Final Exam
Impact of testing different landing pages

May, 2016

Xiaoxiao (Tracy) Liu

Background

X.com marketing director has consulted with a few agencies and discussed potential impact of building a few landing pages to maximize engagement, registrations and net sales on the retail website. As a consultant to help with the decision, I am mainly responsible for test designing along with the execution and measurement planning. The following 4 landing pages will be tested:

A – "On Sale" products without an offer
B – "On Sale" products with a strong offer: 30% off
C – Hot New Merchandise: New Arrivals with an offer 20% off
D – Hot New Merchandise: New Arrivals without an offer
E (Control) – Current front page

There are four profit drives available for testing which are audience, media, offer and message. Statistical tools such as T-Test, Simple Comparison Test and 2-level Full Factorial Design will be utilized to measure the testing results. The main methodologies include ANOVA, T-Test and F-Test.

Below are the assumptions:

1) All landing pages include registration bottom
2) All landing pages have a similar creative design
3) Average order revenue = $100
4) Landing page subsequent click rate off the current front page = 1.1%, and in external media we don't mention any promotions or discounts.
5) COGS = 30% of Gross Revenue
6) Production cost of 1 landing page = $1200

Definitions:

1) Engagement is defined as number of pages per visit.

2) Registrations are defined as number of visitors who registered but not purchased.

3) Net Sales is total revenue minus (COGS, total offer costs and promotion costs).

Test Strategy, design and implementation

The ***objective*** of the testing is to maximize the engagement, registration and net sales of the retail website from June 20th to July 31th by testing the newly launched landing pages.

Here is the layout of the four ***profit drivers*** and their layers.

Message:

Although the texts listed below are not the final version on the site, they are different in tones and wording.

Message A: "On Sale" Products without an offer

Message B: "On Sale" Products with 30% off

Message C: Hot New Merchandise: New Arrivals with 20% off

Message D: Hot New Merchandise: New Arrivals without an offer

Message E: Current version.

Audience:

Audience are segmented into new customers who come to the website for the first time and returning customers who have visited the website before. Returning customers may have or have not registered.

New Customers

Returning Customers

Offer:

No Discount

20% Off Discount

30% Off Discount

Media:

The links of the new landing pages will reach our customers through the following channels:

Email

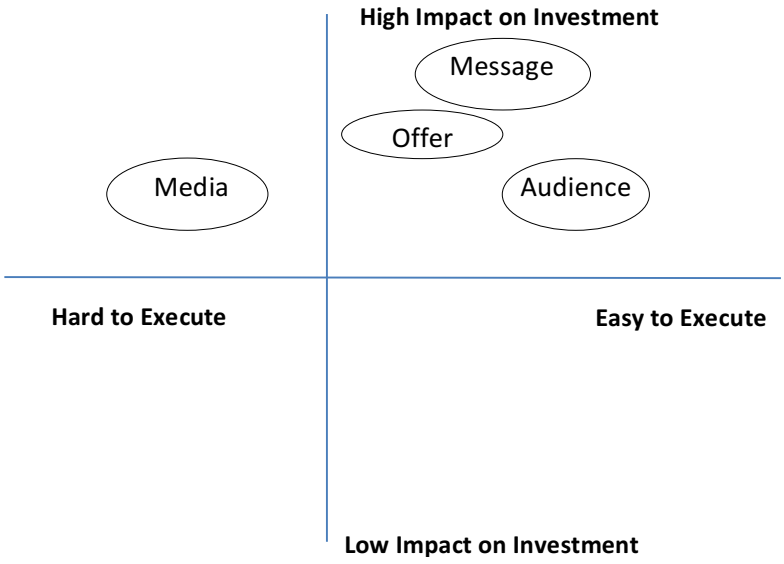
Social Media

Website

KPI

The primary KPIs are number of pages per visit (average), registrations and net revenues (revenues minus GOS, production costs and other related costs). Secondary KPIs include impressions, click rate, clickers, response rate and orders.

