Media Mix Modeling

Marketing 2505: Marketing Analytics
Winter 2020

Using data from the class, try to build a Media Mix model.

Start with the base model

$$Sales_t = \beta_0 + \beta_1 Newspaper_t + \beta_2 Magazine_t + \beta_3 Radio_t + \beta_4 TV_t + \epsilon_t$$

MODEL INFO:Q1
Observations: 48
Dependent Variable: Yval
Type: OLS linear regression

MODEL FIT:
$F(4,43) = 4.70419$, $p = 0.00308$
$R^2 = 0.30440$
Adj. $R^2 = 0.23969$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	2043.83856	43.96858	46.48407	0.00000
Xvarsnewspaper	0.07971	0.18969	0.42023	0.67641
Xvarsmagazine	-0.00192	0.11127	-0.01726	0.98631
Xvarsradio	0.28097	0.06581	4.26921	**0.00011**
Xvarstv	0.08001	0.08524	0.93864	0.35316

Please do not just copy the R output, but put the results in tables or a nice format, ready to present to your managers. Think about what are the important statistics you want to show for a regression model.

1. Add lagged X variable

Besides having all X variables in the current time period, also add the same X variables about media spending in the past time period. In this model, the number of parameters is almost doubled that in the base model. Compare the results from this model with those from the base model. Which one do you like better? Why?

Regression Results

Parameter Estimates	With(Lag)	Without(Lag)
(Intercept)	1990.02515 ***	2043.83856 ***
New_Newspaper	0.32707	
New_Magazine	-0.09842	
New_Radio	0.29759 ***	
New_TV	0.15063	
XvarsNewspaper	0.23321	0.07971
XvarsMagazine	-0.17025	-0.00192
XvarsRadio	0.07809	0.28097 ***
XvarsTV	-0.02035	0.08001
N	47	48
R2	0.43968	0.30440
AdjR2	0.3217	0.23969

*** p < 0.001; ** p < 0.01; * p < 0.05.

By taking a long-term perspective in mind , I feel that we should consider the model with lagged X variables to know the effect of media channels on sales .

The R square and adj R square is greater for the model with lagged variable as compared to the base model **0.3217 > 0.2396 ,** which shows that **32.17%** of the variables follow the model . Both Radio and lagged Radio are significant in the base and the model with lagged variables respectively **.**

2. Stock Variable

Create stock variable with different λ values: $\lambda = 0.1, 0.5, 0.9,$ compare the results from these models with those from the base model by putting them in one big table. Comment on the differences across the models.

Regression Analysis

	Ad Stock (0.1)	Ad Stock (0.5)	Ad Stock (0.9)
(Intercept)	2033.736	2009.986	2084.89155 ***
GGnewspaper	0.09190	0.09313	-0.01655
GGmagazine	-0.01505	-0.05414	-0.04249
GGradio	**0.27231**	**0.18173**	**0.03248** *
GGtv	0.08343	0.06471	0.01605
N	48	48	48
R2	**0.30722**	**0.27479**	**0.20048**
Adj R2	**0.2428**	**0.2073**	**0.1261**

*** p < 0.001; ** p < 0.01; * p < 0.05.

By including **λ (Lambda)** , we are accounting for the effect of investment in the channel in previous time period and its lingering effect on current sales .
We can see in the above table as the **λ increases** the effect of all the medium on the sales decreases , Moreover ,we can see that **newspaper** channel is having more **negative** effect as the **λ increases .** This shows that if we keep investing on channels it will reach the saturation point and start having negative impact on Sales .
Adjusted R Square is greatest for **λ (0.1)**
0.2428 (AdStock (0.1)) > 0.2073 (AdStock (0.5))> 0.1261(AdStock (0.9))

3. Dummy variables

(1) Create one dummy variable for the three months in the summer, from June to August.

MODEL INFO:Q.3(A)
Observations: 48
Dependent Variable: Yval
Type: OLS linear regression

MODEL FIT:
$F(5,42) = 3.80241$, $p = 0.00625$
$R^2 = 0.31161$
Adj. $R^2 = 0.22966$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	2083.99116	74.97247	27.79675	0.00000
Newspaper	0.07442	0.19110	0.38942	0.69893
Magazine	-0.00979	0.11262	-0.08690	0.93116
Radio	0.09048	0.08724	1.03716	0.30559
TV	0.28269	0.06630	4.26399	0.00011
Dummy_Summer	-33.28427	50.16407	-0.66351	0.51063

(2) Create three dummy variables for the three months from June to August, one for each month.

MODEL INFO:Q.3(b)
Observations: 48
Dependent Variable: Yval
Type: OLS linear regression

MODEL FIT:
$F(7,40) = 2.59908$, $p = 0.02617$
$R^2 = 0.31264$
Adj. $R^2 = 0.19235$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	2048.02988	47.87200	42.78137	0.00000
Newspaper	0.07453	0.19674	0.37881	0.70683
Magazine	-0.00633	0.11626	-0.05443	0.95687
Radio	0.09526	0.09157	1.04035	0.30442
TV	**0.28685**	**0.07002**	**4.09639**	**0.00020**
Dummy_June	-23.49183	80.00728	-0.29362	0.77057
Dummy_July	-28.55040	79.67851	-0.35832	0.72199
Dummy_August	-49.04901	83.33770	-0.58856	0.55947

(3) Compare the above models :

Adjusted R square **(Model(3.A)) (0.2296)** with only **single dummy** variable (Dummy_Summer) is **higher** than the **Model(3.B) (0.19235)** with **three dummy** variables (Dummy_June, Dummy_July, Dummy_August). Dummy Variables in both models are insignificant .Here we can see that in the first model we are finding the effect of summer on sales and in the second model we are calculating the effect of three months of summer on sales . So, we can see that when we are adding three different model for the same thing the Adj R square decreases .