Author: Harsh Tandon 1580393

Homework on Media Mix Modeling

Marketing 2505: Marketing Analytics
Winter 2020

NOTE: This homework is due on Monday, Jan. 27, before class. Please type the answers and bring a hard copy, no need to print out all the R outputs.

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

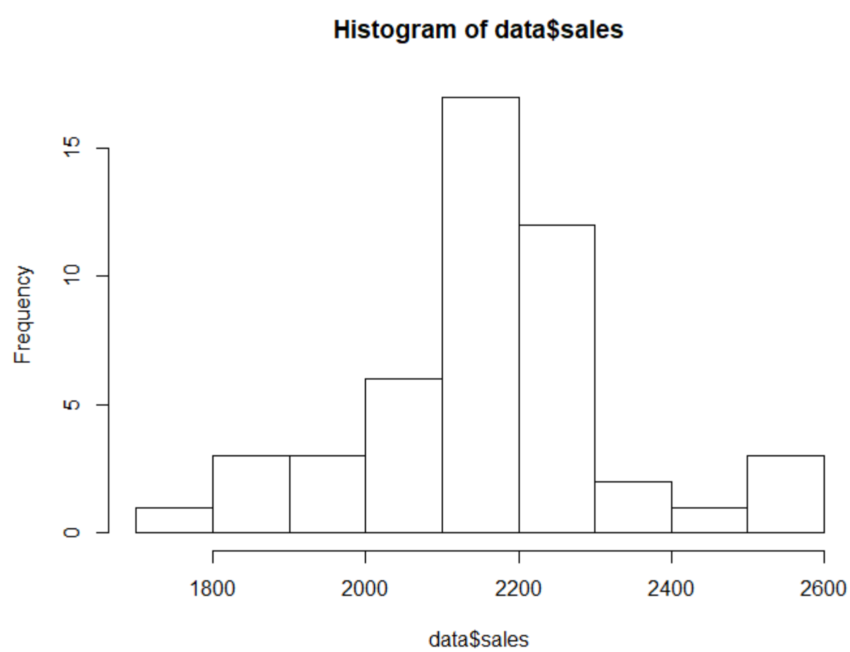
Before running any type of regression, let's look at descriptive statistics of the data. Notice a few things about our data:

- All the attributes have same number of rows. However, the dataset contains only 48 rows.
- We have the data from year 2012 to 2015.
- Even though radio is a traditional media channel, there was no periodic break in the marketing, as seen in newspaper, magazine and tv, where values are 0 periodically.

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
year	48	2,013.50	1.13	2,012	2,012.8	2,013.5	2,014.2	2,015
month	48	6.50	3.49	1	3.8	6.5	9.2	12
sales	48	2,168.69	166.11	1,765	2,099.5	2,171	2,233.5	2,575
newspaper	48	55.44	120.20	0	0	0	46	507
magazine	48	99.48	199.15	0	0	0	130.5	1,038
radio	48	378.19	340.06	2	100.5	265.5	657.8	1,340
tv	48	179.48	262.45	0	0	26.5	279.2	1,303

Let's plot sales – We don't need to log transform sales as the histogram looks normally distributed.



Author: Harsh Tandon 1580393

Start with the base model

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

	Dependent variable:
	sales
newspaper	0.08
magazine	-0.002
radio	0.281***
tv	0.08
Constant	2,043.839***

Model Statistics:

Adjusted R2	0.24
Residual Std. Error	144.843
F Statistic	4.704**

F-statistic shows that our basic model fits the data better than a null model. Adjusted R2 shows a value of 0.24 which is greater than 0.10, suggesting we have a decent model fit.

The only significant independent variable in this model is 'radio'. This means, one unit increase in advertising on radio is associated with 0.281 units increase in sales.

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?

All the media channels in this dataset are traditional channels. And we know that the result of spending on advertising in these traditional channels, is not instantaneous. Therefore, introducing lag variables in our model should improve our results. We introduce lag 1 variables.

Comparing our lagged model with basic model:

- F statistic is significant for all our presented models. This shows, all our models are significantly better than null model.
- Adjusted R2 for our lagged model increased from 0.24 to 0.322, which shows that lagged model fits our data **8.2%** (0.322 – 0.24) better than our basic model.
- Residual Standard error decreases from 144.84 to 135.81.
- In terms of model fit, Lagged Model fits better than Basic Model, however, none of the lag variables are significant.
- The only significant independent variable in the model is current-time-period-'radio', which shows one unit increase in advertising for current time period radio is associated with 0.298 units increase in sales.

