0.267** (0.085)	0.267*** (0.065)
factor_PC20	-0.693*** (0.061)	-0.693*** (0.058)
factor_PC21	-0.602*** (0.058)	-0.602*** (0.059)
avg_female	-0.364*** (0.047)	-0.364*** (0.062)
avg_age	-0.061*** (0.006)	-0.061*** (0.008)

avg_income	0.107***	0.107***
	(0.011)	(0.015)
avg_homeowner	-0.200***	-0.200**
	(0.055)	(0.076)
avg_childowner	0.076	0.076
	(0.052)	(0.071)
factor_Month2	0.415***	0.415***
	(0.055)	(0.058)
factor_Month3	0.045	0.045
	(0.056)	(0.060)
factor_Month4	0.012	0.012
	(0.057)	(0.059)
factor_Month5	0.230***	0.230***
	(0.055)	(0.059)
factor_Month6	-0.0001	-0.0001
	(0.057)	(0.057)
factor_Month7	-0.003	-0.003
	(0.056)	(0.058)
factor_Month8	0.008	0.008
	(0.052)	(0.054)
factor_Month9	0.068	0.068
	(0.053)	(0.055)
factor_Month10	0.006	0.006
	(0.057)	(0.061)
factor_Month11	0.413***	0.413***
	(0.055)	(0.058)
factor_Month12	1.003***	1.003***
	(0.053)	(0.056)
BIE_timeline_PC:group_store_PC	-0.288***	-0.288***
	(0.056)	(0.042)
Constant	6.996***	6.996***
	(0.092)	(0.103)
Observations	21,115	21,115
R2	0.323	0.323
Adjusted R2	0.322	0.322
Residual Std. Error	1.607	1.607
F Statistic	279.284***	279.284***
Note:		*p<0.05; **p<0.01; ***p<0.001

meffects52 <- ggpredict(ols51, terms=c("BIE_timeline_PC", "group_store_PC"))

generates a tidy data frame

BA Project_BOPS_Final Submitted.Rmd × meffects52 × meffects51 ×

Filter

	x	predicted	conf.low	conf.high	group
1	0	6.967346	6.830327	7.104365	0
2	0	8.532798	8.416403	8.649192	1
3	1	7.232166	7.104944	7.359389	0
4	1	8.509448	8.393477	8.625419	1

Visualize the output

```
ggplot(meffects52,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline_PC") + ylab("Sales Value") +
  labs(colour="group_store_PC") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time = 0", "Time = 1")) +
  theme(axis.title.x=element_blank())
```



The interaction coefficient of BIE_timeline_PC:group_store_PC is significant with value -0.288 . This means that BOPS implementation is associated with 28.8% decrease in product-level sales value.

####=====================IMPACT ON RETURN QUANTITY==================####

```
#Count Data model approach for return quantity
#Poisson Model
poisson52 <-
glm(returnquantity~BIE_timeline_PC*group_store_PC+factor_PC+salesquantity+avg_female+avg_age+avg_income+factor_Month, family="poisson", data=sales5)
stargazer(poisson52,
        title="Poisson Results", type="text",
        column.labels=c("Model-1"),
        df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

Poisson Results

	Dependent variable: returnquantity Model-1
BIE_timeline_PC	0.52*** (0.05)
group_store_PC	2.68*** (0.04)
factor_PC2	0.65*** (0.02)
factor_PC3	-0.26*** (0.02)
factor_PC4	1.07*** (0.02)
factor_PC5	0.93*** (0.02)
factor_PC6	-0.02 (0.02)
factor_PC7	-2.30*** (0.04)
factor_PC8	-1.63*** (0.07)
factor_PC9	-0.36*** (0.03)
factor_PC10	-2.25*** (0.14)
factor_PC11	-0.12*** (0.02)
factor_PC12	0.94*** (0.02)
factor_PC13	-0.40*** (0.02)
factor_PC14	-0.52*** (0.03)
factor_PC15	-3.03*** (0.50)
factor_PC17	-1.23*** (0.05)
factor_PC20	0.06** (0.02)
factor_PC21	0.82*** (0.02)
salesquantity	0.002*** (0.0000)
avg_female	-0.17*** (0.02)
avg_age	-0.09*** (0.002)
avg_income	0.09*** (0.004)
factor_Month2	0.04** (0.01)
factor_Month3	-0.27*** (0.02)
factor_Month4	-0.33*** (0.02)
factor_Month5	-0.15*** (0.02)
factor_Month6	-0.23*** (0.02)
factor_Month7	-0.20*** (0.02)
factor_Month8	-0.34*** (0.01)
factor_Month9	-0.38*** (0.02)
factor_Month10	-0.25*** (0.02)
factor_Month11	0.07*** (0.01)
factor_Month12	-0.03* (0.01)
BIE_timeline_PC:group_store_PC	-0.53*** (0.05)
Constant	-1.24*** (0.05)
Observations	21,115
Log Likelihood	-81,838.35
Akaike Inf. Crit.	163,748.70
Note:	*p<0.05; **p<0.01; ***p<0.001

Model fit assessment

poisson52a <- glm(returnquantity~1, data=sales5, family="poisson") # running a comparison with null model.

lrtest(poisson52, poisson52a) # We conclude that the model does not fit because the goodness-of-fit chi-squared test is statistically significant. If the test had not been statistically significant, it would indicate that the data fit the model well.

```
Likelihood ratio test

Model 1: returnquantity ~ BIE_timeline_PC * group_store_PC + factor_PC +
    salesquantity + avg_female + avg_age + avg_income + factor_Month
Model 2: returnquantity ~ 1
  #Df  LogLik  Df  Chisq Pr(>Chisq)
1  36  -81838
2   1 -165851 -35 168025  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Since Poisson doesn't fit the data, we will check for negative binomial model.
#Negative Binomial Model

```
negbin53 <-
glm.nb(returnquantity~BIE_timeline_PC*group_store_PC+factor_PC+salesquantity+avg_female+avg_age+avg_income+factor_Month, data = sales5)
stargazer(negbin53,
        apply.coef = exp, t.auto=F, p.auto = F,
        title="Negative Bionomial Results", type="text",
        column.labels=c("IRRs"),
        df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Bionomial Results

	Dependent variable: returnquantity IRRs
BIE_timeline_PC	1.5183*** (0.0583)
group_store_PC	6.0839*** (0.0511)
factor_PC2	1.2790*** (0.0470)
factor_PC3	0.8321*** (0.0491)
factor_PC4	1.2268*** (0.0446)
factor_PC5	0.8908* (0.0460)
factor_PC6	0.8770** (0.0483)
factor_PC7	0.1834*** (0.0657)
factor_PC8	0.2515*** (0.0968)
factor_PC9	0.4081*** (0.0736)
factor_PC10	0.1385*** (0.1713)
factor_PC11	0.6914*** (0.0495)
factor_PC12	0.9064* (0.0465)
factor_PC13	0.5762*** (0.0507)
factor_PC14	0.5587*** (0.0547)
factor_PC15	0.0575*** (0.5610)
factor_PC17	0.3962*** (0.0728)
factor_PC20	1.0473 (0.0483)
factor_PC21	0.8070*** (0.0462)
salesquantity	1.0085*** (0.0001)
avg_female	1.0975* (0.0393)

avg_age	0.9169*** (0.0052)	factor_Month9	0.7815*** (0.0430)
avg_income	1.0863*** (0.0090)	factor_Month10	0.8517*** (0.0460)
factor_Month2	0.9092* (0.0446)	factor_Month11	0.9205 (0.0444)
factor_Month3	0.8832** (0.0454)	factor_Month12	1.0723 (0.0429)
factor_Month4	0.7917*** (0.0463)	BIE_timeline_PC:group_store_PC	0.6151*** (0.0612)
factor_Month5	0.8488*** (0.0449)	Constant	0.4683*** (0.0824)
factor_Month6	0.8388*** (0.0466)	Observations	21,115
factor_Month7	0.9131* (0.0456)	Log Likelihood	-43,322.6100
factor_Month8	0.8219*** (0.0417)	theta	0.9046*** (0.0139)
		Akaike Inf. Crit.	86,717.2200
		Note:	*p<0.05; **p<0.01; ***p<0.001

negbin53a <- glm.nb(returnquantity~1, data=sales5)

#running a comparison with null model.

lrtest(negbin53, negbin53a) # We conclude that the model fits well.

```
Likelihood ratio test

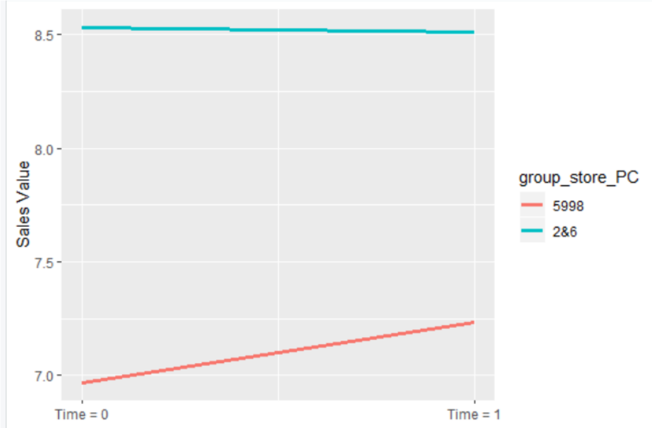
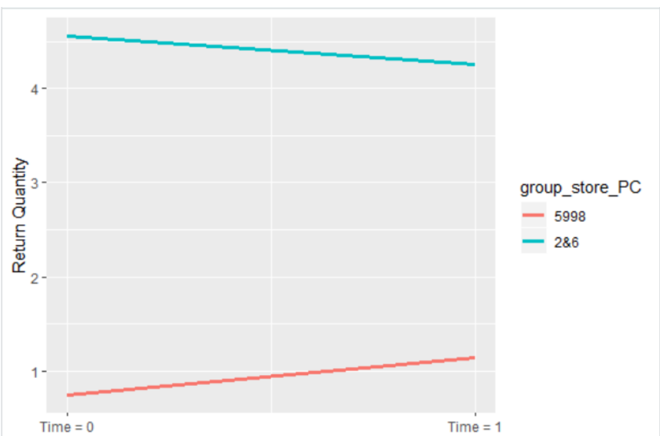
Model 1: returnquantity ~ BIE_timeline_PC * group_store_PC + factor_PC +
    salesquantity + avg_female + avg_age + avg_income + factor_Month
Model 2: returnquantity ~ 1
  #Df LogLik  Df Chisq Pr(>Chisq)
1  37 -43322
2   2 -51169 -35 15694  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Check for heteroskedasticity

gqtest(negbin53) # Goldfeld-Quandt test indicates no heteroskedasticity

bptest(negbin53) # Breusch-Pagan test indicates heteroskedasticity

```
> gqtest(negbin53) # Goldfeld-Quandt test indicates no heteroskedasticity

        Goldfeld-Quandt test

data:  negbin53
GQ = 0.021027, df1 = 10522, df2 = 10521, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin53) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  negbin53
BP = 807.5, df = 35, p-value < 2.2e-16
```

Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.