Qualitative Assessment of Feasibility



Since all of the factors have high impact on investment, all of them will be reserved.

Test Design:

Topline Summary of Test Design			Test A	Test B	Test C	Test D	Test E	Total
Audience	Media	Offer	Message A	Message B	Message C	Message D	Current	
New Customers	Social Media	No Discount	A	D	H	K	N	Q
New Customers	Social Media	20% Off	B	E	I	L	O	R
New Customers	Social Media	30% Off	C	G	J	M	P	S
Total EXP New Customers Social Media ONLY Offer Test			NSMSUMA	NSMSUMB	NSMSUMC	NSMSUMD	NSMSUME	NSMSUMT
New Customers	Email	No Discount	A1	D1	H1	K1	N1	Q1
New Customers	Email	20% Off	B1	E1	I1	L1	O1	R1
New Customers	Email	30% Off	C1	G1	J1	M1	P1	S1
Total EXP New Customers Email ONLY Offer Test			NESUMA	NESUMB	NESUMC	NESUMD	NESUME	NESUMT
New Customers	Website	No Discount	A2	D2	H2	K2	N2	Q2
New Customers	Website	20% Off	B2	E2	I2	L2	O2	R2
New Customers	Website	30% Off	C2	G2	J2	M2	P2	S2
Total EXP New Customers Website ONLY Offer Test			NWSUMA	NWSUMB	NWSUMC	NWSUMD	NWSUME	NWSUMT
Total EXP New Customers ONLY Media and Offer Test			NMOSUMA	NMOSUMB	NMOSUMC	NMOSUMD	NMOSUME	NMOSUMT
Returning Customers	Social Media	No Discount	A3	D3	H3	K3	N3	Q3
Returning Customers	Social Media	20% Off	B3	E3	I3	L3	O3	R3
Returning Customers	Social Media	30% Off	C3	G3	J3	M3	P3	S3
Total EXP Returning Customers Social Media ONLY Offer Test			RSMSUMA	RSMSUMB	RSMSUMC	RSMSUMD	RSMSUME	RSUMT
Returning Customers	Email	No Discount	A4	D4	H4	K4	N4	Q4
Returning Customers	Email	20% Off	B4	E4	I4	L4	O4	R4
Returning Customers	Email	30% Off	C4	G4	J4	M4	P4	S4
Total EXP Returning Customers Email ONLY Offer Test			RESUMA	RESUMB	RESUMC	RESUMD	RESUME	RESUMT
Returning Customers	Website	No Discount	A5	D5	H5	K5	N5	Q5
Returning Customers	Website	20% Off	B5	E5	I5	L5	O5	R5
Returning Customers	Website	30% Off	C5	G5	J5	M5	P5	S5
Total EXP Returning Customers Website ONLY Offer Test			RWSUMA	RWSUMB	RWSUMC	RWSUMD	RWSUME	RWSUMT
Total EXP Returning Customers ONLY Media and Offer Test			RMOSUMA	RMOSUMB	RMOSUMC	RMOSUMD	RMOSUME	RMOSUMT

