The impact of sales promotion on shampoo sales

James/Peng Li 14/09/2024

Contents

Executive Summary: This assignment aims to measure the effects of sales promotions on sales, with a focus on estimating a descriptive model that examines the own and cross effects of price promotions for the brand with the highest market share. Based on this criterion, Brand 2 was selected from 11 brands for analysis. We developed three models progressively, introducing new variables at each stage to assess their impact. For model 1, we initially selected variables for Brand 2 based on primary analyses, such as correlation tests. We found that promotion variables, such as the promotion dummy for the brand, had significant positive effects on sales. For model 2, the cross-brand variables for Brands 3, 4, 5, and 8 were added (as other brands had too many missing values), which improved model performance, as seen in the R-squared and Durbin-Watson statistics. For model 3, cross-store dummy variables were introduced, significantly enhancing the model with a higher R-squared and improved DW values. Building on Model 3, we incorporated time series analysis to further refine the model, splitting the dataset for prediction analysis.Our findings show that special packages and promotional activities had a substantial impact on Brand 2's sales. Additionally, cross-brand promotions influenced sales, and promotions were more effective in high-sales stores, indicating that larger stores should be a focus. However, the model's predictive accuracy for Brand 2 was limited, likely due to missing variables and performance differences across stores. Further store-specific analysis is recommended to improve results.

1. INTRODUCTION

1.1 Main target

The effectiveness of a sales promotion can be examined by decomposing the sales "bump" during the promotion period into sales increase due to brand switching, purchase time acceleration, and stockpiling[1]. As the focus company has seldom studied the impacts of its sales promotion programs, while continuing to launch the promotion, the firm has had to bear a great deal of sales promotion burden[2]. In this assignment, we will measure the effects of sales promotions on sales. The main objective is to estimate a descriptive model that examines the own and cross effects of (price) promotions on sales for the (one) brand with the highest market share.

1.2 Variables description

- Original file: shampoo
- Number of brands: 11
- Number of observations: 5195
- Number of stores: 48
- Number of weeks per store (max.): 109
- Note: not every store has 109 weeks
- Period: week 20 in 2016 until week 24 in 2018
- Missing value indicator: . in SAS

2. Data description

2.1 Descriptive labels for the variables

The original data set lacks descriptive labels for all variables. We will add the appropriate labels using a SAS program. Since there are many variables, we will initially focus on providing descriptive labels for the variables that are most likely to be used, with additional labels to be added later as needed.

2.2 Deal with missing values

From below chart, we can find that for some variables there are missing values. For brands 2 (26 missing), 3(57 missing), 4(57 missing), 5(8 missing) and 8(94 missing), there are few missing values, while for brands 1(1855 missing), 6(1847 missing), 7(677 missing), 9(1488 missing), 10(1470 missing) and 11(1091 missing) there are many missing values. We use “mean” for adding values to the missing values, but only values of brand 2,3,4,5,8 will be used for analysis in next steps.

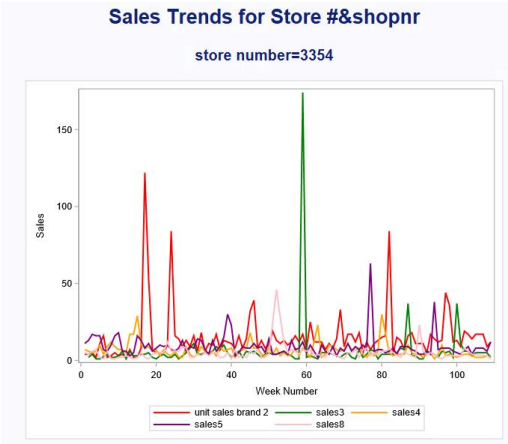
Summary Statistics
Results

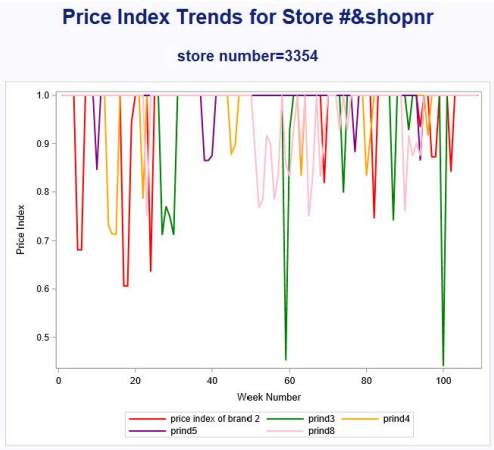
The MEANS Procedure

Variable	Label	Mean	Std Dev	Minimum	Maximum	N	N Miss
shopnr	store number	3468.66	74.7154948	3354.00	3685.00	5195	0
weeknr	week number	55.1424447	31.4365972	1.0000000	109.0000000	5195	0
sales1		4.2946108	3.3649045	1.0000000	48.0000000	3340	1855
actpr1		4.8669401	0.3469892	3.1500000	5.0500000	3340	1855
regpr1		4.9324012	0.1681379	3.3900000	5.0500000	3340	1855
pci1		0.0559880	0.2299330	0	1.0000000	3340	1855
prom1		0.0479042	0.4566684	0	5.0000000	3340	1855
spec11		0.0026946	0.1557289	0	9.0000000	3340	1855
spec21		0	0	0	0	3340	1855
spec31		0	0	0	0	3340	1855
spec41		0	0	0	0	3340	1855
sales2	unit sales brand 2	23.3507448	20.3587651	1.0000000	231.0000000	5169	26
actpr2		5.3439485	0.4534929	3.3300000	6.1600000	5169	26
regpr2		5.4777404	0.0535201	4.7400000	6.1600000	5169	26
pci2	dummy for price discount for brand 2	0.0942155	0.2921588	0	1.0000000	5169	26
prom2	categorical promotion variabele for brand 2	0.5059006	1.5252055	0	7.0000000	5169	26
spec12	intensity of special pack 1 for brand2	0.1118205	0.9396832	0	9.0000000	5169	26
spec22	intensity of special pack 2 for brand2	0	0	0	0	5169	26
spec32	intensity of special pack 3 for brand2	0.0092861	0.0959254	0	1.0000000	5169	26
spec42	intensity of special pack 4 for brand2	0.0835752	0.8633283	0	9.0000000	5169	26
sales3		11.3853640	18.2840637	1.0000000	363.0000000	5138	57
actpr3		6.3853387	0.7219234	2.3500000	10.0000000	5138	57
regpr3		6.6798170	0.2191671	5.4500000	10.0000000	5138	57
pci3		0.1998832	0.3999513	0	1.0000000	5138	57
prom3		0.2162320	0.9555801	0	6.0000000	5138	57
spec13		0.9021020	2.4454236	0	9.0000000	5138	57
spec23		0.0805761	0.8478398	0	9.0000000	5138	57
spec33		0.000389257	0.0197277	0	1.0000000	5138	57
spec43		0	0	0	0	5138	57
sales4		9.5515765	7.3287713	1.0000000	99.0000000	5138	57
actpr4		6.2123161	0.5058719	4.9400000	6.9900000	5138	57
regpr4		6.2704048	0.4576098	4.9900000	6.9900000	5138	57
pci4		0.0628649	0.2427435	0	1.0000000	5138	57
prom4		0.1930712	0.8956097	0	5.0000000	5138	57
spec14		0	0	0	0	5138	57