HWrobstder <- sqrt(diag(vcovHC(negbin53, type="HC1"))) # produces Huber-White robust standard errors

Values	
HWrobstder	Named num [1:36] 0.074 0.059 0.0496 0.0418 0.0437 ...

```
stargazer(negbin53, negbin53,
        apply.coef = exp, t.auto=F, p.auto = F,
        se=list(NULL, HWrobstder),
        title="Negative Binomial Results", type="text",
        column.labels=c("Normal SE", "HW-Robust SE"),
        df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Negative Binomial Results

	Dependent variable:	
	returnquantity	
	Normal SE (1)	HW-Robust SE (2)
BIE_timeline_PC	1.518*** (0.058)	1.518*** (0.059)
group_store_PC	6.084*** (0.051)	6.084*** (0.050)
factor_PC2	1.279*** (0.047)	1.279*** (0.042)
factor_PC3	0.832*** (0.049)	0.832*** (0.044)
factor_PC4	1.227*** (0.045)	1.227*** (0.039)
factor_PC5	0.891* (0.046)	0.891* (0.042)
factor_PC6	0.877** (0.048)	0.877** (0.044)
factor_PC7	0.183*** (0.066)	0.183*** (0.074)
factor_PC8	0.251*** (0.097)	0.251*** (0.083)
factor_PC9	0.408*** (0.074)	0.408*** (0.098)
factor_PC10	0.139*** (0.171)	0.139*** (0.151)
factor_PC11	0.691*** (0.049)	0.691*** (0.045)
factor_PC12	0.906* (0.046)	0.906* (0.041)
factor_PC13	0.576*** (0.051)	0.576*** (0.045)
factor_PC14	0.559*** (0.055)	0.559*** (0.054)
factor_PC15	0.057*** (0.561)	0.057*** (0.431)
factor_PC17	0.396*** (0.073)	0.396*** (0.060)
factor_PC20	1.047 (0.048)	1.047 (0.044)
factor_PC21	0.807*** (0.046)	0.807*** (0.042)
salesquantity	1.009*** (0.0001)	1.009*** (0.0001)
avg_female	1.098* (0.039)	1.098* (0.031)
avg_age	0.917*** (0.005)	0.917*** (0.004)
avg_income	1.086*** (0.009)	1.086*** (0.007)
factor_Month2	0.909* (0.045)	0.909* (0.039)
factor_Month3	0.883** (0.045)	0.883** (0.040)
factor_Month4	0.792*** (0.046)	0.792*** (0.040)
factor_Month5	0.849*** (0.045)	0.849*** (0.040)
factor_Month6	0.839*** (0.047)	0.839*** (0.047)
factor_Month7	0.913* (0.046)	0.913* (0.048)
factor_Month8	0.822*** (0.042)	0.822*** (0.038)
factor_Month9	0.782*** (0.043)	0.782*** (0.037)
factor_Month10	0.852*** (0.046)	0.852*** (0.040)
factor_Month11	0.921 (0.044)	0.921 (0.039)
factor_Month12	1.072 (0.043)	1.072 (0.039)
BIE_timeline_PC:group_store_PC	0.615*** (0.061)	0.615*** (0.062)
Constant	0.468*** (0.082)	0.468*** (0.074)
Observations	21,115	21,115
Log Likelihood	-43,322.610	-43,322.610
theta	0.905*** (0.014)	0.905*** (0.014)
Akaike Inf. Crit.	86,717.220	86,717.220
Note:	*p<0.05; **p<0.01; ***p<0.001	

Visualize the output

meffects53 <- ggpredict(negbin53, terms=c("BIE_timeline_PC", "group_store_PC")) # generates a tidy data frame at two different values of BIE_timeline - before and after BOPS implementation.



ggplot(meffects53,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline_PC") + ylab("Return Quantity") +
  labs(colour="group_store_PC") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time = 0", "Time = 1")) +
  theme(axis.title.x=element_blank())



The interaction coefficient of BIE_timeline_PC:group_store_PC is significant with IRR value 0.615. This means that BOPS implementation is associated with 38.5% decrease in product-level return quantity.

####=====================IMPACT ON RETURN VALUE==============================####

#Since return value is a dollar value, we will use a linear interaction model with log transformed dependent variable.

sales5$log_returnvalue <- log(1+sales5$returnvalue)

BA Project_BOPS_Final Submitted.Rmd | meffects52 | meffects51 | sales5

Filter

	avg_residency	avg_childowner	BIE_timeline_PC	group_store_PC	factor_PC	factor_Month	log_salesvalue	log_returnvalue
296	7.037037	0.36363637	1	1	5	3	10.575784	8.482552
000	4.000000	0.00000000	0	1	4	8	5.309158	0.000000
429	8.428572	0.28571430	0	1	5	8	6.730517	0.000000
000	10.000000	0.00000000	0	1	13	8	5.319492	0.000000
000	8.000000	1.00000000	0	1	2	8	3.911823	3.911823
508	7.628415	0.39071038	0	1	12	8	10.138572	7.918010

Showing 1 to 7 of 21,115 entries

```
ols54 =
lm(log_returnvalue~BIE_timeline_PC*group_store_PC+log_salesvalue+factor_PC+avg_female+avg_age+
avg_income+factor_Month, data = sales5)
stargazer(ols54,
      title="Regression Results", type="text",
      column.labels=c("Model-1"),
      df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))

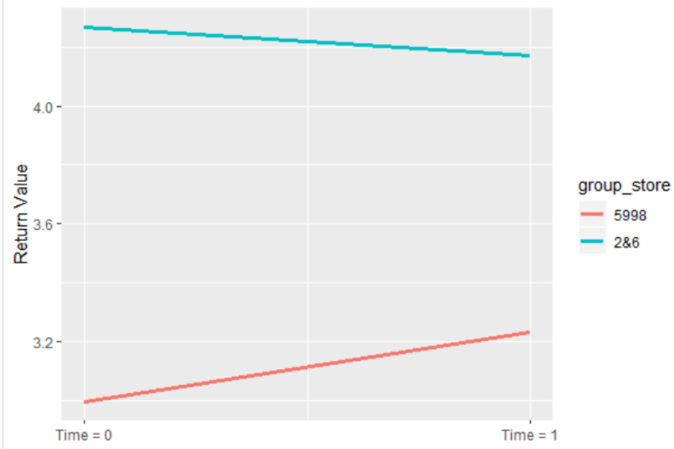
# Check for heteroskedasticity
gqtest(ols54 ) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols54) # Breusch-Pagan test indicates heteroskedasticity
```

```
> # Check for heteroskedasticity
> gqtest(ols54 ) # Goldfeld-Quandt test indicates no heteroskedasticity

        Goldfeld-Quandt test

data:  ols54
GQ = 1.6498, df1 = 10522, df2 = 10521, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols54) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols54
BP = 3378.7, df = 35, p-value < 2.2e-16
```

```
# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols54, type="HC1"))) # produces Huber-White robust standard errors
```

Values	
HWrobstder	Named num [1:36] 0.1636 0.0768 0.0698 0.0115 0.091…

```
stargazer(ols54, ols54,
      se=list(NULL, HWrobstder),
      title="Regression Results", type="text",
      column.labels=c("Normal SE", "HW-Robust SE"),
      df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

Regression Results

	Dependent variable:	
	log_returnvalue	
	Normal SE	HW-Robust SE
	(1)	(2)
BIE_timeline_PC	0.236*** (0.070)	0.236** (0.077)
group_store_PC	1.271*** (0.064)	1.271*** (0.070)
log_salesvalue	1.349*** (0.010)	1.349*** (0.011)
factor_PC2	1.129*** (0.085)	1.129*** (0.091)
factor_PC3	0.041 (0.085)	0.041 (0.103)
factor_PC4	0.675*** (0.079)	0.675*** (0.088)
factor_PC5	0.584*** (0.080)	0.584*** (0.087)
factor_PC6	0.659*** (0.085)	0.659*** (0.095)
factor_PC7	0.591*** (0.093)	0.591*** (0.099)
factor_PC8	0.428** (0.140)	0.428* (0.172)
factor_PC9	0.608*** (0.118)	0.608*** (0.123)
factor_PC10	0.371 (0.210)	0.371 (0.257)
factor_PC11	0.184* (0.084)	0.184 (0.097)
factor_PC12	0.705*** (0.080)	0.705*** (0.086)
factor_PC13	0.295*** (0.085)	0.295** (0.096)
factor_PC14	1.184*** (0.095)	1.184*** (0.101)
factor_PC15	0.163 (0.450)	0.163 (0.489)
factor_PC17	-1.063*** (0.117)	-1.063*** (0.176)
factor_PC20	0.633*** (0.085)	0.633*** (0.096)
factor_PC21	0.567*** (0.080)	0.567*** (0.087)
avg_female	0.381*** (0.065)	0.381*** (0.075)
avg_age	-0.059*** (0.008)	-0.059*** (0.010)
avg_income	0.042** (0.015)	0.042* (0.017)
factor_Month2	-0.456*** (0.076)	-0.456*** (0.076)
factor_Month3	-0.427*** (0.077)	-0.427*** (0.079)
factor_Month4	-0.492*** (0.078)	-0.492*** (0.078)
factor_Month5	-0.501*** (0.076)	-0.501*** (0.077)
factor_Month6	-0.535*** (0.078)	-0.535*** (0.078)
factor_Month7	-0.425*** (0.077)	-0.425*** (0.078)
factor_Month8	-0.497*** (0.071)	-0.497*** (0.071)
factor_Month9	-0.533*** (0.073)	-0.533*** (0.073)
factor_Month10	-0.399*** (0.079)	-0.399*** (0.079)
factor_Month11	-0.391*** (0.076)	-0.391*** (0.074)
factor_Month12	-0.546*** (0.074)	-0.546*** (0.073)
BIE_timeline_PC:group_store_PC	-0.333*** (0.078)	-0.333*** (0.083)
Constant	-7.316*** (0.142)	-7.316*** (0.164)
Observations	21,115	21,115
R2	0.604	0.604
Adjusted R2	0.603	0.603
Residual Std. Error	2.218	2.218
F Statistic	918.503***	918.503***
Note:	*p<0.05; **p<0.01; ***p<0.001	

```
>
```

meffects56 <- ggpredict(ols54, terms=c("BIE_timeline_PC", "group_store_PC"))

BA Project_BOPS_Final Submitted.Rmd × | meffects56 × | meffects52

Filter

	x	predicted	conf.low	conf.high	group
1	0	2.996820	2.807414	3.186226	0
2	0	4.267977	4.106324	4.429630	1
3	1	3.232636	3.056981	3.408291	0
4	1	4.170777	4.009743	4.331812	1

```
ggplot(meffects56,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Return Value") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
```

scale_x_continuous(breaks=c(0,1), labels=c("Time = 0", "Time = 1")) +
theme(axis.title.x=element_blank())



The interaction coefficient of BIE_timeline_PC:group_store_PC is significant with value -0.333. This means that BOPS implementation is associated with 33.3% decrease in product-level return value.

```

#==================================================================================#

## Question6: How does the impact of implementing BOPS strategy vary across product categories?

#==================================================================================#

```{r}

#Descriptive stactistics

stargazer(ODS_prodcat, type="text", median=TRUE, iqr=TRUE,digits=3, title="Descriptive Statistics")

Descriptive Statistics

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| store_number | 30,475 | 1,398.810 | 2,533.160 | 2 | 2 | 6 | 6 | 5,998 |
| year | 30,475 | 2,011.630 | 0.949 | 2,010 | 2,011 | 2,012 | 2,012 | 2,013 |
| month_index | 30,475 | 31.096 | 10.292 | 13 | 22 | 31 | 40 | 48 |
| product_category | 30,475 | 9.335 | 6.316 | 1 | 4 | 7 | 13 | 21 |
| month_dummy | 30,475 | 6.541 | 3.554 | 1 | 3 | 6 | 10 | 12 |
| bops_in_effect | 30,475 | 0.595 | 0.491 | 0 | 0 | 1 | 1 | 1 |
| day | 30,475 | 566.639 | 313.415 | 1 | 292 | 569 | 846 | 1,096 |
| salesvalue | 30,475 | 9,342.367 | 20,865.000 | 0.000 | 546.295 | 1,950.680 | 8,952.230 | 424,294.600 |
| returnvalue | 30,475 | 1,439.134 | 3,166.514 | 0.000 | 0.000 | 199.000 | 1,398.860 | 52,508.210 |
| salesquantity | 30,475 | 54.848 | 162.805 | 1 | 3 | 9 | 34 | 5,111 |
| returnquantity | 30,475 | 5.540 | 13.509 | 0 | 0 | 1 | 5 | 380 |
| avg_female | 23,277 | 0.501 | 0.285 | 0.000 | 0.333 | 0.500 | 0.667 | 1.000 |
| avg_age | 23,795 | 4.935 | 2.115 | 0.000 | 4.000 | 4.907 | 5.913 | 13.000 |
| avg_income | 23,763 | 5.255 | 1.194 | 1.000 | 4.808 | 5.333 | 5.800 | 9.000 |
| avg_homeowner | 23,763 | 0.645 | 0.251 | 0.000 | 0.534 | 0.667 | 0.778 | 1.000 |
| avg_residency | 23,763 | 6.893 | 2.815 | 0.000 | 5.615 | 6.910 | 7.958 | 15.000 |
| avg_childowner | 23,763 | 0.377 | 0.250 | 0.000 | 0.250 | 0.368 | 0.481 | 1.000 |

describeBy(ODS_prodcat, ODS_prodcat$product_category)

```
 Descriptive statistics by group
group: 1
                vars   n   mean   sd median trimmed   mad min   max   range skew kurtosis   se
store_number   1 2059 1150.74 2358.54   6.00   690.95   5.93   2   5998.00   5996.00 1.57   0.46 51.98
year   2 2059 2011.67   0.95 2012.00 2011.71   1.48 2010   2013.00   3.00 -0.15   -0.92   0.02
month*   3 2059   NaN   NA   NA   NaN   NA Inf   -Inf   -Inf   NA   NA   NA
month_index   4 2059   31.61   10.29   32.00   31.78   13.34   13   48.00   35.00 -0.13   -1.19   0.23
product_category   5 2059   1.00   0.00   1.00   1.00   0.00   1   1.00   0.00   NaN   NaN   0.00
month_dummy   6 2059   6.59   3.56   7.00   6.59   4.45   1   12.00   11.00 0.01   -1.27   0.08
bops_in_effect   7 2059   0.63   0.48   1.00   0.67   0.00   0   1.00   1.00 -0.56   -1.69   0.01
day   8 2059   582.14   313.31 594.00   587.64 401.78   2   1096.00   1094.00 -0.13   -1.19   6.90
salesvalue   9 2059 16232.94 22216.16 4458.22 11848.96 5669.52   0 188843.50 188843.50 2.20   7.11 489.60
returnvalue   10 2059 3406.79 5082.82 869.98 2307.84 1289.83   0 31211.98 31211.98 2.00   4.22 112.02
salesquantity   11 2059   17.54   24.44   4.00   12.58   4.45   1   203.00   202.00 2.28   7.51   0.54
returnquantity   12 2059   3.32   4.76   1.00   2.34   1.48   0   31.00   31.00 2.00   4.67   0.10
avg_female   13 1626   0.35   0.27   0.33   0.32   0.22   0   1.00   1.00 0.79   0.42   0.01
avg_age   14 1671   4.54   2.12   4.39   4.47   1.32   0   13.00   13.00 0.54   1.91   0.05
avg_income   15 1668   4.81   1.26   4.87   4.85   0.83   1   9.00   8.00 -0.33   2.06   0.03
avg_homeowner   16 1668   0.59   0.27   0.60   0.61   0.18   0   1.00   1.00 -0.40   0.03   0.01
avg_residency   17 1668   6.60   3.04   6.50   6.45   1.88   0   15.00   15.00 0.59   1.28   0.07
avg_childowner   18 1668   0.36   0.27   0.34   0.33   0.18   0   1.00   1.00 0.77   0.55   0.01