Hypothesis 1: Audience

H0: NMOSUMT = RMOSUMT
H1: NMOSUMT ≠RMOSUMT

Hypothesis 2: Media

H0: NSMSUMT + RSUMT = NESUMT + RESUMT = NWSUMT + RWSUMT
H1: NSMSUMT + RSUMT ≠ NESUMT + RESUMT ≠ NWSUMT + RWSUMT

Hypothesis 3: Offer

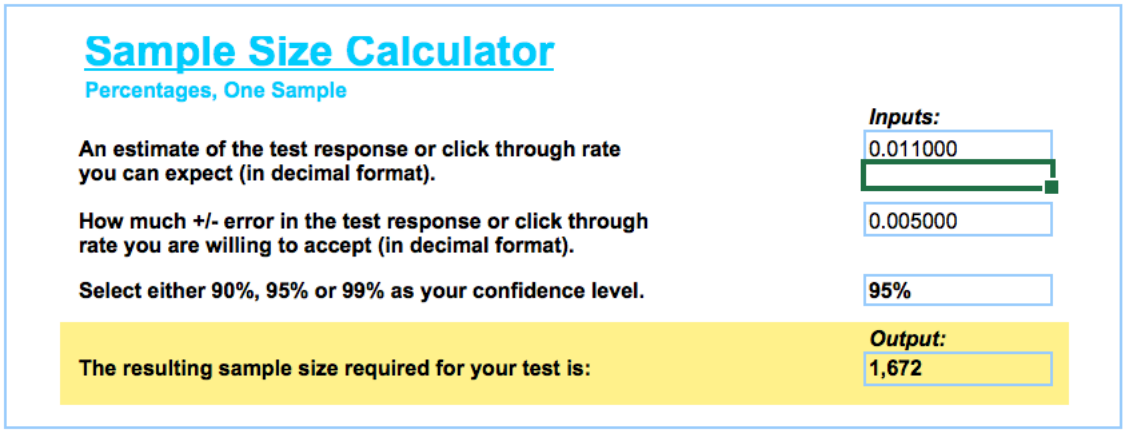
H0: Q+Q1+Q2+Q3+Q4+Q5 = R+R1+R2+R3+R4+R5 = S+S1+S2+S3+S4+S5
H1: Q+Q1+Q2+Q3+Q4+Q5 ≠ R+R1+R2+R3+R4+R5≠S+S1+S2+S3+S4+S5

Hypothesis 4: Message

H0: NMOSUMA+RMOSUMA = NMOSUMB+RMOSUMB = NMOSUMC+RMOSUMC = NMOSUMD+RMOSUMD= NMOSUME+RMOSUME
H1: NMOSUMA+RMOSUMA ≠ NMOSUMB+RMOSUMB ≠ NMOSUMC+RMOSUMC ≠NMOSUMD+RMOSUMD ≠ NMOSUME+RMOSUME

Test Administration and Implementation:

The test will be conducted at 95% confidence level. Based on a click rate of 1.1%, the resulted sample size is 1672 with a confidence level of 95%.

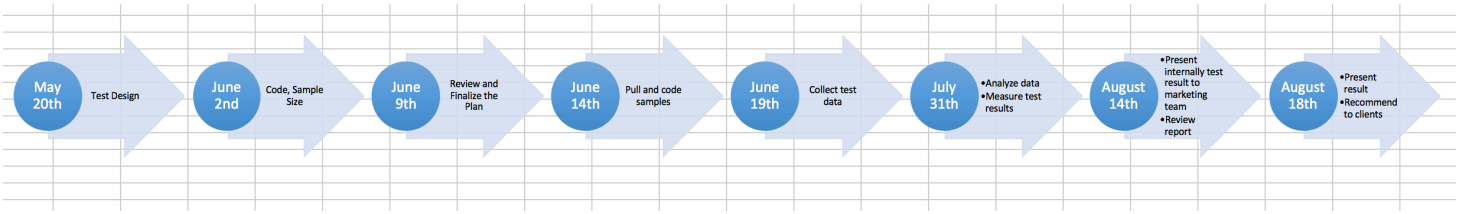


With a cushion of 10% added, the minimum sample size will be:

1672 ×(1+10%) = 1839.2 = 1840

For simplicity, all of the tests will use the same sample size of 1840.

Timeline of the testing:



The test will start from May 20th and last until August 18th when we present to our clients.