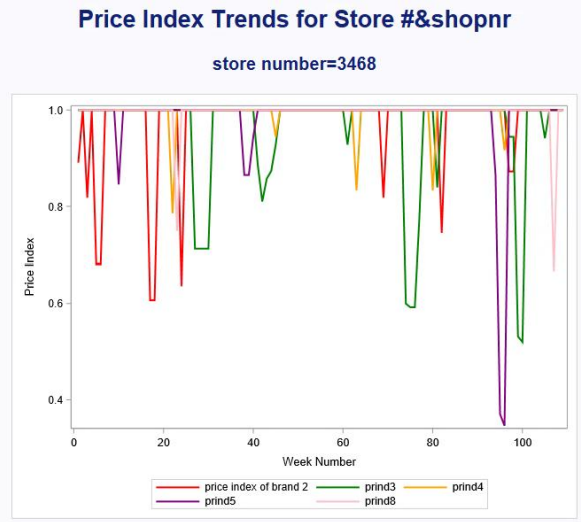
2.3 Make plots of the key variables

We will plot “salesi,” “prindi,” and “promi_j” over time to assess whether the patterns are consistent and to identify any potential outliers. Since this is panel data, we will generate separate plots for each store. Taking store 3354 as an example below, we observe sudden spikes in sales for the 11 brands during certain weeks. Correspondingly, there are noticeable drops in “prindi” (price) and increase in “promi_j” (promotion), suggesting that price reductions and promotions are likely the main drivers of these sales increases.





After reviewing all 48 stores, we found that nearly all exhibit the same pattern as store 3354, where sudden sales increases are typically below 200 units. However, two data points from store 3468 show unusually large sales spikes for sales5, around 800 units, which could be potential outliers. We will remove these two outliers for further analysis.



2.4 Descriptive statistics for data set

To select the brand with the highest market share, we analyzed unit sales and total sales in dollars. From the charts below, we observed that both sales2 and sales5 have the highest unit sales, with approximately 120,000 units each. By calculating a new variable, "SalesT" (based on unit sales "sales" and actual unit price "actpri"), we found that Brand 2 has the highest market share in terms of total dollar sales. Additionally, Brand 2 has very few missing values and high data quality, making it an ideal choice for analysis. Labels were added for the relevant variables associated with Brand 2.

Variable	Mean	Minimum	Maximum	Sum	N
sales1	4.2946108	1.0000000	48.0000000	14344.00	3340
sales2	23.3507448	1.0000000	231.0000000	120700.00	5169
sales3	11.3853640	1.0000000	363.0000000	58498.00	5138
sales4	9.5515765	1.0000000	99.0000000	49076.00	5138
sales5	23.6148062	1.0000000	818.0000000	122490.00	5187
sales6	7.4877539	1.0000000	68.0000000	25069.00	3348
sales7	4.4338203	1.0000000	53.0000000	20032.00	4518
sales8	11.5153891	1.0000000	218.0000000	58740.00	5101
sales9	3.4451039	1.0000000	19.0000000	12771.00	3707
sales10	2.8676510	1.0000000	16.0000000	10682.00	3725
sales11	5.4522417	1.0000000	68.0000000	22376.00	4104

Variable	Mean	Std Dev	Minimum	Maximum	Sum	N
SalesT1	20.5961168	14.7638457	3.3900000	167.5200000	68791.03	3340
SalesT2	121.6874637	96.2103295	3.3300000	1005.76	629002.50	5169
SalesT3	65.7258097	67.9903177	4.9500000	1099.89	337699.21	5138
SalesT4	58.8537524	42.5263579	4.9400000	494.0100000	302390.58	5138
SalesT5	65.0351321	44.7303610	2.5900000	875.2600000	337337.23	5187
SalesT6	43.8219176	31.4595686	4.9900000	339.3200000	146715.78	3348
SalesT7	26.1245418	19.6618871	3.9900000	253.0000000	118030.68	4518
SalesT8	43.9513821	37.7387951	2.6600000	660.5400000	224196.00	5101
SalesT9	13.6233936	8.6830527	1.9900000	75.2400000	50501.92	3707
SalesT10	17.0091758	12.5806126	2.7500000	95.8400000	63359.18	3725
SalesT11	32.2044371	22.5755609	4.4000000	355.6400000	132167.01	4104

Correlation analysis: we can find that the variables "prom" and "spec" also have good correlation with sales2T, so we will consider these variables in the model.