```

#Similarly group 2 to group 20 results are also displayed

# creating groups of store with BOPS implementation
sales6 <- ODS_prodcat[(ODS_prodcat$day<786),]
#Add a BIE timeline variable dividing the data into two groups, one before Aug.1, 2011 and the other after.
#Add a grouping variable for stores dividing the data into two groups, one for 2&6 and the other for 5998.
sales6$BIE_timeline <- ifelse(sales6$day<366,0,1)
sales6$group_store <- ifelse((sales6$store_number==2) | (sales6$store_number==6),1,0)

#Assuming mean value for all 'na' values
sales6$avg_female[is.na(sales6$avg_female)] <- mean(sales6$avg_female, na.rm = TRUE)
sales6$avg_age[is.na(sales6$avg_age)] <- mean(sales6$avg_age, na.rm = TRUE)
sales6$avg_income[is.na(sales6$avg_income)] <- mean(sales6$avg_income, na.rm = TRUE)
sales6$avg_homeowner[is.na(sales6$avg_homeowner)] <- mean(sales6$avg_homeowner, na.rm = TRUE)
sales6$avg_residency[is.na(sales6$avg_residency)] <- mean(sales6$avg_residency, na.rm = TRUE)
sales6$avg_childowner[is.na(sales6$avg_childowner)] <- mean(sales6$avg_childowner, na.rm = TRUE)

BA Project_BOPS_Final Submitted.Rmd × | sales6 × | meffects56 × | meffects52 × | meffects51 × | sales5 ×

Filter

	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store
0	29	0.5092251	4.993311	5.612794	0.6296296	7.037037	0.36363637	1	1
1	0	1.0000000	0.000000	8.000000	1.0000000	4.000000	0.00000000	0	1
7	0	0.5000000	3.714286	4.428571	0.8571429	8.428572	0.28571430	0	1
2	0	0.0000000	2.500000	4.500000	0.5000000	10.000000	0.00000000	0	1
1	1	0.0000000	4.000000	2.000000	0.0000000	8.000000	1.00000000	0	1
7	34	0.6012461	5.046448	5.691257	0.6229508	7.628415	0.39071038	0	1
8	25	0.6894410	5.393259	4.966292	0.5842696	5.707865	0.34269664	0	1

Showing 1 to 8 of 21,003 entries

stargazer(sales6, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

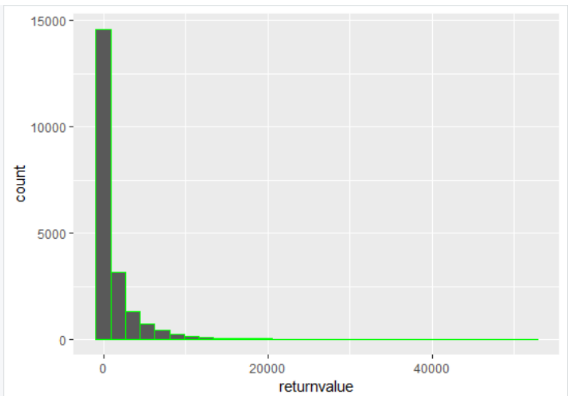
Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
store_number	21,003	1,283.7	2,456.6	2	2	6	6	5,998
year	21,003	2,011.2	0.7	2,010	2,011	2,011	2,012	2,012
month_index	21,003	25.6	7.3	13	19	26	32	38
product_category	21,003	9.4	6.3	1	4	7	13	21
month_dummy	21,003	6.7	3.5	1	4	7	10	12
bops_in_effect	21,003	0.4	0.5	0	0	0	1	1
day	21,003	399.7	222.2	1	204	408	587	785
salesvalue	21,003	8,734.2	19,168.4	0.0	507.9	1,857.3	8,538.8	402,473.4
returnvalue	21,003	1,375.3	2,997.0	0.0	0.0	194.5	1,376.5	51,973.1
salesquantity	21,003	49.9	139.9	1	2	9	32	4,474
returnquantity	21,003	5.3	12.6	0	0	1	5	380
avg_female	21,003	0.5	0.3	0.0	0.4	0.5	0.6	1.0
avg_age	21,003	5.0	1.9	0.0	4.4	5.0	5.7	13.0
avg_income	21,003	5.3	1.1	1.0	5.0	5.3	5.7	9.0
avg_homeowner	21,003	0.6	0.2	0.0	0.6	0.6	0.7	1.0
avg_residency	21,003	6.9	2.5	0.0	6.0	6.9	7.5	15.0
avg_childowner	21,003	0.4	0.2	0.0	0.3	0.4	0.4	1.0
BIE_timeline	21,003	0.5	0.5	0	0	1	1	1
group_store	21,003	0.8	0.4	0	1	1	1	1

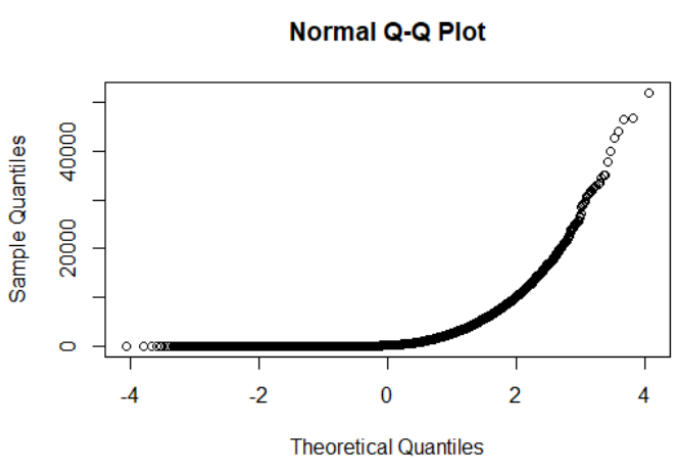
####=====================IMPACT ON RETURN VALUE======================####

#check normalization

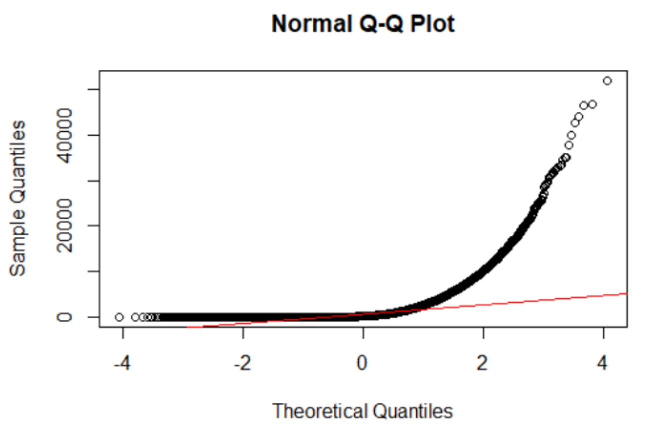
ggplot(sales6, aes(x=returnvalue)) + geom_histogram(colour="green", bins = 30)



qqnorm(sales6$returnvalue)



qqline(sales6$returnvalue, col=2) #distribution is not normal



#analyzing general trend between key independent and dependent variable using box plot

```
df61 <- data.frame(returnvalue=sales6$returnvalue, BIE_timeline=as.factor(sales6$BIE_timeline))
```

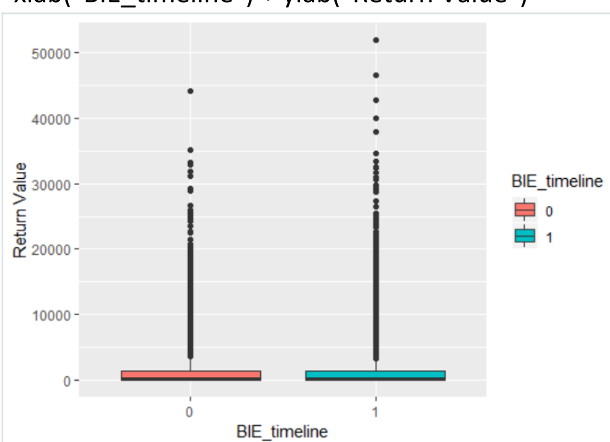
BA Project_BOPS_Final Submitted.Rmd

Filter

	returnvalue	BIE_timeline
1	4828.76	1
2	0.00	0
3	0.00	0
4	0.00	0
5	48.99	0

Showing 1 to 6 of 21,003 entries

```
ggplot(df61, aes(x=BIE_timeline, y=returnvalue, fill=BIE_timeline)) + geom_boxplot() +
  xlab("BIE_timeline") + ylab("Return Value")
```



test for multicollinearity

```
df62=sales6[c("BIE_timeline","group_store","month_dummy","product_category","salesvalue",
"avg_female","avg_age","avg_income")]
```

BA Project_BOPS_Final Submitted.Rmd × | df62 × | df61 × | sales6 × | meffects56 × | meffects52 × | meffects51 × | sales5 ×

Filter

	BIE_timeline	group_store	month_dummy	product_category	salesvalue	avg_female	avg_age	avg_income
1	1	1	3	5	39173.61	0.5092251	4.993311	5.612794
2	0	1	8	4	201.18	1.0000000	0.000000	8.000000
3	0	1	8	5	836.58	0.5000000	3.714286	4.428571
4	0	1	8	13	203.28	0.0000000	2.500000	4.500000
5	0	1	8	2	48.99	0.0000000	4.000000	2.000000

Showing 1 to 6 of 21,003 entries

stargazer(df62, type="text", median=TRUE, iqr=TRUE,digits=1, title="Descriptive Statistics")

```
Descriptive Statistics
================================================================================
Statistic          N      Mean    St. Dev. Min Pctl(25) Median  Pctl(75)    Max
--------------------------------------------------------------------------------
BIE_timeline     21,003   0.5      0.5      0     0        1       1         1
group_store      21,003   0.8      0.4      0     1        1       1         1
month_dummy      21,003   6.7      3.5      1     4        7      10        12
product_category 21,003   9.4      6.3      1     4        7      13        21
salesvalue       21,003 8,734.2 19,168.4 0.0   507.9   1,857.3 8,538.8 402,473.4
avg_female       21,003   0.5      0.3     0.0   0.4      0.5     0.6       1.0
avg_age          21,003   5.0      1.9     0.0   4.4      5.0     5.7      13.0
avg_income       21,003   5.3      1.1     1.0   5.0      5.3     5.7       9.0
--------------------------------------------------------------------------------
```

round(cor(df62),3)

```
> round(cor(df62),3)
                 BIE_timeline group_store month_dummy product_category salesvalue avg_female avg_age avg_income
BIE_timeline            1.000      -0.096       0.039           -0.007      0.007     -0.018  -0.085     -0.032
group_store            -0.096       1.000       0.021           -0.007      0.200      0.012   0.012     -0.012
month_dummy             0.039       0.021       1.000           -0.001      0.091     -0.030   0.050      0.043
product_category       -0.007      -0.007      -0.001            1.000     -0.073      0.066   0.070      0.093
salesvalue              0.007       0.200       0.091           -0.073      1.000     -0.129  -0.047      0.031
avg_female             -0.018       0.012      -0.030            0.066     -0.129      1.000   0.077     -0.011
avg_age                -0.085       0.012       0.050            0.070     -0.047      0.077   1.000      0.172
avg_income             -0.032      -0.012       0.043            0.093      0.031     -0.011   0.172      1.000
```

vifcor(df62)# No variable from the 9 input variables has collinearity problem.

```
No variable from the 8 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( avg_female ~ group_store ):  0.004545106
max correlation ( salesvalue ~ group_store ):  0.2054456

---------- VIFs of the remained variables --------
         Variables      VIF
1     BIE_timeline 1.019542
2      group_store 1.057619
3      month_dummy 1.012736
4 product_category 1.021246
5       salesvalue 1.086274
6       avg_female 1.039312
7          avg_age 1.059098
8       avg_income 1.039270
```

