Assignment 4

Group 9

Ankit Bassi

Navya Tanuja Gollapalli

Sagar Jadhav

Piyush Kulkarni

Arti Nanote

Divya Virender Kumar

MIS 6344.002 Web Analytics- S17

The University of Texas at Dallas – Naveen Jindal School of Management

PART A

First we created a group variable which is equal to 1 (irrespective of video switch on and switch off peroid) when sku is in treatement group and 0 when it is in control group. We created 296 dummies for 297 skus.

We created a promotion variable whose value is 1 when there is any kind of promotion going on for that sku in that week. Further we interacted this promotion variable with video week to get the effect of any kind of promotion with the effect of video at the same time (var13).

We considered following variables in our regression model:

Sales as independent variable, 5 promotion varaibles, 296 product dummies, vidwk and price as dependent/predictors. The regression equation is as follows:

Y= 34.48 + 49.40*var13 + 111.8*catalog + 36.44*deptpg + 67.22*email + 3.66*group + 71.22*homepg + 12.8*price + 21.56*vidwk + sku coefficients.

So the model tells us that video is increasing the sales by 21.56 with p-value 0.0001. Resulting in significant output.

The model also tells us that the sales significantly increases when there is a pramotion going on and the product also has a video. (sales increases by 49 and p-val of 0.0008)

Further, to analyse the impact of the video in switch off and switch on period, we ran a separate regression on the treatment group.

We can say that within the treatment group also that the sales is increased by video. Vid week has a significant p-value.

Regression Output for compelete data set (Not showing all the sku pages)

Model Fit Statistics

R-Square	0.4459	Adj R-Sq	0.4202
AIC	62507.9258	BIC	62538.2526
SBC	64583.8771	C(p)	304.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	34.4848	17.9954	1.92	0.0554
VAR13	1	49.4080	14.7788	3.34	0.0008
cat	1	111.8	18.9533	5.90	<.0001
deptpg	1	36.4492	11.5704	3.15	0.0016
email	1	67.2230	24.5048	2.74	0.0061
group	1	3.6623	29.3837	0.12	0.9008
homepg	1	71.2288	10.4509	6.82	<.0001
price	1	12.8082	6.0223	2.13	0.0335
sku2143	1	160.8	25.4411	6.32	<.0001
sku2413	1	36.9795	25.4411	1.45	0.1461
sku2461	1	15.3343	25.4369	0.60	0.5466
sku2532	1	64.1818	28.7563	2.23	0.0257
sku2637	1	15.8009	25.4411	0.62	0.5346
sku2663	1	55.5152	25.4411	2.18	0.0291
sku2700	1	22.6223	25.4411	0.89	0.3739
sku2716	1	66.8009	25.4411	2.63	0.0087
sku2800	1	105.8	25.4411	4.16	<.0001
sku3133	1	28.8310	28.2922	1.02	0.3082
sku3164	1	88.0152	25.4411	3.46	0.0005
sku3251	1	48.9231	26.7798	1.83	0.0678
sku3253	1	90.9932	25.4338	3.58	0.0003
sku3261	1	263.2	27.8653	9.44	<.0001
sku3269	1	76.0652	27.8678	2.73	0.0064
sku3292	1	-3.7705	25.4411	-0.15	0.8822
sku3311	1	89.9437	25.4411	3.54	0.0004
sku3329	1	36.6223	25.4411	1.44	0.1501
sku3339	1	88.4795	25.4411	3.48	0.0005
sku3364	1	55.3366	25.4411	2.18	0.0297
sku3381	1	40.6767	29.3264	1.39	0.1655
sku3383	1	67.7374	25.6754	2.64	0.0084
sku3408	1	31.1417	25.4525	1.22	0.2212
sku3409	1	52.5542	25.4540	2.06	0.0390
sku3413	1	-34.4848	69.6539	-0.50	0.6206
sku3416	1	183.8	25.4411	7.22	<.0001
sku3418	1	160.1	25.4411	6.29	<.0001
sku3435	1	26.1580	25.4411	1.03	0.3039
sku3479	1	25.2170	27.1723	0.93	0.3534
sku3511	1	52.4918	25.4357	2.06	0.0391
sku3561	1	29.5000	25.4329	1.16	0.2461
sku3618	1	-2.6991	25.4411	-0.11	0.9155
sku3619	1	8.9527	29.8297	0.30	0.7641

```
sku4248       1      17.2652      32.8401       0.53      0.5991
sku4249       1      -0.6723      29.8297      -0.02      0.9820
sku4250       1      22.2652      29.8297       0.75      0.4554
sku4251       1       215.0       31.4637       6.83     <.0001
sku4263       1      -7.8420      31.1555      -0.25      0.8013
vidwk         1      21.5623       5.6561       3.81      0.0001
```