TYPE	_NAME_	sales2T	sales2	actpr2	regpr2	pci2	prom2	spec12	spec22	spec32	spec42	prind2	prom2_1	prom2_2	prom2_3	prom2_4	prom2_5	prom2_6	prom2_7
MEAN		121.70	23.35	5.34	5.48	0.09	0.51	0.11	0	0.01	0.08	0.98	0.02	0.00	0.02	0.02	0.04	0.01	0.02
STD		95.97	20.31	0.45	0.05	0.29	1.52	0.94	0	0.10	0.86	0.08	0.12	0.04	0.13	0.13	0.19	0.09	0.13
N		5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00	5195.00
CORR	sales2T	1.00	0.97	-0.18	-0.03	0.25	0.33	0.11	.	-0.09	0.15	-0.17	-0.04	0.04	0.01	0.14	0.26	0.17	0.11
CORR	sales2	0.97	1.00	-0.34	-0.03	0.38	0.43	0.09	.	-0.08	0.16	-0.33	-0.01	0.06	0.02	0.14	0.38	0.22	0.11
CORR	actpr2	-0.18	-0.34	1.00	0.07	-0.91	-0.65	-0.03	.	0.03	-0.12	0.99	-0.26	-0.12	-0.29	-0.07	-0.56	-0.23	-0.15
CORR	regpr2	-0.03	-0.03	0.07	1.00	0.04	0.05	-0.02	.	0.01	0.01	-0.05	0.06	0.01	0.01	0.03	0.03	0.02	0.01
CORR	pci2	0.25	0.38	-0.91	0.04	1.00	0.69	0.05	.	-0.03	0.30	-0.92	0.23	0.13	0.25	0.21	0.57	0.27	0.13
CORR	prom2	0.33	0.43	-0.65	0.05	0.69	1.00	0.33	.	-0.03	0.26	-0.66	0.04	0.04	0.22	0.31	0.58	0.31	0.59
CORR	spec12	0.11	0.09	-0.03	-0.02	0.05	0.33	1.00	.	-0.01	-0.01	-0.03	-0.01	-0.00	0.14	-0.02	-0.02	-0.01	0.51
CORR	spec22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CORR	spec32	-0.09	-0.08	0.03	0.01	-0.03	-0.03	-0.01	.	1.00	-0.01	0.03	-0.01	-0.00	-0.01	-0.01	-0.02	-0.01	-0.01
CORR	spec42	0.15	0.16	-0.12	0.01	0.30	0.26	-0.01	.	-0.01	1.00	-0.12	0.07	-0.00	-0.01	0.03	0.40	0.02	-0.01
CORR	prind2	-0.17	-0.33	0.99	-0.05	-0.92	-0.66	-0.03	.	0.03	-0.12	1.00	-0.27	-0.12	-0.29	-0.08	-0.57	-0.24	-0.16
CORR	prom2_1	-0.04	-0.01	-0.26	0.06	0.23	0.04	-0.01	.	-0.01	0.07	-0.27	1.00	-0.01	-0.02	-0.02	-0.02	-0.01	-0.02
CORR	prom2_2	0.04	0.06	-0.12	0.01	0.13	0.04	-0.00	.	-0.00	-0.00	-0.12	-0.01	1.00	-0.01	-0.01	-0.01	-0.00	-0.01
CORR	prom2_3	0.01	0.02	-0.29	0.01	0.25	0.22	0.14	.	-0.01	-0.01	-0.29	-0.02	-0.01	1.00	-0.02	-0.03	-0.01	-0.02
CORR	prom2_4	0.14	0.14	-0.07	0.03	0.21	0.31	-0.02	.	-0.01	0.03	-0.08	-0.02	-0.01	-0.02	1.00	-0.03	-0.01	-0.02
CORR	prom2_5	0.26	0.38	-0.56	0.03	0.57	0.58	-0.02	.	-0.02	0.40	-0.57	-0.02	-0.01	-0.03	-0.03	1.00	-0.02	-0.03
CORR	prom2_6	0.17	0.22	-0.23	0.02	0.27	0.31	-0.01	.	-0.01	0.02	-0.24	-0.01	-0.00	-0.01	-0.01	-0.02	1.00	-0.01
CORR	prom2_7	0.11	0.11	-0.15	0.01	0.13	0.59	0.51	.	-0.01	-0.01	-0.16	-0.02	-0.01	-0.02	-0.02	-0.03	-0.01	1.00

2.5 Measurement Equivalence

After checking the overall statistic summary for all the variables, we find that the variables such like sales in unites, price related variables are in reasonable range, so they should be in equivalent units across all brands.

3. Model specification

3.1 Methodology

The Scan*Pro model developed by Wittink et al. in 1988 is a widely recognized store-level sales response model used to analyze and predict the effect of marketing activities (like pricing, promotions, and advertising) on product sales. This model was particularly useful for FMCG (Fast-Moving Consumer Goods) companies looking to optimize marketing efforts based on scanner data collected from retail outlets. Firms spend a significant part of their marketing budgets on sales promotions. The Trade Promotion report (2005) indicates that during 1997–2004, promotion accounted for roughly 75% of marketing expenditures for US packaged goods manufacturers; the other 25% was for advertising. In 2004, 59% of the budget was spent

on promotion to the trade (i.e., from manufacturers to retailers), and 16% on manufacturer promotions to consumers[3].

3.2 Variables selection and and preparation

Based on the analysis, we selected the appropriate variables and opted to use a multiplicative model. To prepare the data for this model, we transformed continuous variables using logarithmic functions and converted integer variables into dummy variables.