#converting into factor variables
sales6$BIE_timeline<- as.factor(sales6$BIE_timeline)
sales6$month_dummy<- as.factor(sales6$month_dummy)
sales6$product_category<- as.factor(sales6$product_category)
#Since sales value is a dollar value, we will use a linear interaction model with log transformed dependent variable.

sales6$log_salesvalue <- log(1+sales6$salesvalue)
sales6$log_returnvalue <- log(1+sales6$returnvalue)

BA Project_BOPS_Final Submitted.Rmd × sales6 ×

Filter

ity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue	log_returnvalue
29	0.5092251	4.993311	5.612794	0.6296296	7.037037	0.36363637	1	1	10.575784	8.482552
0	1.0000000	0.000000	8.000000	1.0000000	4.000000	0.00000000	0	1	5.309158	0.000000
0	0.5000000	3.714286	4.428571	0.8571429	8.428572	0.28571430	0	1	6.730517	0.000000
0	0.0000000	2.500000	4.500000	0.5000000	10.000000	0.00000000	0	1	5.319492	0.000000
1	0.0000000	4.000000	2.000000	0.0000000	8.000000	1.00000000	0	1	3.911823	3.911823

Showing 1 to 6 of 21,003 entries

we want to analyse the imapct of BOPS implementation across all the product category, so we will use triple interaction

ols61 = lm(log_returnvalue~BIE_timeline*group_store*product_category+log_salesvalue+ month_dummy+ avg_female+avg_age+avg_income, data = sales6)
stargazer(ols61,
title="Regression Results", type="text",
column.labels=c("Model-Return value "),
df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))

Check for heteroskedasticity

gqtest(ols61) # Goldfeld-Quandt test indicates heteroskedasticity
bptest(ols61) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  ols61
GQ = 1.6404, df1 = 10415, df2 = 10414, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols61) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols61
BP = 3471.4, df = 78, p-value < 2.2e-16
```

Since there is heteroskedasticity in the data, we will replace SEs with clustered robust SEs.

HWrobstder <- sqrt(diag(vcovHC(ols61, type="HC1"))) # produces Huber-White robust standard errors
clusrobstder <- sqrt(diag(cluster.vcov(ols61, sales6$store_number)))
# produces clustered robust standard errors using store_number as clustering variable

Values	
clusrobstder	Named num [1:79] 0.37116 0.02824 0.570...
HWrobstder	Named num [1:79] 0.334 0.385 0.323 0.3...

stargazer(ols61, ols61, ols61,
se=list(NULL, HWrobstder, clusrobstder),
title="Regression Results", type="text",
column.labels=c("Normal SE", "HW-Robust SE", "Clustered SE"),

df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))

Regression Results

	Dependent variable:		
	log_returnvalue		
	Normal SE	HW-Robust SE	Clustered SE
	(1)	(2)	(3)
BIE_timeline1	0.450	0.450	0.450***
	(0.319)	(0.385)	(0.028)
group_store	2.853***	2.853***	2.853***
	(0.282)	(0.323)	(0.570)
product_category2	2.503***	2.503***	2.503***
	(0.362)	(0.395)	(0.078)
product_category3	1.293***	1.293**	1.293***
	(0.355)	(0.443)	(0.004)
product_category4	1.659***	1.659***	1.659***
	(0.315)	(0.367)	(0.008)
product_category5	2.131***	2.131***	2.131***
	(0.311)	(0.353)	(0.028)
product_category6	1.771***	1.771***	1.771***
	(0.347)	(0.372)	(0.057)
product_category7	2.562***	2.562***	2.562***
	(0.319)	(0.330)	(0.114)
product_category8	0.406*	0.406	0.406
	(0.175)	(0.217)	(0.580)
product_category9	3.278***	3.278***	3.278***
	(0.717)	(0.504)	(0.154)
product_category10	0.425	0.425	0.425
	(0.240)	(0.296)	(0.664)
product_category11	1.841***	1.841***	1.841***
	(0.328)	(0.378)	(0.019)
product_category12	2.336***	2.336***	2.336***
	(0.312)	(0.347)	(0.036)
product_category13	0.992**	0.992**	0.992***
	(0.335)	(0.381)	(0.028)
product_category14	3.367***	3.367***	3.367***
	(0.388)	(0.341)	(0.150)
product_category15	0.158	0.158	0.158
	(0.493)	(0.570)	(0.589)
product_category17	-1.180***	-1.180***	-1.180*
	(0.172)	(0.253)	(0.487)
product_category20	1.564***	1.564***	1.564***
	(0.334)	(0.374)	(0.032)
product_category21	2.065***	2.065***	2.065***
	(0.314)	(0.345)	(0.051)
log_salesvalue	1.348***	1.348***	1.348***
	(0.010)	(0.012)	(0.053)
month_dummy2	-0.449***	-0.449***	-0.449***
	(0.076)	(0.075)	(0.036)
month_dummy3	-0.423***	-0.423***	-0.423***
	(0.077)	(0.078)	(0.116)
month_dummy4	-0.492***	-0.492***	-0.492***
	(0.078)	(0.077)	(0.116)
month_dummy5	-0.494***	-0.494***	-0.494***
	(0.076)	(0.076)	(0.094)
month_dummy6	-0.534***	-0.534***	-0.534***
	(0.078)	(0.078)	(0.081)
month_dummy7	-0.431***	-0.431***	-0.431***
	(0.077)	(0.077)	(0.055)
month_dummy8	-0.497***	-0.497***	-0.497***
	(0.071)	(0.071)	(0.145)
month_dummy9	-0.529***	-0.529***	-0.529***
	(0.074)	(0.074)	(0.054)
month_dummy10	-0.400***	-0.400***	-0.400***
	(0.079)	(0.079)	(0.103)
month_dummy11	-0.380***	-0.380***	-0.380***
	(0.075)	(0.074)	(0.062)
month_dummy12	-0.524***	-0.524***	-0.524***
	(0.073)	(0.073)	(0.064)
avg_female	0.441***	0.441***	0.441
	(0.066)	(0.077)	(0.274)
avg_age	-0.056***	-0.056***	-0.056***
	(0.008)	(0.010)	(0.016)
avg_income	0.049**	0.049**	0.049*
	(0.015)	(0.017)	(0.020)
BIE_timeline1:group_store	-0.519	-0.519	-0.519***
	(0.344)	(0.415)	(0.047)
BIE_timeline1:product_category2	-0.306	-0.306	-0.306***
	(0.438)	(0.498)	(0.006)
BIE_timeline1:product_category3	-0.348	-0.348	-0.348***
	(0.427)	(0.550)	(0.015)
BIE_timeline1:product_category4	0.499	0.499	0.499***
	(0.383)	(0.464)	(0.010)
BIE_timeline1:product_category5	-0.522	-0.522	-0.522***
	(0.382)	(0.452)	(0.013)
BIE_timeline1:product_category6	-0.404	-0.404	-0.404***
	(0.429)	(0.480)	(0.016)
BIE_timeline1:product_category7	-0.380	-0.380	-0.380***
	(0.394)	(0.419)	(0.017)
BIE_timeline1:product_category8	-0.784*	-0.784*	-0.784***
	(0.305)	(0.355)	(0.141)
BIE_timeline1:product_category9	-1.386	-1.386*	-1.386***
	(0.766)	(0.592)	(0.045)
BIE_timeline1:product_category10	-1.621**	-1.621**	-1.621***
	(0.526)	(0.542)	(0.039)
BIE_timeline1:product_category11	-0.868*	-0.868	-0.868***
	(0.401)	(0.478)	(0.028)
BIE_timeline1:product_category12	-0.195	-0.195	-0.195***
	(0.382)	(0.444)	(0.011)
BIE_timeline1:product_category13	0.415	0.415	0.415***
	(0.412)	(0.481)	(0.038)
BIE_timeline1:product_category14	-0.796	-0.796	-0.796***
	(0.509)	(0.455)	(0.003)
BIE_timeline1:product_category15	-1.509	-1.509*	-1.509***
	(1.213)	(0.626)	(0.109)
BIE_timeline1:product_category17	-0.223	-0.223	-0.223***
	(0.238)	(0.357)	(0.033)
BIE_timeline1:product_category20	0.293	0.293	0.293***
	(0.410)	(0.486)	(0.015)
BIE_timeline1:product_category21	-0.141	-0.141	-0.141***
	(0.384)	(0.444)	(0.007)
group_store:product_category2	-1.769***	-1.769***	-1.769***
	(0.385)	(0.418)	(0.216)
group_store:product_category3	-1.426***	-1.426**	-1.426***
	(0.381)	(0.470)	(0.041)
group_store:product_category4	-1.287***	-1.287***	-1.287*
	(0.340)	(0.390)	(0.518)
group_store:product_category5	-1.806***	-1.806***	-1.806***
	(0.338)	(0.377)	(0.477)
group_store:product_category6	-1.320***	-1.320***	-1.320***
	(0.372)	(0.399)	(0.156)
group_store:product_category7	-2.511***	-2.511***	-2.511***
	(0.353)	(0.364)	(0.695)
group_store:product_category8			
group_store:product_category9	-2.564***	-2.564***	-2.564***
	(0.750)	(0.537)	(0.694)
group_store:product_category10			
group_store:product_category11	-1.884***	-1.884***	-1.884***
	(0.356)	(0.408)	(0.050)
group_store:product_category12	-1.988***	-1.988***	-1.988**
	(0.338)	(0.371)	(0.664)
group_store:product_category13	-0.951**	-0.951*	-0.951***
	(0.362)	(0.409)	(0.162)
group_store:product_category14	-2.324***	-2.324***	-2.324***
	(0.412)	(0.366)	(0.489)
group_store:product_category15			
group_store:product_category17			
group_store:product_category20	-1.130**	-1.130**	-1.130***
	(0.362)	(0.402)	(0.203)
group_store:product_category21	-1.734***	-1.734***	-1.734***
	(0.340)	(0.369)	(0.462)

```
BIE_timeline1:group_store:product_category2     0.611      0.611      0.611***
                                                (0.474)    (0.531)    (0.013)

BIE_timeline1:group_store:product_category3     0.338      0.338      0.338
                                                (0.467)    (0.592)    (0.264)

BIE_timeline1:group_store:product_category4     -0.448     -0.448     -0.448*
                                                (0.422)    (0.499)    (0.175)

BIE_timeline1:group_store:product_category5     0.540      0.540      0.540***
                                                (0.421)    (0.486)    (0.054)

BIE_timeline1:group_store:product_category6     0.465      0.465      0.465***
                                                (0.465)    (0.519)    (0.064)

BIE_timeline1:group_store:product_category7     0.465      0.465      0.465*
                                                (0.444)    (0.470)    (0.188)

BIE_timeline1:group_store:product_category8

BIE_timeline1:group_store:product_category9     0.715      0.715      0.715
                                                (0.813)    (0.643)    (0.602)

BIE_timeline1:group_store:product_category10

BIE_timeline1:group_store:product_category11    0.911*     0.911      0.911***
                                                (0.442)    (0.522)    (0.078)

BIE_timeline1:group_store:product_category12    0.216      0.216      0.216
                                                (0.421)    (0.477)    (0.128)

BIE_timeline1:group_store:product_category13    -0.267     -0.267     -0.267***
                                                (0.452)    (0.521)    (0.026)

BIE_timeline1:group_store:product_category14    0.433      0.433      0.433***
                                                (0.544)    (0.497)    (0.121)

BIE_timeline1:group_store:product_category15

BIE_timeline1:group_store:product_category17

BIE_timeline1:group_store:product_category20    -0.315     -0.315     -0.315***
                                                (0.451)    (0.527)    (0.028)

BIE_timeline1:group_store:product_category21    0.007      0.007      0.007
                                                (0.422)    (0.479)    (0.174)

Constant                                        -8.745***  -8.745***  -8.745***
                                                (0.292)    (0.334)    (0.371)