Regression Results

	Dependent variable:	
	sales Basic Model (1)	Lag_sales Lagged Model (2)
newspaper	0.080 (0.190)	0.327 (0.227)
lag_newspaper		0.233 (0.187)
magazine	-0.002 (0.111)	-0.098 (0.116)
lag_magazine		-0.170 (0.141)
radio	0.281*** (0.066)	0.298*** (0.071)
lag_radio		0.078 (0.071)
tv	0.080 (0.085)	0.151 (0.096)
lag_tv		-0.020 (0.093)
Constant	2,043.839*** (43.969)	1,990.025*** (56.197)

	Basic Model	Lagged Model
Adjusted R2	0.24	0.322
Residual Std. Error	144.843	135.809
F Statistic	4.704**	3.727**

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.

We include Ad Stock Variable to capture the historical impact of marketing expenses. λ captures the discounted values.

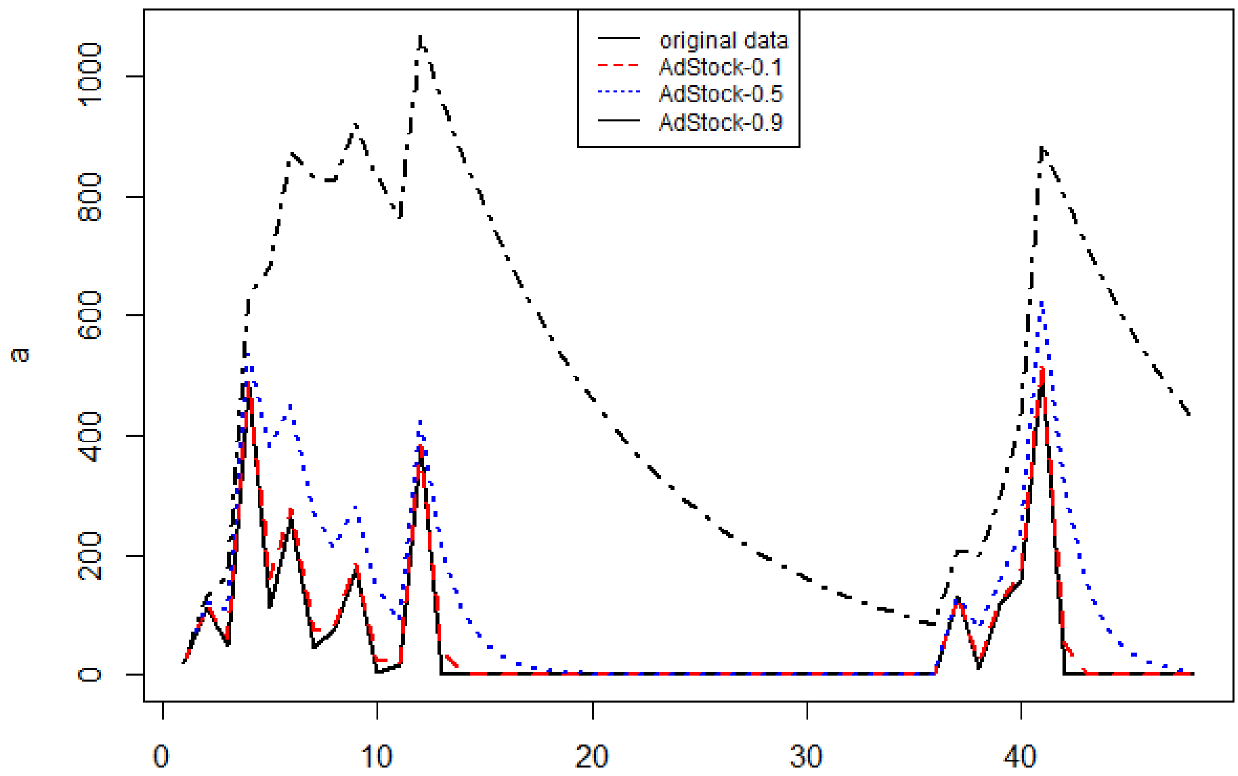
Comparing our three Ad Stock Variables with basic model:

- *Model fit in decreasing order: AdStock 1 Model > Basic Model > AdStock 5 Model > AdStock 9 Model.*
- F statistic is significant for all our presented models. This shows, all our models are significantly better than null model.
- Adjusted R2 for our AdStock1 Model (where $\lambda = 0.1$) increases from 0.24 to 0.243, which shows that AdStock1 Model fits our data **0.3%** (0.243 – 0.24) better than our basic model. However, for our dataset, as λ increases, the Adjusted R2 decreases.
- Residual Standard error decreases from 144.84 to 144.55 for AdStock1 Model. But for AdStock 5 and AdStock 9 models, residual standard errors increase.
- In terms of model fit, Ad Stock 1 Model fits the data best.
- The only significant independent variable in the model is AdStock scale variable 'radio'
    - AdStock1 Model - One unit increase in stock variable for radio channel is associated with 0.272 units increase in sales.
    - AdStock5 Model - One unit increase in stock variable for radio channel is associated with 0.182 units increase in sales.
    - AdStock9 Model - One unit increase in stock variable for radio channel is associated with 0.032 units increase in sales.