Only for brand 2

- dependent variable: LN(SalesT2)=LN(sales2*actpr2)
- independent variable for price: LN(regpr2), LN(prind2)
- independent variables for non-price promotions:

  - Special pack variables: spec12_dummy, spec32 _dummy, spec42_dummy ; (spec22 are all 0);

  - promotion dummy variables: prom2_1, prom2_2, prom2_3, prom2_4, prom2_5, prom2_6, prom2_7;

Cross brand (only choose 3, 4,5,8)

- LN(prind3), LN(prind4), LN(prind5), LN(prind8)
- prom3_1, prom3_2, prom3_4, prom3_5, prom3_6, prom4_1, prom4_4, prom4_5, prom5_1, prom5_2, prom5_4, prom5_5, prom5_6, prom8_1, prom8_4, prom8_5

Cross store

- Stores: invent dummy variables for each store

4. Results

4.1 Model 1: only include the variables of brand 2

- ***use independent variable for price***

LN(SalesT2)=λ+β1LN(regpr2)+ LN(prind2)+μ

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LNsales2T

Number of Observations Read	5195
Number of Observations Used	5195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	34.62222	17.31111	23.65	<.0001
Error	5192	3800.88396	0.73207		
Corrected Total	5194	3835.50618			

Root MSE	0.85561	R-Square	0.0090
Dependent Mean	4.48798	Adj R-Sq	0.0086
Coeff Var	19.06446		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	13.24770	2.03087	6.52	<.0001	0
LNregpr2	1	-5.16206	1.19425	-4.32	<.0001	1.00209
LNprind2	1	-0.65443	0.11811	-5.54	<.0001	1.00209

The R-Sq is very low, so the model needs to be improved with adding and testing more variables.

- ***add special pack variables***

LN(SalesT2)=λ+β1*LN(regpr2)+β2*LN(prind2)+γ1*spec12_dummy+γ2*spec32_dummy+γ3*spec42_dummy+μ

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LNsales2T

Number of Observations Read	5195
Number of Observations Used	5195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	182.25707	36.45141	51.77	<.0001
Error	5189	3653.24911	0.70404		
Corrected Total	5194	3835.50618			

Root MSE	0.83907	R-Square	0.0475
Dependent Mean	4.48798	Adj R-Sq	0.0466
Coeff Var	18.69594		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	13.35591	1.99244	6.70	<.0001	0
LNregpr2	1	-5.22658	1.17168	-4.46	<.0001	1.00298
LNprind2	1	-0.54439	0.11648	-4.67	<.0001	1.01343
spec12_dummy	1	0.79822	0.11930	6.69	<.0001	1.00098
spec32_dummy	1	-1.25034	0.12175	-10.27	<.0001	1.00119
spec42_dummy	1	0.92977	0.12231	7.60	<.0001	1.01053

Although the R-squared value has increased, it remains relatively low, indicating that further improvements are needed. The coefficients for LNregpr2 and LNprind2 are still negative, which is consistent with expectations: actual price increases or lack of price decreases, tend to have a negative impact on sales.

- ***add promotion dummy variables (only choose brand 3, 4,5,8)***

Prom2_1, prom2_2, prom2_3, prom2_4, prom2_5, prom2_6, prom2_7

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LNsales2T

Number of Observations Read	5195
Number of Observations Used	5195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	383.62460	31.96872	47.99	<.0001
Error	5182	3451.88158	0.66613		
Corrected Total	5194	3835.50618			

Root MSE	0.81617	R-Square	0.1000
Dependent Mean	4.48798	Adj R-Sq	0.0979
Coeff Var	18.18565		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	13.61663	1.94111	7.01	<.0001	0
LNregpr2		1	-5.39149	1.14155	-4.72	<.0001	1.00623
LNprind2		1	1.05326	0.19638	5.36	<.0001	3.04437
prom2_1	promotion 1 dummy for brand 2	1	-0.26416	0.10581	-2.50	0.0126	1.32388
prom2_2	promotion 2 dummy for brand 2	1	0.93666	0.27830	3.37	0.0008	1.04457
prom2_3	promotion 3 dummy for brand 2	1	0.22737	0.10099	2.25	0.0244	1.38349
prom2_4	promotion 4 dummy for brand 2	1	0.70876	0.08662	8.18	<.0001	1.03953
prom2_5	promotion 5 dummy for brand 2	1	1.09034	0.09590	11.37	<.0001	2.56509
prom2_6	promotion 6 dummy for brand 2	1	1.45416	0.14491	10.04	<.0001	1.18905
prom2_7	promotion 7 dummy for brand 2	1	0.60029	0.10953	5.48	<.0001	1.69680
spec12_dummy		1	0.34158	0.14544	2.35	0.0189	1.57255
spec32_dummy		1	-1.23035	0.11844	-10.39	<.0001	1.00155
spec42_dummy		1	0.23182	0.13517	1.72	0.0864	1.30444

The R-Sq is increased but still low, so further improvement is needed. We can find that prom2_6 (display with feature with picture) and prom2_5(display with feature with just text) have better effects on sales than other ways. spec32_dummy (two bottles with free other goods) has negative effects on sales, and spec12_dummy (two bottles with price discount) has positive effects on sales.

4.2 Model 2: add cross brand variables

- LN(prind3), LN(prind4), LN(prind5), LN(prind8)

- prom3_1, prom3_2, prom3_4, prom3_5, prom3_6, prom4_1, prom4_4, prom4_5, prom5_1, prom5_2, prom5_4, prom5_5, prom5_6, prom8_1, prom8_4, prom8_5

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LNsales2T

Number of Observations Read	5195
Number of Observations Used	5195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	32	543.56231	16.98632	26.64	<.0001
Error	5162	3291.94387	0.63773		
Corrected Total	5194	3835.50618			

Root MSE	0.79858	R-Square	0.1417
Dependent Mean	4.48798	Adj R-Sq	0.1364
Coeff Var	17.79372		

Durbin-Watson D	0.625
Number of Observations	5195
1st Order Autocorrelation	0.687

The R-Sq is increasing to 0.1364 and DW is also increasing a bit to 0.625. After adding the cross brand variables, the model is improved, which means the price change and promotion of other brands have relatively high influence on brand 2's sales.