----------------------------------------------------------------------------------
Observations                                    21,003     21,003     21,003
R2                                              0.608      0.608      0.608
Adjusted R2                                     0.607      0.607      0.607
Residual Std. Error                             2.208      2.208      2.208
F Statistic                                     416.729*** 416.729*** 416.729***
==================================================================================
Note:                                           *p<0.05; **p<0.01; ***p<0.001
>
```

displays normal/HW robust/clustered robust standard errors.

###----------Marginal Effects --------------------###

meffects61 <- ggpredict(ols61, terms=c("BIE_timeline", "group_store","product_category"))

BA Project_BOPS_Final Submitted.Rmd × | meffects61 ×

Filter

	x	predicted	conf.low	conf.high	group	facet
1	0	1.6393230	1.11125151	2.167395	0	1
2	0	4.1420814	3.64492469	4.639238	0	2
3	0	2.9322202	2.45686513	3.407575	0	3
4	0	3.2982594	2.94806578	3.648453	0	4
5	0	3.7701993	3.43128916	4.109109	0	5
6	0	3.4102912	2.95623243	3.864350	0	6

Showing 1 to 6 of 72 entries

Replacing by actual variable names for convenience

meffects61_colnames <- meffects61[c("x","predicted","group", "facet")]

BA Project_BOPS_Final Submitted.Rmd × | meffects61_colnames ×

Filter

	x	predicted	group	facet
1	0	1.6393230	0	1
2	0	4.1420814	0	2
3	0	2.9322202	0	3
4	0	3.2982594	0	4
5	0	3.7701993	0	5
6	0	3.4102912	0	6
7	0	4.2015701	0	7
8	0	2.0454564	0	8

Showing 1 to 9 of 72 entries

colnames(meffects61_colnames) <- c("BIE_timeline", "predicted_return_value","group_store" , "product_category")

BA Project_BOPS_Final Submitted.Rmd × | meffects61_colnames ×

Filter

	BIE_timeline	predicted_return_value	group_store	product_category
1	0	1.6393230	0	1
2	0	4.1420814	0	2
3	0	2.9322202	0	3
4	0	3.2982594	0	4
5	0	3.7701993	0	5
6	0	3.4102912	0	6
7	0	4.2015701	0	7
8	0	2.0454564	0	8

Showing 1 to 9 of 72 entries

levels(meffects61_colnames$product_category) <- c('1-Bridal', '2-Gold Wed Bands', '3-Solitaires', '4-Diamond Fashion', '5-Semi Precious', '6-Mens', '7-Gold Earrings', '8-In House Special Event', '9-Beads','10-Piercings / Close Out', '11-Diamond Solitaires Jewelry', '12-Gold Chain / Jewelry ', '13-Watches', '14-Pre-Owned', '15-Specialized Jewelry', '17-Events', '19-Repair / Warranty','20-Diamond Wedding Bands','21-Sterling Silver')

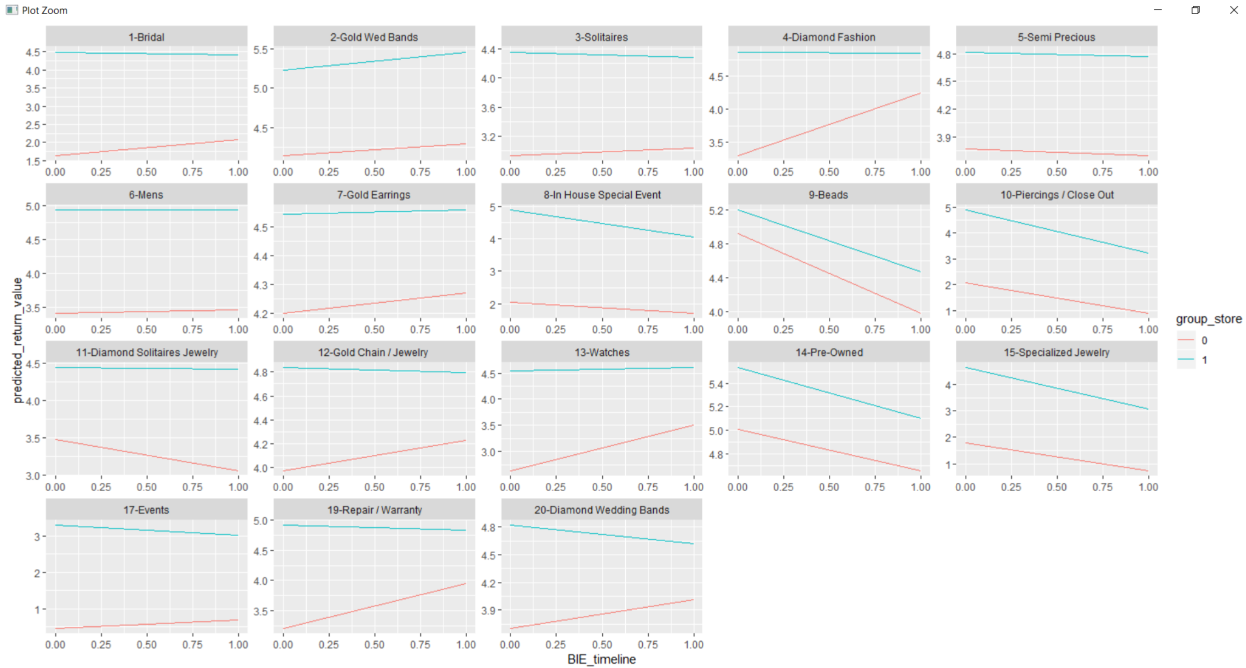
BA Project_BOPS_Final Submitted.Rmd × | meffects61_colnames ×

Filter

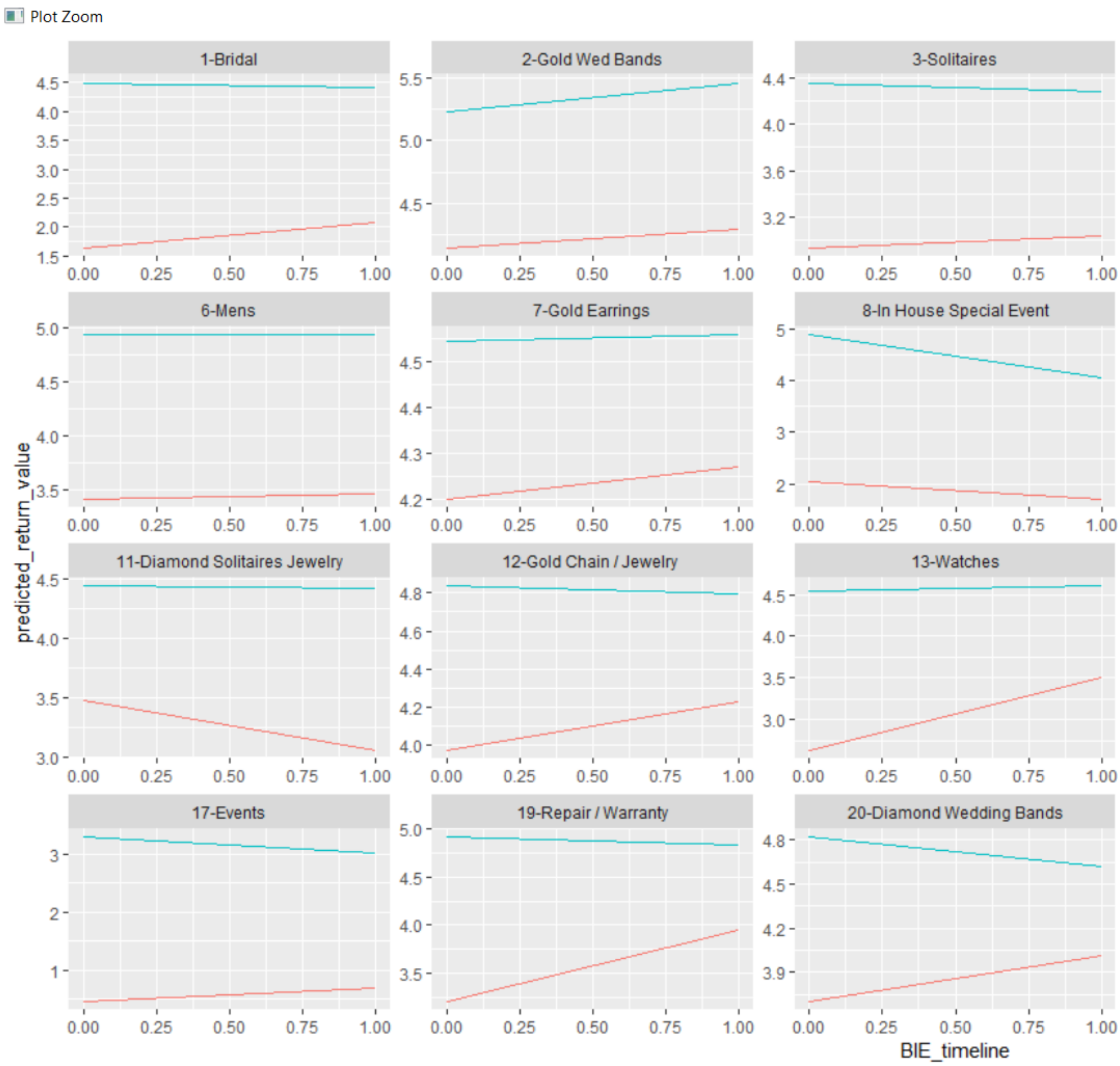
	BIE_timeline	predicted_return_value	group_store	product_category
1	0	1.6393230	0	1-Bridal
2	0	4.1420814	0	2-Gold Wed Bands
3	0	2.9322202	0	3-Solitaires
4	0	3.2982594	0	4-Diamond Fashion
5	0	3.7701993	0	5-Semi Precious
6	0	3.4102912	0	6-Mens
7	0	4.2015701	0	7-Gold Earrings
8	0	2.0454564	0	8-In House Special Event

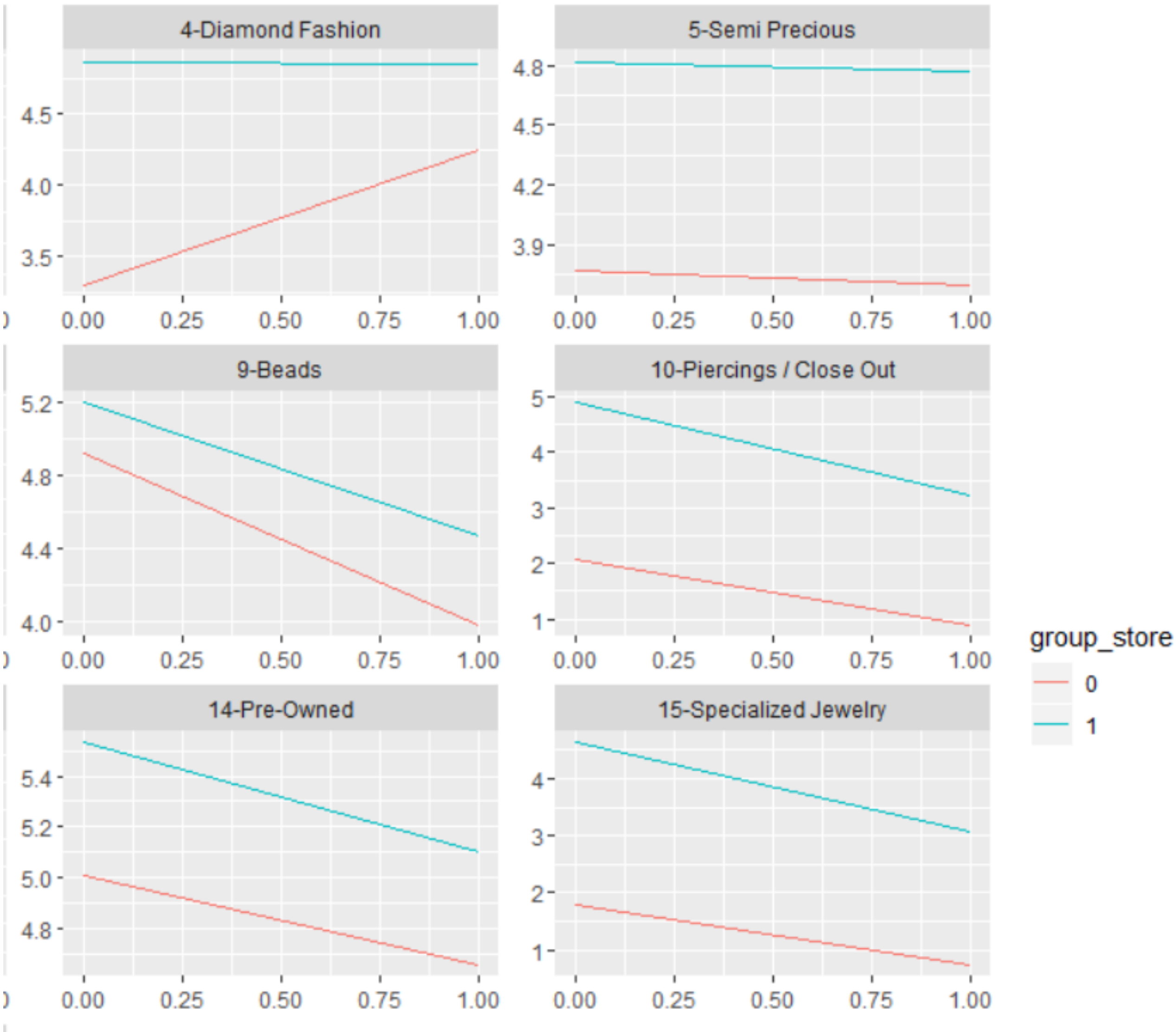
Showing 1 to 9 of 72 entries