Regression Output for only treatment group

	Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	116.08002	13.32003	8.71	<.0001
vidwk	vidwk	1	16.88779	7.40963	2.28	0.0228
price	price	1	9.51484	17.89345	0.53	0.5950
email	email	1	25.10626	62.46238	0.40	0.6878
cat	cat	1	-13.93109	34.58442	-0.40	0.6871
homepg	homepg	1	47.20674	25.76080	1.83	0.0671
deptpg	deptpg	1	10.62283	22.94692	0.46	0.6435
Interaction_vidwk_promo_	Interaction(vidwk*promo)	1	72.48217	23.15970	3.13	0.0018
sku2143	sku2143	0	0	.	.	.
sku2413	sku2413	0	0	.	.	.
sku2461	sku2461	0	0	.	.	.
sku2532	sku2532	0	0	.	.	.
sku2637	sku2637	0	0	.	.	.
sku2663	sku2663	0	0	.	.	.
sku2700	sku2700	0	0	.	.	.
sku2716	sku2716	0	0	.	.	.
sku2800	sku2800	0	0	.	.	.

Ideal Solution:

Final Regression Model

$$Sales_{it} = \alpha_i + \alpha_t + \delta \times Vid_{it} + \boldsymbol{\beta} \times \boldsymbol{X_{it}} + \varepsilon_{it}$$

Where, $i\epsilon\{1, 2, 3, \ldots\ldots 297\}$ denotes the 297 principal products $\quad t\epsilon\{1, 2, 3, \ldots\ldots.. 28\}$ denotes the 28 weeks of the study period

$\boldsymbol{X_{it}}$ denotes the control variables

Vid_{it} is 1 if video is on for product i in week t, and 0 o.w.

α_i and α_t are products and week fixed effects

There are two sets of fixed effects. In order to control for these we could use either dummy variables, or the fixed effects approach

In general, using fixed effects approach is a better than dummy variables because the number of variables in the model is smaller when we use FE

Fewer variables are good as standard errors of the coefficients are lower as (n-k-1) higher, n= no. of observations and k= number of variables in the model. Hence, we are in a better position to judge whether the treatment has a significant impact

Results

Dependent Variable (Weekly sales in numbers)	Coefficient Estimates
	Full dataset
Product video	16.70** (5.08)
Catalogue	103.80** (17.01)
Website home page	61.08*** (9.41)
Category front page	29.44 (10.15)
Price markdown	76.62*** (5.82)
Email promotion	68.91* (22.04)
No. of product-weeks (No. of products)	6828 (297)

Product video leads to 15% increase in focal product sales

Part B

By considering a pre-video period and post video switch off period in addition to the switch on period for our test group we can analyze the impact on sales by video. This would allow us to see if sales actually increase during the switch on period. If the video increases sales then we should see low sales in pre video period and post video switch off period and more sales in video switch on period. There could be other factors that affect the test group and so, to make sure that these don't influence our results,

we would ideally see an increase in sales once the video is shown and a decrease during the switch off period.

Additional data that can improve the regression model:

1. Promotions: There could be a case where the products are promoted more during treatment period than control period. Because of this, sales are more for treatment group than control group. So, there is a need to maintain the degree to which the promotions are done in these two groups. We can implement this by introducing an interactive variable that will nullify this effect.
2. Price: If the price associated with the treatment group products are more than the control group. In this case the sales would be more for treatment group than control group. We can improve the model performance by controlling the price associated with the products.
3. Catalog position/the web page position: the position of product could also impact the sales as in weather the product was above the fold or below the fold.
4. Additional information about the effect of a focal product on a co-ordination product or vice versa can improve the model