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	12.55650	1.94725	6.45	<.0001	0
LNregpr2		1	-4.75302	1.14532	-4.15	<.0001	1.05759
LNprind2		1	0.99100	0.19371	5.12	<.0001	3.09294
LNprind3		1	0.36850	0.10266	3.59	0.0003	1.62253
LNprind4		1	0.14277	0.40528	0.35	0.7247	2.30283
LNprind5		1	0.44724	0.57858	0.77	0.4396	3.06857
LNprind8		1	0.61280	0.33056	1.85	0.0638	3.47296
prom2_1	promotion 1 dummy for brand 2	1	-0.30510	0.10404	-2.93	0.0034	1.33626
prom2_2	promotion 2 dummy for brand 2	1	0.89735	0.27253	3.29	0.0010	1.04599
prom2_3	promotion 3 dummy for brand 2	1	0.25262	0.09942	2.54	0.0111	1.40008
prom2_4	promotion 4 dummy for brand 2	1	0.68625	0.08507	8.07	<.0001	1.04690
prom2_5	promotion 5 dummy for brand 2	1	1.04372	0.09446	11.05	<.0001	2.59858
prom2_6	promotion 6 dummy for brand 2	1	1.41055	0.14221	9.92	<.0001	1.19570
prom2_7	promotion 7 dummy for brand 2	1	0.63241	0.10841	5.83	<.0001	1.73561
prom3_1		1	-0.14056	0.14351	-0.98	0.3274	1.02655
prom3_2		1	0.33650	0.11055	3.04	0.0023	1.22938
prom3_4		1	-0.05363	0.08771	-0.61	0.5410	1.18248
prom3_5		1	-0.52430	0.11513	-4.55	<.0001	1.04913
prom3_6		1	0.82002	0.12560	6.53	<.0001	1.20005
prom4_1		1	-0.01499	0.16691	-0.09	0.9285	1.60410
prom4_4		1	-0.25420	0.08146	-3.12	0.0018	1.26867
prom4_5		1	0.11794	0.11718	1.01	0.3142	1.92406
prom5_1		1	0.23150	0.15543	1.49	0.1364	1.39115
prom5_2		1	-1.09045	0.31700	-3.44	0.0006	1.10113
prom5_4		1	-0.08377	0.09802	-0.85	0.3928	1.31744
prom5_5		1	0.35372	0.11954	2.96	0.0031	1.76432
prom5_6		1	-1.21699	0.15897	-7.66	<.0001	1.57222
prom8_1		1	0.15267	0.14560	1.05	0.2944	1.57994
prom8_4		1	0.36118	0.10654	3.39	0.0007	1.48702
prom8_5		1	0.38969	0.13714	2.84	0.0045	2.46408
spec12_dummy		1	0.29320	0.15854	1.85	0.0645	1.95103
spec32_dummy		1	-1.24938	0.11610	-10.76	<.0001	1.00478
spec42_dummy		1	0.24193	0.13240	1.83	0.0677	1.30666

From the coefficients, we can observe that if Brand 3 (where the coefficient for LNprind3 is positive and other coefficients are not significant) decreases its price, it will result in a decrease in sales for Brand 2. Additionally, for Brand 5, the variables prom5_2 (feature-only with picture) and prom5_6 (display with feature and picture) have a significant negative impact on the sales of Brand 2.

4.3 Model 3: add cross store variables

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LNsales2T

Number of Observations Read	5193
Number of Observations Used	5193

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	1995.06357	25.25397	70.20	<.0001
Error	5113	1839.24013	0.35972		
Corrected Total	5192	3834.30369			

Root MSE	0.59977	R-Square	0.5203
Dependent Mean	4.48827	Adj R-Sq	0.5129
Coeff Var	13.36294		

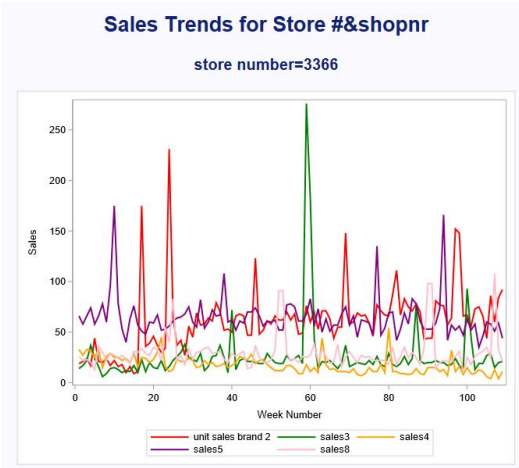
Durbin-Watson D	1.092
Number of Observations	5195
1st Order Autocorrelation	0.454

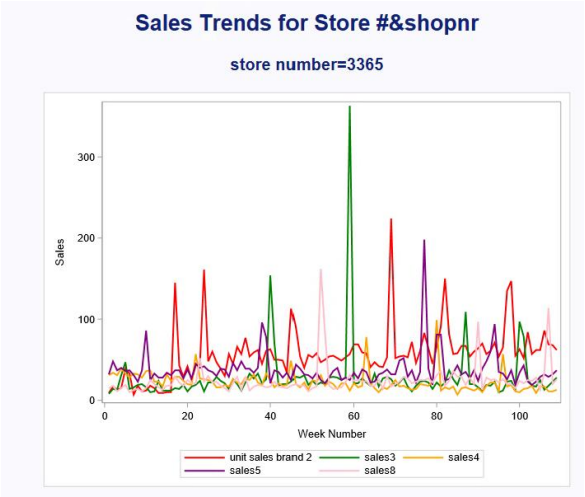
After adding the cross-store variables, we observed a significant increase in the R-squared value to 0.5129 and a slight increase in the Durbin-Watson statistic to 1.09. This indicates that store-specific factors play a major role in influencing sales. It suggests that there are substantial differences between stores, even when similar promotions or pricing strategies are applied.