```
ggplot(meffects61_colnames, aes(BIE_timeline,
 y= predicted_return_value, colour =group_store)) +
 geom_line() +
 facet_wrap(product_category ~ ., scales="free")
```



Zoomed version:


Plot Zoom
1-Bridal
2-Gold Wed Bands
3-Solitaires
6-Mens
7-Gold Earrings
8-In House Special Event
11-Diamond Solitaires Jewelry
12-Gold Chain / Jewelry
13-Watches
17-Events
19-Repair / Warranty
20-Diamond Wedding Bands
predicted_return_value
BIE_timeline



####------Subsetting data based on product categories-------###

#Checking predicted values for all the product categories and analyzing impact of BOPS implementation for both storegroups(group 0 = 5998 and group 1 = 2 and 6)
#Change product_category epeating the code for all

df_pc1<- subset(sales6, product_category==1)

BA Project_BOPS_Final Submitted.Rmd | df_pc1 | meffects61_colnames

Filter

	store_number	year	month	month_index	product_category	month_dummy	bops_in_effect	day	salesvalue	returnvalue	salesquantity	returnquantity	avg_female
13	2	2010	AUG	13	1	8	0	2	12253.53	1478.82	13	2	0.454
33	2	2010	AUG	13	1	8	0	3	7989.74	1287.40	22	3	0.555
45	2	2010	AUG	13	1	8	0	4	26309.46	10614.14	26	5	0.440
61	2	2010	AUG	13	1	8	0	5	6399.69	771.66	12	1	0.500
76	2	2010	AUG	13	1	8	0	6	14782.14	3143.39	11	4	0.363
92	2	2010	AUG	13	1	8	0	9	19136.12	1360.78	19	3	0.533
116	2	2010	AUG	13	1	8	0	10	10068.10	2699.42	21	6	0.476

Showing 1 to 9 of 1,381 entries

BA Project_BOPS_Final Submitted.Rmd × | df_pc1 × | meffects61_colnames ×

Filter

esquantity	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue	log_returnvalue
13	2	0.4545455	4.615385	5.769231	0.3076923	6.846154	0.07692308	0	1	9.413651	7.299676
22	3	0.5555556	3.727273	4.590909	0.4090909	4.863636	0.27272728	0	1	8.986039	7.161156
26	5	0.4400000	5.461538	5.115385	0.5769231	6.000000	0.42307693	0	1	10.177722	9.270037
12	1	0.5000000	7.250000	3.750000	0.5833333	7.333333	0.25000000	0	1	8.764161	6.649839
11	4	0.3636364	6.272727	3.727273	0.5454546	5.727273	0.18181819	0	1	9.601243	8.053375
19	3	0.5333334	3.473684	4.736842	0.6315789	4.473684	0.31578946	0	1	9.859385	7.216548
21	6	0.4761905	4.142857	6.333333	0.8095238	6.142857	0.42857143	0	1	9.217227	7.901163

Showing 1 to 9 of 1,381 entries

ols_pc1= lm(log_returnvalue~BIE_timeline*group_store+log_salesvalue+ month_dummy+ avg_female+avg_age+avg_income, data = df_pc1)

stargazer(ols_pc1,
    title="Regression Results", type="text",
    column.labels=c("ME-pc1"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))

Check for heteroskedasticity
gqtest(ols_pc1) # Goldfeld-Quandt test indicates heteroskedasticity
bptest(ols_pc1) # Breusch-Pagan test indicates heteroskedasticity

```
        Goldfeld-Quandt test

data:  ols_pc1
GQ = 2.3917, df1 = 672, df2 = 671, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols_pc1) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols_pc1
BP = 178.49, df = 18, p-value < 2.2e-16
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols_pc1, type="HC1"))) # produces Huber-White robust standard errors
clusrobstder <- sqrt(diag(cluster.vcov(ols_pc1, sales6$store_number))) # produces clustered robust standard errors

Values	
clusrobstder	Named num [1:79] 0.37116 0.02824 0.…
HWrobstder	Named num [1:19] 0.69 0.4 0.361 0.0…

stargazer(ols_pc1, ols_pc1, ols_pc1,
    se=list(NULL, HWrobstder, clusrobstder),
    title="REgression Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE", "Clustered SE"),
    df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001)) # displays normal/HW robust/clustered robust standard errors. With clustered robust standard errors

```
REgression Results
=================================================================
                              Dependent variable:
                    ---------------------------------------------
                                 log_returnvalue
                     Normal SE  HW-Robust SE  Clustered SE
                        (1)         (2)           (3)
-----------------------------------------------------------------
BIE_timeline1          0.378       0.378        0.378***
                      (0.378)     (0.400)      (0.028)

group_store           2.535***    2.535***     2.535***
                      (0.343)     (0.361)      (0.570)

log_salesvalue        1.666***    1.666***     1.666***
                      (0.044)     (0.057)      (0.053)

month_dummy2          -0.325      -0.325      -0.325***
                      (0.356)     (0.350)      (0.036)

month_dummy3          -0.121      -0.121       -0.121
                      (0.360)     (0.369)      (0.116)

month_dummy4          -0.230      -0.230       -0.230*
                      (0.365)     (0.360)      (0.116)

month_dummy5          -0.171      -0.171       -0.171
                      (0.359)     (0.363)      (0.094)

month_dummy6          -0.046      -0.046       -0.046
                      (0.362)     (0.363)      (0.081)

month_dummy7          -0.105      -0.105       -0.105
                      (0.356)     (0.349)      (0.055)

month_dummy8          -0.289      -0.289       -0.289*
                      (0.336)     (0.347)      (0.145)

month_dummy9          -0.500      -0.500      -0.500***
                      (0.348)     (0.337)      (0.054)

month_dummy10          0.489       0.489       0.489***
                      (0.371)     (0.383)      (0.103)

month_dummy11          0.266       0.266       0.266***
                      (0.352)     (0.347)      (0.062)

month_dummy12         -0.150      -0.150       -0.150*
                      (0.341)     (0.346)      (0.064)

avg_female            1.271***    1.271***     1.271***
                      (0.282)     (0.337)      (0.274)

avg_age               -0.094**    -0.094*     -0.094***
                      (0.036)     (0.043)      (0.016)

avg_income             0.084       0.084       0.084***
                      (0.060)     (0.071)      (0.020)

BIE_timeline1:group_store -0.429  -0.429      -0.429***
                      (0.406)     (0.425)      (0.047)

Constant             -11.822***  -11.822***  -11.822***
                      (0.627)     (0.690)      (0.371)

-----------------------------------------------------------------
Observations           1,381       1,381        1,381
R2                     0.591       0.591        0.591
Adjusted R2            0.586       0.586        0.586
Residual Std. Error    2.589       2.589        2.589
F Statistic          109.308***  109.308***   109.308***
=================================================================
Note:                        *p<0.05; **p<0.01; ***p<0.001
```

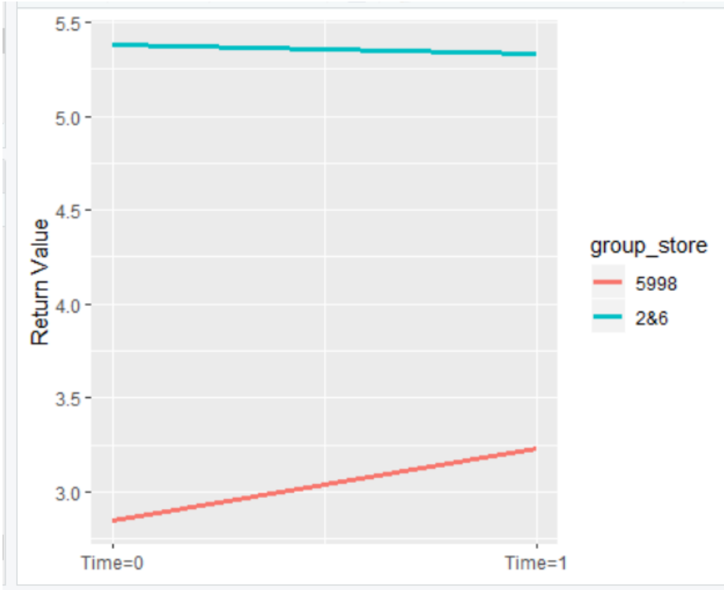
meffects_pc1 <- ggpredict(ols_pc1, terms=c("BIE_timeline", "group_store")) # generates a tidy data frame

BA Project_BOPS_Final Submitted.Rmd × | meffects_pc1 × | df_pc1 ×

Filter

	x	predicted	conf.low	conf.high	group
1	0	2.848693	2.041666	3.655719	0
2	0	5.383446	4.848068	5.918823	1
3	1	3.226764	2.575525	3.878003	0
4	1	5.332706	4.803626	5.861786	1

```
ggplot(meffects_pc1,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Return Value") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
  theme(axis.title.x=element_blank())
```



##Interaction coefficients are insignificant for following categories- 5,6,7.
##Data available only for one store_group:8,10,15,17.

Interaction coefficient is significant for categories - 1,2,3,4,9,11,12,13,14,20,21.
Therefore, BOPS implementation for these categories is associated with:
Product Category 1: 42.9% decrease in return value
Product Category 3: 93.9% decrease in return value
Product Category 4: 95.0% decrease in return value
Product Category 9: 49.3% decrease in return value
Product Category 12: 34.6% decrease in return value
Product Category 13: 81.3% decrease in return value
Product Category 14: 41.4% decrease in return value
Product Category 20: 92.7% decrease in return value
Product Category 21: 54.0% decrease in return value

Product Category 2: 19.1% increase in return value
Product Category 11: 37.4% increase in return value

In order to further access the impact of BOPS implementation on return value, we divided these in three broad groups:
No Impact: 5,6,7
Low Impact(Decrease<50%): 1,9,12,14,
High Impact (Decrease>50%): cat. 3,4,13,20,21.

Also, we see an increase in return value for categories 2 and 11.

####===================IMPACT ON SALES VALUE=====================####

```
ols62 = lm(log_salesvalue~BIE_timeline*group_store*product_category+ month_dummy+
avg_female+avg_age+avg_income + avg_homeowner + avg_childowner, data = sales6)

stargazer(ols62,
    title="Regression Results", type="text",
    column.labels=c("Model-sales value "),
    df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))

# Check for heteroskedasticity
gqtest(ols62) # Goldfeld-Quandt test indicates heteroskedasticity
bptest(ols62) # Breusch-Pagan test indicates heteroskedasticity
```

```
        Goldfeld-Quandt test

data:  ols62
GQ = 0.44219, df1 = 10414, df2 = 10413, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols62) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols62
BP = 1104.4, df = 79, p-value < 2.2e-16
```

```
# Since there is heteroskedasticity in the data, we will replace SEs with clustered robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols62, type="HC1"))) # produces Huber-White robust standard errors
clusrobstder <- sqrt(diag(cluster.vcov(ols62, sales6$store_number ))) # produces clustered robust
standard errors using store_number as clustering variable
```

Values	
clusrobstder	Named num [1:80] 0.05942 0.01047 0.9756...
HWrobstder	Named num [1:80] 0.141 0.119 0.123 0.13...

```
stargazer(ols62, ols62, ols62,
    se=list(NULL, HWrobstder, clusrobstder),
    title="Regression Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE", "Clustered SE"),
    df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001)) # displays normal/HW robust/clustered robust
standard errors.
```