Regression Results

	Dependent variable:			
	sales			
	Basic Model (1)	AdStock1 Model (2)	AdStock5 Model (3)	AdStock9 Model (4)
newspaper	0.080	0.092	0.093	-0.017
	(0.190)	(0.191)	(0.187)	(0.180)
magazine	-0.002	-0.015	-0.054	-0.042
	(0.111)	(0.110)	(0.103)	(0.103)
radio	0.281***	0.272***	0.182***	0.032*
	(0.066)	(0.064)	(0.049)	(0.014)
tv	0.080	0.083	0.065	0.016
	(0.085)	(0.082)	(0.060)	(0.020)
Constant	2,043.839***	2,033.736***	2,009.986***	2,084.892***
	(43.969)	(47.124)	(64.600)	(82.616)

	Basic Model	AdStock1 Model	AdStock 5 Model	AdStock 9 Model
Adjusted R2	0.24	0.243	0.207	0.126
Residual Std. Error	144.843	144.548	147.893	155.285
F Statistic	4.704**	4.767**	4.073**	2.696*



3. <u>Dummy variables</u>

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

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

Compare the results of the above two models in their dummy variable estimates.

Comparing our two different dummy variable models:

- F statistic is significant for all our presented models. This shows, all our models are significantly better than null model.
- Adjusted R2 for 3 dummy variable model decreases from 0.23 to 0.192, which shows that model with one dummy variable fits our data **3.8%** (0.23 – 0.192) better than model with three dummy variables.
- Residual Standard error increases from 145.79 to 149.28.
- The only significant independent variable in the model is 'radio':
  - 1 Dummy Variable: One unit increase in advertising for radio is associated with 0.283 units increase in sales.
  - 3 Dummy Variables: One unit increase in advertising for radio is associated with 0.287 units increase in sales.
- Here both the models essentially capture the same data 'sales in summer vs sales in non-summer months'. However, these extra dummy variables in 3 dummy variable model are not increasing the model fit as much as Adjusted R2 penalizes for adding extra variables.

Author: Harsh Tandon 1580393

```
Regression Results
==================================================
                         Dependent variable:
                    ------------------------------
                                sales
                        One-Dummy     Three-Dummy
                           (1)            (2)
--------------------------------------------------
newspaper                 0.074          0.075
                         (0.191)        (0.197)

magazine                 -0.010          -0.006
                         (0.113)        (0.116)

radio                   0.283***       0.287***
                         (0.066)        (0.070)

tv                        0.090          0.095
                         (0.087)        (0.092)

isSummer1               -33.284
                        (50.164)

isJune                                  -23.492
                                        (80.007)

isJuly                                  -28.550
                                        (79.679)

isAugust                                -49.049
                                        (83.338)

Constant              2,050.707***   2,048.030***
                        (45.452)        (47.872)

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

	1 Dummy Variable	3 Dummy Variables
Adjusted R2	0.23	0.192
Residual Std. Error	145.795	149.284
F Statistic	3.802**	2.599*

4. Nonlinear transformation of the X variables

(1) Add the square terms of the X variables, together with the linear form. How do you like the estimated results?
(2) Use ln-transformation for all the X variables, and include only the ln(X+1) into the model, without the linear term, and how do you like the results?
(3) Do these nonlinear transformation make sense in this context? Do they provide better fit comparing to the basic model?

Comparing our models:

- *Model fit in decreasing order: Basic Model > Sq Terms Model > Log Transformed Sq Terms Model*.
- F statistic is significant for all our presented models. This shows, all our models are significantly better than null model.
- Adjusted R2 for Basic model is greater, which shows that basic model fits our data **2.8%** (0.24 – 0.212) better than sq terms model.
- Residual Standard error increases from 144.84 to 146.36.
- In Sq Terms Model, both radio and radio square are significant.

Author: Harsh Tandon 1580393

Regression Results

	Dependent variable:		
	sales	Lag_sales	
	Basic Model	Sq Terms	Log-Trans Sq Terms
	(1)	(2)	(3)
newspaper	0.080	-0.275	
	(0.190)	(0.708)	
newspaper2		0.001	-7.217
		(0.002)	(9.767)
magazine	-0.002	-0.306	
	(0.111)	(0.278)	
magazine2		0.0003	-0.234
		(0.0003)	(8.297)
radio	0.281***	0.803**	
	(0.066)	(0.228)	
radio2		-0.001**	12.620
		(0.0002)	(8.909)
tv	0.080	-0.077	
	(0.085)	(0.188)	
tv2		0.0001	-4.188
		(0.0002)	(4.332)
Constant	2,043.839***	2,029.081***	2,076.807***
	(43.969)	(63.820)	(109.512)
Observations	48	47	47
R2	0.304	0.349	0.122
Adjusted R2	0.240	0.212	0.039
Residual Std. Error	144.843	146.357	161.685
F Statistic	4.704**	2.549*	1.462