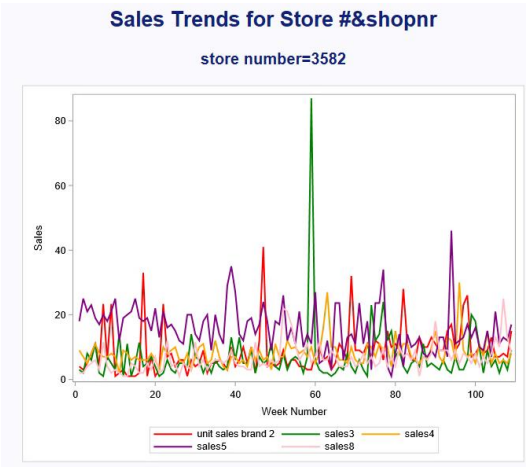
Parameter Estimates

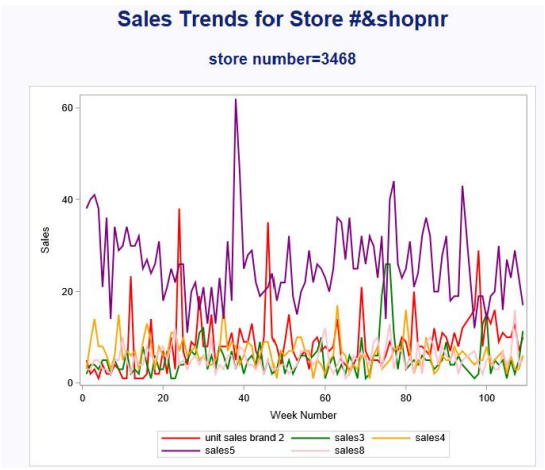
Variable	Label	DF	Parameter	Standard Error	tValue	Pr>|t|	Variance Inflation
shopnr3366	store number 3366	B	0.8262	0.08375	9.86	<.0001	2.08089
shopnr3365	store number 3365	B	0.76089	0.08354	9.11	<.0001	2.07031
...	...	...	...	...	...	...	...
shopnr3582	store number 3582	B	-1.2007	0.08339	-14.4	<.0001	2.06279
shopnr3468	store number 3468	B	-1.21627	0.08373	-14.53	<.0001	2.04221

From the ranking of the coefficients across all stores, we observe that some stores are highly affected by promotions, such as Shopnr 3366 and Shopnr 3365. In contrast, other stores, such as Shopnr 3582 and Shopnr 3468, show a lower sensitivity to promotions.









Model: Yule Walker/Prais Winsten Estimates

Regression Analysis with Autoregressive Errors

The AUTOREG Procedure

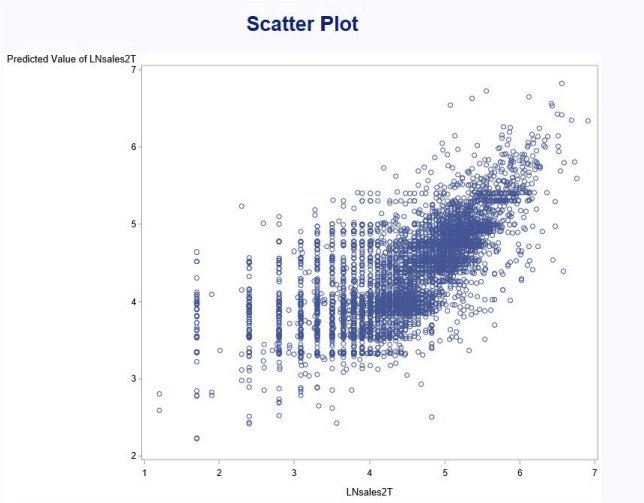
Yule-Walker Estimates			
SSE	1338.95652	DFE	5111
MSE	0.26198	Root MSE	0.51184
SBC	8400.14044	AIC	7862.62496
MAE	0.36469104	AICC	7865.28875
MAPE	9.76214229	HQC	8050.65485
Durbin-Watson	1.9735	Transformed Regression R-Square	0.3892
		Total R-Square	0.6508

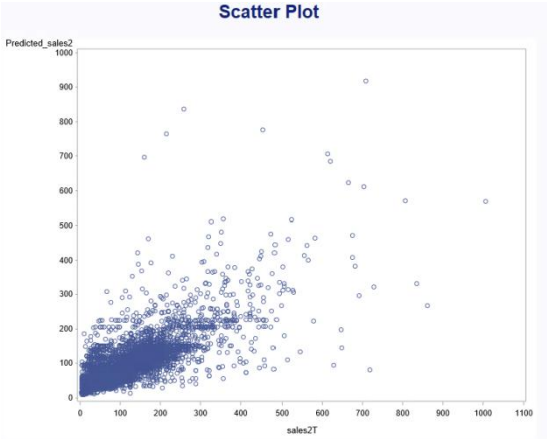
Durbin-Watson Statistics	
Order	DW
1	1.9735

We can find that the Yule Walker/Prais Winsten Estimates lead to significant increase of R-squared values and DW**.**

Model: split the data set for prediction and validation

Split original data set into 75% training and 25% validation. And below we compared the real value and predicted values related to sales of brand 2.





Overall, the predicted results are generally accurate; however, there are still some noticeable outliers. These outliers may arise due to various factors, such as noise in the data, limitations in the model, or the presence of unaccounted variables.

5. Discussion

In this report, we focus exclusively on Brand 2 for analysis, as it has the largest market share and fewer missing values. Our key findings are as follows:

- **Special Packs:** The impact of special packs is inconsistent, particularly the "two bottles with free additional goods" offer for Brand 2. The free items don't appear to influence purchasing decisions, and if customers don't need the free goods, it may even discourage them from buying. However, the "two bottles with a price discount" package has shown positive effects on sales, making it a more effective promotion strategy for Brand 2.

- **Promotion Effectiveness:** Promotions 6 (display with a featured picture) and Promotions 5 (display with a text-only feature) have demonstrated stronger effects on sales compared to other methods. Therefore, Brand 2 should focus on utilizing these two promotional strategies to maximize sales.

- **Cross-Brand Effects:** Analysis of the coefficients reveals that when Brand 3 decreases its price (with a significant positive LNprind3 coefficient), it negatively impacts the sales of Brand 2. Brand 3's (Garnier Fructis) average price is approximately $6.68 for 250ml, while Brand 2 (Herbal Essences) is priced at around $5.48 for 300ml. When Brand 3 reduces its price, it creates direct competition for Brand 2 and influences its sales. Additionally, certain promotions from Brand 5, specifically prom5_2 (feature-only with a picture) and prom 5_6 (display with feature and picture), have a significant negative effect on Brand 2's sales. Further analysis of Brand 5's promotion details is recommended.