Regression Results

	Dependent variable: log_salesvalue		
	Normal SE (1)	HW-Robust SE (2)	Clustered SE (3)
BIE_timeline1	0.493*	0.493***	0.493***
	(0.230)	(0.119)	(0.010)
group_store	1.533***	1.533***	1.533
	(0.203)	(0.123)	(0.976)
product_category2	-1.268***	-1.268***	-1.268***
	(0.261)	(0.135)	(0.003)
product_category3	0.037	0.037	0.037***
	(0.256)	(0.136)	(0.006)
product_category4	-0.084	-0.084	-0.084***
	(0.227)	(0.126)	(0.008)
product_category5	-0.421	-0.421***	-0.421***
	(0.224)	(0.121)	(0.008)
product_category6	-1.076***	-1.076***	-1.076***
	(0.250)	(0.129)	(0.004)
product_category7	-2.098***	-2.098***	-2.098***
	(0.230)	(0.126)	(0.011)
product_category8	-1.650***	-1.650***	-1.650
	(0.126)	(0.106)	(0.915)
product_category9	-2.800***	-2.800***	-2.800***
	(0.517)	(0.368)	(0.015)
product_category10	-2.814***	-2.814***	-2.814**
	(0.172)	(0.146)	(0.946)
product_category11	-0.400	-0.400**	-0.400***
	(0.237)	(0.135)	(0.014)
product_category12	-0.573*	-0.573***	-0.573***
	(0.225)	(0.127)	(0.011)
product_category13	-0.550*	-0.550***	-0.550***
	(0.241)	(0.131)	(0.004)
product_category14	-2.772***	-2.772***	-2.772***
	(0.279)	(0.130)	(0.004)
product_category15	-3.708***	-3.708***	-3.708***
	(0.355)	(0.361)	(0.975)
product_category17	0.492***	0.492***	0.492
	(0.124)	(0.093)	(0.973)
product_category20	-0.636**	-0.636***	-0.636***
	(0.241)	(0.125)	(0.008)
product_category21	-0.894***	-0.894***	-0.894***
	(0.226)	(0.121)	(0.007)
month_dummy2	0.413***	0.413***	0.413***
	(0.055)	(0.057)	(0.113)
month_dummy3	0.034	0.034	0.034
	(0.055)	(0.059)	(0.062)
month_dummy4	-0.001	-0.001	-0.001
	(0.056)	(0.058)	(0.109)
month_dummy5	0.216***	0.216***	0.216***
	(0.054)	(0.059)	(0.060)
month_dummy6	0.005	0.005	0.005
	(0.056)	(0.057)	(0.143)
month_dummy7	-0.005	-0.005	-0.005
	(0.056)	(0.057)	(0.050)
month_dummy8	0.012	0.012	0.012
	(0.051)	(0.054)	(0.107)
month_dummy9	0.056	0.056	0.056
	(0.053)	(0.055)	(0.049)
month_dummy10	0.004	0.004	0.004
	(0.057)	(0.060)	(0.059)
month_dummy11	0.410***	0.410***	0.410***
	(0.054)	(0.057)	(0.057)
month_dummy12	1.014***	1.014***	1.014***
	(0.052)	(0.055)	(0.090)
avg_female	-0.394***	-0.394***	-0.394***
	(0.048)	(0.064)	(0.087)
avg_age	-0.064***	-0.064***	-0.064***
	(0.006)	(0.008)	(0.008)
avg_income	0.104***	0.104***	0.104***
	(0.011)	(0.015)	(0.008)
avg_homeowner	-0.194***	-0.194*	-0.194***
	(0.055)	(0.077)	(0.042)
avg_childowner	0.062	0.062	0.062
	(0.051)	(0.071)	(0.114)
BIE_timeline1:group_store	-0.377	-0.377*	-0.377***
	(0.248)	(0.155)	(0.019)
BIE_timeline1:product_category2	-0.212	-0.212	-0.212***
	(0.316)	(0.171)	(0.005)
BIE_timeline1:product_category3	-0.407	-0.407*	-0.407***
	(0.308)	(0.167)	(0.004)
BIE_timeline1:product_category4	0.089	0.089	0.089***
	(0.276)	(0.157)	(0.003)
BIE_timeline1:product_category5	-0.387	-0.387*	-0.387***
	(0.275)	(0.152)	(0.005)
BIE_timeline1:product_category6	-0.316	-0.316*	-0.316***
	(0.309)	(0.156)	(0.002)
BIE_timeline1:product_category7	-0.403	-0.403**	-0.403***
	(0.284)	(0.156)	(0.007)
BIE_timeline1:product_category8	-1.349***	-1.349***	-1.349***
	(0.220)	(0.194)	(0.109)
BIE_timeline1:product_category9	0.821	0.821*	0.821***
	(0.552)	(0.388)	(0.013)
BIE_timeline1:product_category10	-0.016	-0.016	-0.016
	(0.380)	(0.298)	(0.008)
BIE_timeline1:product_category11	-0.430	-0.430**	-0.430***
	(0.289)	(0.163)	(0.011)
BIE_timeline1:product_category12	-0.318	-0.318*	-0.318***
	(0.275)	(0.158)	(0.008)
BIE_timeline1:product_category13	-0.831**	-0.831***	-0.831***
	(0.297)	(0.163)	(0.002)
BIE_timeline1:product_category14	0.067	0.067	0.067***
	(0.367)	(0.183)	(0.014)
BIE_timeline1:product_category15	1.304	1.304**	1.304***
	(0.874)	(0.478)	(0.001)
BIE_timeline1:product_category17	-0.397*	-0.397**	-0.397***
	(0.172)	(0.135)	(0.028)
BIE_timeline1:product_category20	-0.214	-0.214	-0.214***
	(0.295)	(0.151)	(0.006)
BIE_timeline1:product_category21	-0.055	-0.055	-0.055***
	(0.277)	(0.151)	(0.001)
group_store:product_category2	0.286	0.286	0.286***
	(0.278)	(0.166)	(0.021)
group_store:product_category3	-0.204	-0.204	-0.204***
	(0.274)	(0.168)	(0.027)
group_store:product_category4	0.320	0.320*	0.320***
	(0.245)	(0.159)	(0.036)
group_store:product_category5	0.302	0.302	0.302**
	(0.243)	(0.157)	(0.110)
group_store:product_category6	-0.017	-0.017	-0.017
	(0.268)	(0.163)	(0.047)
group_store:product_category7	-0.896***	-0.896***	-0.896
	(0.255)	(0.166)	(0.557)
group_store:product_category8			
group_store:product_category9	-1.501**	-1.501***	-1.501
	(0.541)	(0.393)	(0.872)
group_store:product_category10			
group_store:product_category11	-0.383	-0.383*	-0.383***
	(0.257)	(0.170)	(0.056)
group_store:product_category12	0.520*	0.520**	0.520**
	(0.244)	(0.164)	(0.197)
group_store:product_category13	-0.554*	-0.554***	-0.554***
	(0.261)	(0.161)	(0.139)
group_store:product_category14	0.651*	0.651***	0.651
	(0.297)	(0.182)	(0.584)
group_store:product_category15			
group_store:product_category17			
group_store:product_category20	-0.070	-0.070	-0.070
	(0.261)	(0.157)	(0.081)
group_store:product_category21	0.403	0.403**	0.403***
	(0.245)	(0.156)	(0.033)

```
BIE_timeline1:group_store:product_category2     -0.083      -0.083      -0.083
                                                (0.342)     (0.220)     (0.044)

BIE_timeline1:group_store:product_category3      0.296       0.296       0.296***
                                                (0.337)     (0.218)     (0.003)

BIE_timeline1:group_store:product_category4     -0.183      -0.183      -0.183
                                                (0.304)     (0.210)     (0.121)

BIE_timeline1:group_store:product_category5      0.201       0.201       0.201***
                                                (0.303)     (0.206)     (0.007)

BIE_timeline1:group_store:product_category6      0.343       0.343       0.343***
                                                (0.335)     (0.207)     (0.041)

BIE_timeline1:group_store:product_category7      0.220       0.220       0.220
                                                (0.320)     (0.220)     (0.206)

BIE_timeline1:group_store:product_category8

BIE_timeline1:group_store:product_category9      1.032       1.032*      1.032***
                                                (0.586)     (0.437)     (0.205)

BIE_timeline1:group_store:product_category10

BIE_timeline1:group_store:product_category11     0.238       0.238       0.238***
                                                (0.319)     (0.219)     (0.030)

BIE_timeline1:group_store:product_category12     0.153       0.153       0.153***
                                                (0.303)     (0.213)     (0.022)

BIE_timeline1:group_store:product_category13     0.509       0.509*      0.509***
                                                (0.326)     (0.211)     (0.041)

BIE_timeline1:group_store:product_category14    -0.790*     -0.790**     -0.790
                                                (0.393)     (0.245)     (0.782)

BIE_timeline1:group_store:product_category15

BIE_timeline1:group_store:product_category17

BIE_timeline1:group_store:product_category20     0.286       0.286       0.286**
                                                (0.325)     (0.204)     (0.104)

BIE_timeline1:group_store:product_category21     0.062       0.062       0.062
                                                (0.304)     (0.205)     (0.088)

Constant                                         6.970***    6.970***    6.970***
                                                (0.205)     (0.141)     (0.059)

--------------------------------------------------------------------------------
Observations                                     21,003      21,003      21,003
R2                                               0.338       0.338       0.338
Adjusted R2                                      0.335       0.335       0.335
Residual Std. Error                              1.592       1.592       1.592
F Statistic                                    135.036***  135.036***  135.036***
================================================================================
Note:                                           *p<0.05; **p<0.01; ***p<0.001
>
```

Since we cannot interpret the results of a triple interaction model, we will analyse the imapact of BOPS implementation on individual categories by subsetting the data further.This will help us understand how sales value is changing for each category w.r.t store groups(group 0 = 5998 and group 1 = 2&6).

###-----------Marginal Effects ---------------------###

meffects62 <- ggpredict(ols62, terms=c("BIE_timeline", "group_store","product_category"))

BA Project_BOPS_Final Submitted.Rmd × | meffects62 ×

Filter

	x	predicted	conf.low	conf.high	group	facet
1	0	6.899337	6.518658	7.280016	0	1
2	0	5.631111	5.273560	5.988662	0	2
3	0	6.936172	6.593506	7.278839	0	3
4	0	6.815450	6.563122	7.067778	0	4
5	0	6.478612	6.234665	6.722559	0	5

Showing 1 to 6 of 72 entries

# Replacing by actual variable names for convenience
meffects62_colnames <- meffects62[c("x","predicted","group", "facet")]

BA Project_BOPS_Final Submitted.Rmd × | meffects62_colnames × | meffects62 ×

Filter

	x	predicted	group	facet
1	0	6.899337	0	1
2	0	5.631111	0	2
3	0	6.936172	0	3
4	0	6.815450	0	4
5	0	6.478612	0	5

Showing 1 to 6 of 72 entries

colnames(meffects62_colnames) <- c("BIE_timeline", "predicted_sales_value","group_store" , "product_category")

BA Project_BOPS_Final Submitted.Rmd × | meffects62_colnames × | meffects62 ×

Filter

	BIE_timeline	predicted_sales_value	group_store	product_category
1	0	6.899337	0	1
2	0	5.631111	0	2
3	0	6.936172	0	3
4	0	6.815450	0	4
5	0	6.478612	0	5

Showing 1 to 6 of 72 entries

levels(meffects62_colnames$product_category) <- c('1-Bridal', '2-Gold Wed Bands', '3-Solitaires', '4-Diamond Fashion', '5-Semi Precious', '6-Mens', '7-Gold Earrings', '8-In House Special Event', '9-Beads','10-Piercings / Close Out', '11-Diamond Solitaires Jewelry', '12-Gold Chain / Jewelry ', '13-Watches', '14-Pre-Owned', '15-Specialized Jewelry', '17-Events', '19-Repair / Warranty','20-Diamond Wedding Bands','21-Sterling Silver')