- **Cross-Store Effects:** In some stores, promotional efforts show better results, particularly in stores with higher overall sales. This suggests that demand is higher in these locations and that consumers may be more responsive to promotions. Strengthening promotional activities in larger, high-performing stores could yield better results for Brand 2. Moreover, performance varies significantly across stores, suggesting that a store-by-store analysis may be more effective. Collecting more detailed data for specific stores could enable deeper analysis and more tailored promotional strategies. Given the differences in store locations and customer demographics, promotional outcomes are likely to vary, necessitating customized promotion plans for each store based on its unique characteristics.

- **Model Prediction Accuracy:** The overall predictive power of the model for Brand 2 is not particularly strong, which aligns with the relatively low R-squared value. There may be additional hidden variables that could provide valuable insights but are not currently included in the analysis.

Reference:

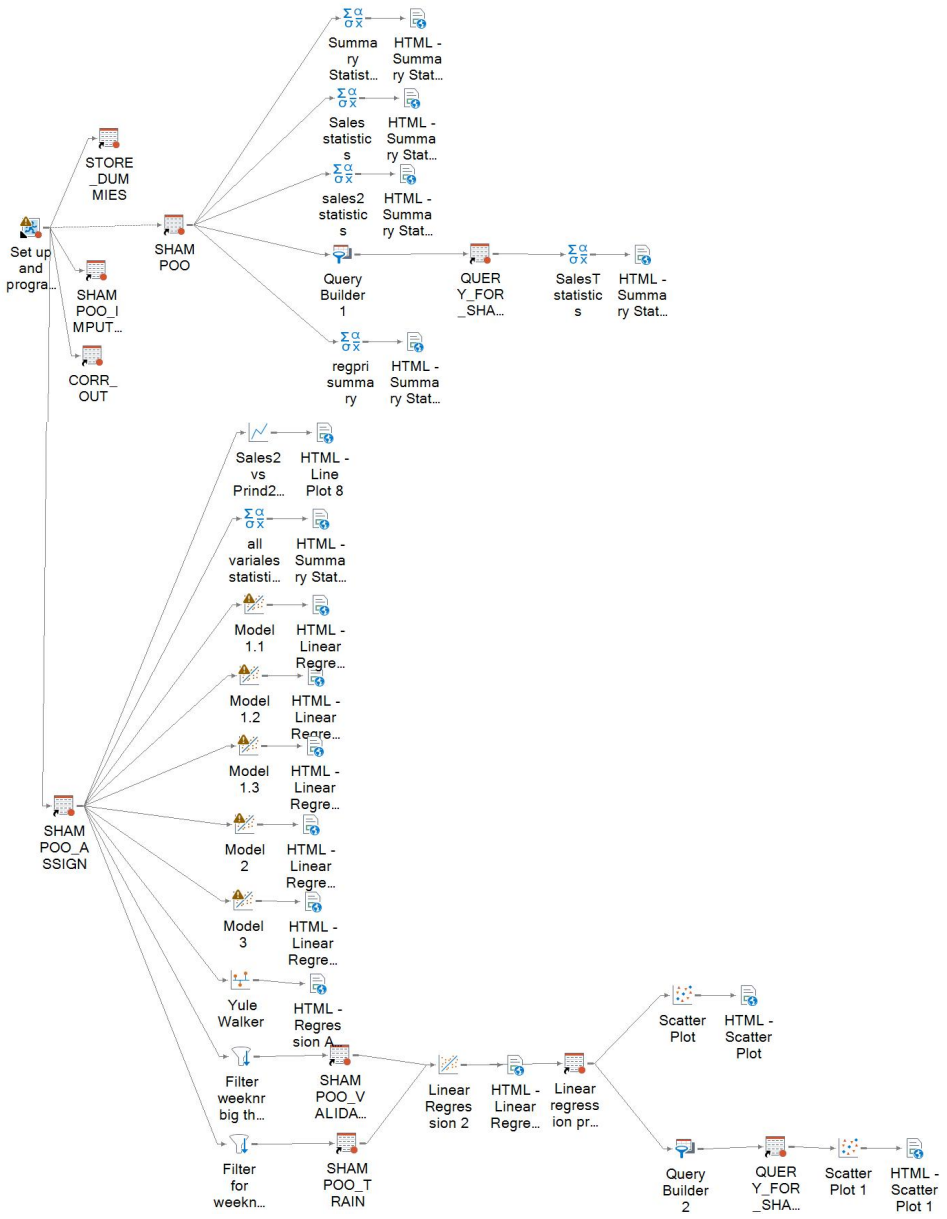
[1] Gupta, S. (1988). Impact of Sales Promotions on when, what, and how Much to Buy. *Journal of Marketing Research*, 25(4), 342–355. https://doi.org/10.1177/002224378802500402
[2] Kansa, M. (2018, October 24). *ANALYSIS OF THE EFFECT OF PROMOTION ON SALES*. https://jscm.au.edu/index.php/jscm/article/view/150
[3] Van Heerde, H. J., & Neslin, S. A. (2008). Sales promotion models. In *International series in management science/operations research/International series in operations research & management science* (pp. 107–162). https://doi.org/10.1007/978-0-387-78213-3_5