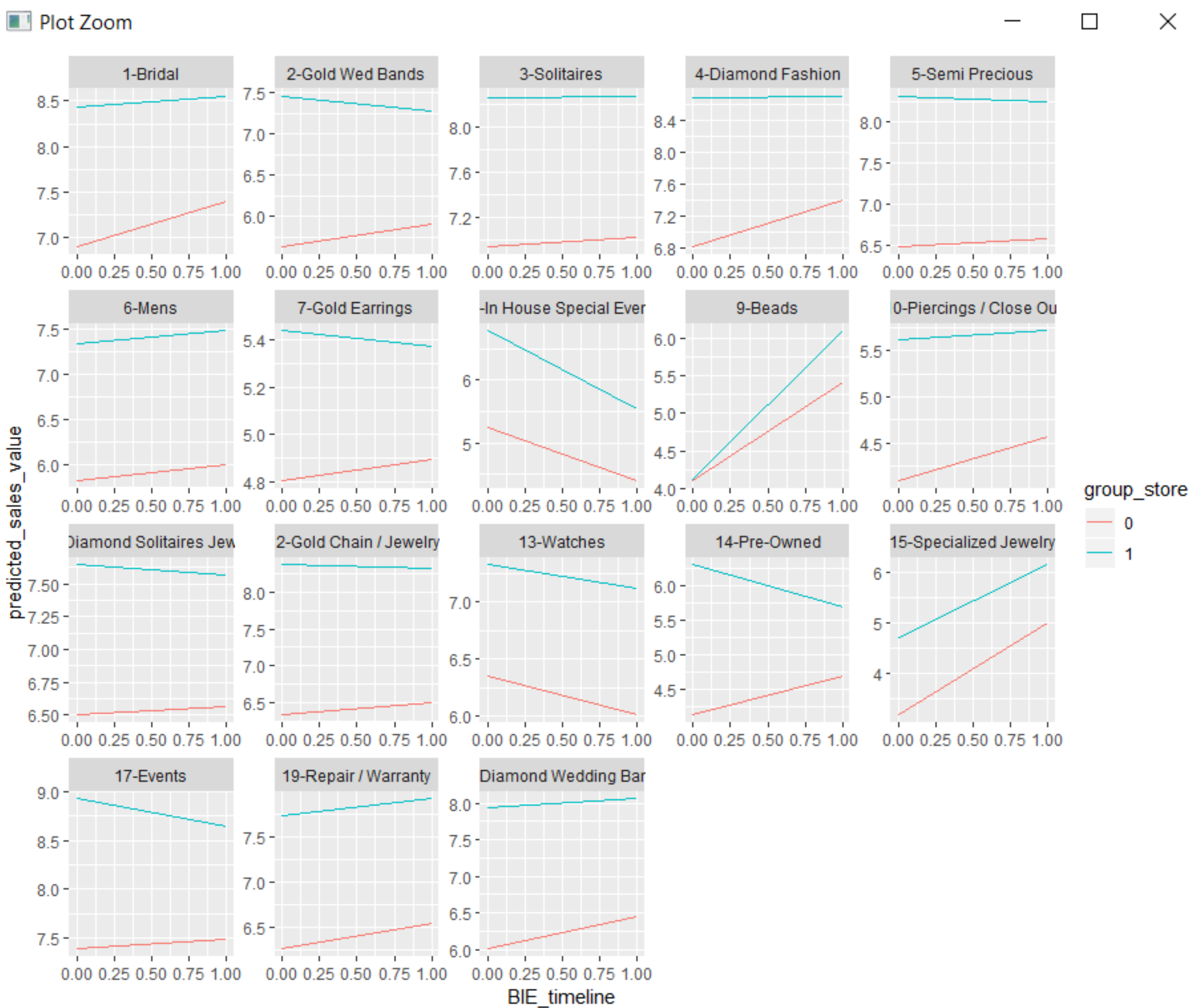
BA Project_BOPS_Final Submitted.Rmd × | meffects62_colnames × | meffects62 ×

Filter

	BIE_timeline	predicted_sales_value	group_store	product_category
1	0	6.899337	0	1-Bridal
2	0	5.631111	0	2-Gold Wed Bands
3	0	6.936172	0	3-Solitaires
4	0	6.815450	0	4-Diamond Fashion
5	0	6.478612	0	5-Semi Precious

Showing 1 to 6 of 72 entries

```
ggplot(meffects62_colnames, aes(BIE_timeline,
 y= predicted_sales_value, colour =group_store)) +
 geom_line() +
 facet_wrap(product_category ~ ., scales="free")
```



Checking predicted values for all the product categories (from 1 to 21) and analyzing impact of BOPS implementation for both storegroups(group 0 = 5998 and group 1 = 2 and 6).

df2_pc1<- subset(sales6, product_category==1)

BA Project_BOPS_Final Submitted.Rmd | df2_pc1 | meffects62_colnames | meffects62

	store_number	year	month	month_index	product_category	month_dummy	bops_in_effect	day	salesvalue	returnvalue	salesquantity	returnquantity	avg_fem
13	2	2010	AUG	13	1	8	0	2	12253.53	1478.82	13	2	
33	2	2010	AUG	13	1	8	0	3	7989.74	1287.40	22	3	
45	2	2010	AUG	13	1	8	0	4	26309.46	10614.14	26	5	
61	2	2010	AUG	13	1	8	0	5	6399.69	771.66	12	1	

Showing 1 to 6 of 1,381 entries

BA Project_BOPS_Final Submitted.Rmd | df2_pc1 | meffects62_colnames | meffects62

ntity	returnquantity	avg_female	avg_age	avg_income	avg_homeowner	avg_residency	avg_childowner	BIE_timeline	group_store	log_salesvalue	log_returnvalue
13	2	0.4545455	4.615385	5.769231	0.3076923	6.846154	0.07692308	0	1	9.413651	7.299676
22	3	0.5555556	3.727273	4.590909	0.4090909	4.863636	0.27272728	0	1	8.986039	7.161156
26	5	0.4400000	5.461538	5.115385	0.5769231	6.000000	0.42307693	0	1	10.177722	9.270037
12	1	0.5000000	7.250000	3.750000	0.5833333	7.333333	0.25000000	0	1	8.764161	6.649839

Showing 1 to 6 of 1,381 entries

ols2_pc1= lm(log_salesvalue~BIE_timeline*group_store+ month_dummy+ avg_female+avg_age+avg_income+ avg_homeowner + avg_childowner, data = df2_pc1)

```
stargazer(ols2_pc1,
    title="Regression Results", type="text",
    column.labels=c("ME-pc1"),
    df=FALSE, digits=4, star.cutoffs = c(0.05,0.01,0.001))
```

Check for heteroskedasticity
gqtest(ols2_pc1) # Goldfeld-Quandt test indicates no heteroskedasticity
bptest(ols2_pc1) # Breusch-Pagan test indicates heteroskedasticity

```
> gqtest(ols2_pc1) # Goldfeld-Quandt test indicates no heteroskedasticity

        Goldfeld-Quandt test

data:  ols2_pc1
GQ = 0.44456, df1 = 671, df2 = 670, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(ols2_pc1) # Breusch-Pagan test indicates heteroskedasticity

        studentized Breusch-Pagan test

data:  ols2_pc1
BP = 95.957, df = 19, p-value = 0.00000000000287
```

# Since there is heteroskedasticity in the data, we will replace SEs with robust SEs.
HWrobstder <- sqrt(diag(vcovHC(ols2_pc1, type="HC1"))) # produces Huber-White robust standard errors
clusrobstder <- sqrt(diag(cluster.vcov(ols2_pc1, sales6$store_number))) # produces clustered robust standard errors

Values	
clusrobstder	Named num [1:80] 0.05942 0.01047 0.97567 0.00339 0.…
HWrobstder	Named num [1:20] 0.343 0.118 0.127 0.205 0.242 ...

```
stargazer(ols2_pc1, ols2_pc1, ols2_pc1,
    se=list(NULL, HWrobstder, clusrobstder),
    title=" Regression Results", type="text",
    column.labels=c("Normal SE", "HW-Robust SE", "Clustered SE"),
    df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001)) # displays normal/HW robust/clustered robust standard errors. With clustered robust standard errors
```

Regression Results

	Dependent variable:		
	log_salesvalue		
	Normal SE	HW-Robust SE	Clustered SE
	(1)	(2)	(3)
BIE_timeline1	0.434	0.434***	0.434***
	(0.230)	(0.118)	(0.010)
group_store	1.501***	1.501***	1.501
	(0.205)	(0.127)	(0.976)
month_dummy2	0.405	0.405*	0.405***
	(0.217)	(0.205)	(0.113)
month_dummy3	0.025	0.025	0.025
	(0.220)	(0.242)	(0.062)
month_dummy4	0.050	0.050	0.050
	(0.223)	(0.212)	(0.109)
month_dummy5	0.008	0.008	0.008
	(0.219)	(0.236)	(0.060)
month_dummy6	-0.117	-0.117	-0.117
	(0.221)	(0.219)	(0.143)
month_dummy7	-0.214	-0.214	-0.214***
	(0.218)	(0.223)	(0.050)
month_dummy8	-0.130	-0.130	-0.130
	(0.205)	(0.209)	(0.107)
month_dummy9	0.100	0.100	0.100*
	(0.213)	(0.206)	(0.049)
month_dummy10	-0.011	-0.011	-0.011
	(0.227)	(0.217)	(0.059)
month_dummy11	0.121	0.121	0.121*
	(0.215)	(0.209)	(0.057)
month_dummy12	0.552**	0.552**	0.552***
	(0.208)	(0.200)	(0.090)
avg_female	-0.086	-0.086	-0.086
	(0.172)	(0.211)	(0.087)
avg_age	-0.054*	-0.054	-0.054***
	(0.023)	(0.028)	(0.008)
avg_income	0.108**	0.108*	0.108***
	(0.039)	(0.053)	(0.008)
avg_homeowner	-0.222	-0.222	-0.222***
	(0.188)	(0.259)	(0.042)
avg_childowner	0.075	0.075	0.075
	(0.181)	(0.256)	(0.114)
BIE_timeline1:group_store	-0.313	-0.313*	-0.313***
	(0.248)	(0.154)	(0.019)
Constant	6.960***	6.960***	6.960***
	(0.334)	(0.343)	(0.059)
Observations	1,381	1,381	1,381
R2	0.104	0.104	0.104

Observations	1,381	1,381	1,381
R2	0.104	0.104	0.104
Adjusted R2	0.091	0.091	0.091
Residual Std. Error	1.581	1.581	1.581
F Statistic	8.289***	8.289***	8.289***
Note:	*p<0.05; **p<0.01; ***p<0.001		

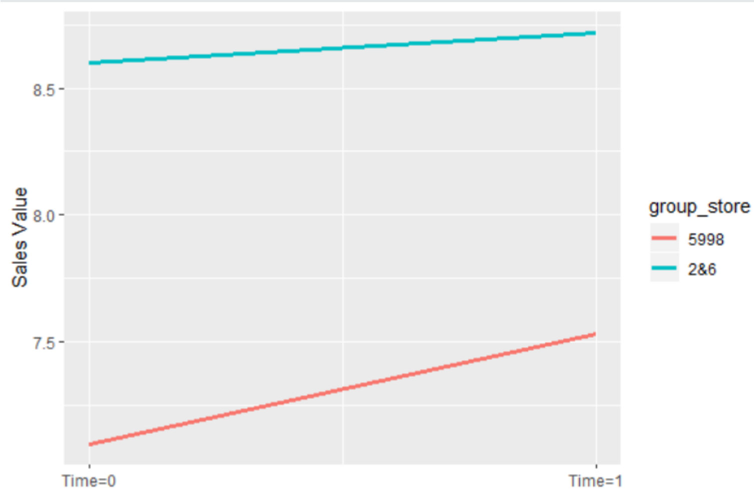
meffects2_pc1 <- ggpredict(ols2_pc1, terms=c("BIE_timeline", "group_store")) # generates a tidy data frame

BA Project_BOPS_Final Submitted.Rmd | meffects2_pc1 | df2_pc1

Filter

	x	predicted	conf.low	conf.high	group
1	0	7.097919	6.610979	7.584860	0
2	0	8.598620	8.270955	8.926285	1
3	1	7.532005	7.137703	7.926307	0
4	1	8.719334	8.396289	9.042378	1

```
ggplot(meffects2_pc1,aes(x, predicted, colour=group)) + geom_line(size=1.3) +
  xlab("BOPS Timeline") + ylab("Sales Value") +
  labs(colour="group_store") +
  scale_colour_discrete(labels=c("5998","2&6")) +
  scale_x_continuous(breaks=c(0,1), labels=c("Time=0", "Time=1")) +
  theme(axis.title.x=element_blank())
```



Interaction coefficient is insignificant for categories - 6,7
However, there was no coefficient displayed for categories - 8,10,15,17, because data is available for only one store group. So, we cannot analyse these categories.

Interaction coefficient is significant for categories - 1,2,3,4,5,9,11,12,13,14,20,21
Therefore, BOPS implementation for these categories is aasociated with:
Product Category 1: 31.3% decrease in sales value
Product Category 2: 38.8% decrease in sales value
Product Category 3: 6.4% decrease in sales value
Product Category 4: 62% decrease in sales value
Product Category 5: 27.2% decrease in sales value
Product Category 11: 16.9% decrease in sales value
Product Category 12: 31.7% decrease in sales value
Product Category 14: 114% decrease in sales value
Product Category 20: 5.7% decrease in sales value
Product Category 21: 39.3% decrease in sales value

Product Category 9: 35% increase in sales value
Product Category 13: 12.2% increase in sales value

In order to further access the impact of BOPS implementation on sales value, we divided these in three broad groups:
No Impact(insignificant): cat. 6,7
Low Impact(Decrease<30%): cat. 3,5, 11,20
High Impact (Decrease>50%): cat. 1,2,12,21,4,14

Also, we see an increase in sales value for categories 9 and 13.

```
```