Complete code:

```
/*set up libname to access chocolate data set*/
/*path file of where it is located on your computer should be used*/
libname choc "H:\Return of marketing investment";

* Deal with missing values;
proc contents data=CHOC.SHAMPOO;
run;
proc means data=CHOC.SHAMPOO;

/* Simple mean imputation for missing values */
proc stdize data=CHOC.SHAMPOO reponly method=mean out=SHAMPOO_imputed_data;
run;

* plot to find outliers;
/* Sort the data by store number */
proc sort data=CHOC.SHAMPOO;
   by shopnr;
run;
/* Generate plots for each store */
proc sgplot data=CHOC.SHAMPOO;
   by shopnr; /* Separate plot for each store */
   series x=weeknr y=sales2 / lineattrs=(color=red thickness=2);
   series x=weeknr y=sales3 / lineattrs=(color=green thickness=2);
   series x=weeknr y=sales4 / lineattrs=(color=orange thickness=2);
   series x=weeknr y=sales5 / lineattrs=(color=purple thickness=2);
   series x=weeknr y=sales8 / lineattrs=(color=pink thickness=2);
   xaxis label="Week Number";
   yaxis label="Sales";
   title "Sales Trends for Store #&shopnr";
run;

/* Sort the data by store number */
proc sort data=CHOC.SHAMPOO;
   by shopnr;
run;
/* Generate plots for each store */
proc sgplot data=CHOC.SHAMPOO;
   by shopnr; /* Separate plot for each store */
   series x=weeknr y=prind2 / lineattrs=(color=red thickness=2);
   series x=weeknr y=prind3 / lineattrs=(color=green thickness=2);
   series x=weeknr y=prind4 / lineattrs=(color=orange thickness=2);
   series x=weeknr y=prind5 / lineattrs=(color=purple thickness=2);
   series x=weeknr y=prind8 / lineattrs=(color=pink thickness=2);
   xaxis label="Week Number";
   yaxis label="Price Index";
   title "Price Index Trends for Store #&shopnr";
run;

/* Sort the data by store number */
proc sort data=CHOC.SHAMPOO;
   by shopnr;
run;
/* Generate plots for each store */
```

```
proc sgplot data=CHOC.SHAMPOO;
   by shopnr; /* Separate plot for each store */
   series x=weeknr y=prom2_1/ lineattrs=(color=red thickness=2);
   series x=weeknr y=prom2_2 / lineattrs=(color=green thickness=2);
   series x=weeknr y=prom2_3 / lineattrs=(color=orange thickness=2);
   series x=weeknr y=prom2_4 / lineattrs=(color=purple thickness=2);
   series x=weeknr y=prom2_5 / lineattrs=(color=pink thickness=2);
   series x=weeknr y=prom2_6 / lineattrs=(color=pink thickness=2);
   series x=weeknr y=prom2_7 / lineattrs=(color=pink thickness=2);
   xaxis label="Week Number";
   yaxis label="Price Index";
   title "Price Index Trends for Store #&shopnr";
run;


* select the brand with the highest market share for modeling, and provide
descriptive labels for brand 2 and other related variables;
data WORK.SHAMPOO_imputed_data;
   set WORK.SHAMPOO_imputed_data;
   label
shopnr='store number'
weeknr='week number'
sales2='unit sales brand 2'
salml2='sales in ml for brand 2'
actpr2='actual unit price brand 2'
regpr2='regular (normal) price per unit brand 2'
prind2='price index of brand 2'
pci2='dummy for price discount for brand 2'
prom2='categorical promotion variabele for brand 2'
prom2_1='promotion 1 dummy for brand 2'
prom2_2='promotion 2 dummy for brand 2'
prom2_3='promotion 3 dummy for brand 2'
prom2_4='promotion 4 dummy for brand 2'
prom2_5='promotion 5 dummy for brand 2'
prom2_6='promotion 6 dummy for brand 2'
prom2_7='promotion 7 dummy for brand 2'
spec12='intensity of special pack 1 for brand2'
spec22='intensity of special pack 2 for brand2'
spec32='intensity of special pack 3 for brand2'
spec42='intensity of special pack 4 for brand2';
run;


* show the correlation;
proc corr data=CHOC.SHAMPOO pearson nosimple noprob plots=none outp=corr_out;
 var prind1-prind11;
proc print data=corr_out noobs;
run;

/* Create dummy variables for stores */
proc transreg data=SHAMPOO_imputed_data design;
   model class(shopnr / zero=none);
   output out=store_dummies(drop=_:);
run;
data SHAMPOO_Assign;
   merge SHAMPOO_imputed_data store_dummies;
```

```
run;

/* Create dummy variables for spec */
data WORK.SHAMPOO_Assign;
    set SHAMPOO_Assign;
    if spec12 = 9 then spec12_dummy = 1;
    else spec12_dummy = 0;
    if spec32 = 1 then spec32_dummy = 1;
    else spec32_dummy = 0;
    if spec42 = 9 then spec42_dummy = 1;
    else spec42_dummy = 0;
run;

/* Create LN */
data WORK.SHAMPOO_Assign;
    set SHAMPOO_Assign;
    sales2T = sales2 * actpr2;
    prind2_scaled = 1000 * prind2;
    LNactpr2=log(actpr2);
    LNregpr2= log(regpr2);
    LNsales2T = log(sales2 * actpr2);
    LNprind1= log(prind1);
    LNprind2= log(prind2);
    LNprind3= log(prind3);
    LNprind4= log(prind4);
    LNprind5= log(prind5);
    LNprind6= log(prind6);
    LNprind7= log(prind7);
    LNprind8= log(prind8);
    LNprind9= log(prind9);
    LNprind10= log(prind10);
    LNprind11= log(prind11);
run;

/* delete the outliers */
data WORK.SHAMPOO_Assign;
    set SHAMPOO_Assign;
    if sales5 not in (818, 782);
run;
```


Summary Statist...
HTML - Summary Stat...
Sales statistics
HTML - Summary Stat...
sales2 statistics
HTML - Summary Stat...
STORE_DUMMIES
Set up and progra...
SHAM POO
Query Builder 1
QUERY_FOR_SHA...
SalesT statistics
HTML - Summary Stat...
SHAM POO_IMPUT...
CORR_OUT
regpri summary
HTML - Summary Stat...
Sales2 vs Prind2...
HTML - Line Plot 8
all variales statisti...
HTML - Summary Stat...
Model 1.1
HTML - Linear Regre...
Model 1.2
HTML - Linear Regre...
Model 1.3
HTML - Linear Regre...
SHAM POO_ASSIGN
Model 2
HTML - Linear Regre...
Model 3
HTML - Linear Regre...
Yule Walker
HTML - Regression A...
Filter weeknr big th...
SHAM POO_VALIDA...
Linear Regression 2
HTML - Linear Regre...
Linear regression pr...
Scatter Plot
HTML - Scatter Plot
Filter for weekn...
SHAM POO_TRAIN
Query Builder 2
QUERY_FOR_SHA...
Scatter Plot 1
HTML - Scatter Plot 1