Organizational coordination and collaboration

Date	Task	Staff
5/20/16	Test Design	Tracy
6/2/16	Code and Decide Sample Size	Paul
6/9/16	Review and Finalize the Plan	Tracy
6/14/16	Pull and Code Samples	Paul
6/19/16	Collect Test Data	Steven
7/31/16	Analyze Data and Measure Test Results	Steven
8/14/16	Present internally test result to the marketing team and review the report	Tracy
8/18/16	Present results and make recommendations to our client	Tracy,Steven and Paul

The whole process has three segments. The first part of the process is the preparation which include test design, sample coding, plan review and finalization. Tracy and Paul will be responsible in the part. The second segment is the data collection and measurement. As the data scientist, Steven will be in charge of the process. The third segment is to write report and make recommendations to our client. Tracy, Steven and Paul will participate in the process.

Data collection

The data will be collected through online tracking. After the new landing pages are launched, we will start to collect data of our website performance in real time. Potential customers will be selected on a random basis. Emails will be sent out to a random group of people. Social media advertisements or activities will appear in front of another group of people. The four new landing pages will appear differently for people in different samples.

Proforma P&L

Media

Social media works best in registration while customers driven to the website directly has the most average number of pages per visit. To sum it, website has the largest net revenues compared with social media and email channel. A possible explanation to the result may be that people who go to the website directly or through search engine have more interests to the brand than people who are driven to the website by advertisement or information in emails and social media.

	Control	Test 2	Test1		
	H0	H1	H2	Difference 1	Difference 2
What do we test?	**Website**	**Social Media**	**Email**		
Circulation/Impressions	50,000	50,000	50,000		
Number of Pages Per Visit (Average)	3	2	2	1	1
Click Rate	1.10%	1.10%	1.10%	0.00%	0.00%
Clickers	275		275	-	275
Response Rate	2.50%	3.00%	2.45%	0.05%	2.40%
Registrations	1250	1550	1225	25	1,200
Orders	1300	1210	1080	220	860
Total Gross Revenue	$130,000	$121,000	$108,000	$ 22,000	$ 22,000
Avg. Gross Rev./Order	$100.00	$100.00	$100.00	$0.0	$100.0
COGS @ 30%	$39,000.00	$36,300.00	$32,400.00	$2,700	$6,600
Total Margin $$	$91,000.00	$84,700.00	$75,600.00	$6,300	$15,400
Production Cost of 1 Landing Page	$1,200	$1,200	$1,200	$0	$0
Total Net Margin $$ (Net of Offer, COGS and Promotion Costs)	$89,800.00	$83,500.00	$74,400.00	$6,300	$15,400

Audience

Returning customers outperform new customers in all aspects which is within our expectation.

	Control	Test1	
	H0	H1	Difference
What do we test?	**Returning Customers**	**New Customers**	
Circulation/Impressions	50,000	50,000	
Number of Pages Per Visit (Average)	3	1	2
Click Rate	1.10%	1.10%	0.00%
Clickers	275	275	-
Response Rate	2.50%	2.45%	0.05%
Registrations	1250	1225	25
Orders	1400	1364	36
Total Gross Revenue	$140,000	$136,400	$ 3,600
Avg. Gross Rev./Order	$100.00	$100.00	$0.0
COGS @ 30%	$42,000.00	$40,920.00	$1,080
Total Margin $$	$98,000.00	$95,480.00	$2,520
Production Cost of 1 Landing Page	$1,200	$1,200	$0
Total Net Margin $$ (Net of Offer, COGS and Promotion Costs)	$96,800.00	$94,280.00	$2,520

Offer

The strongest offer has the best results compared to no discount and 20% off. However, the difference between 20% off and 30% off is still to be tested to see if the difference is statistically significant.

	Control	Test 2	Test1		
	H0	H1	H2	Difference 1	Difference 2
What do we test?	**No Discount**	**20% Off**	**30% Off**		
Circulation/Impressions	50,000	50,000	50,000		
Number of Pages Per Visit (Average)	1	3	4	-2	-3
Click Rate	1.10%	1.10%	1.10%	0.00%	0.00%
Clickers	275	275	275	-	-
Response Rate	2.00%	2.45%	3.00%	-0.45%	-1.00%
Registrations	1000	1225	1500	(225)	(500)
Orders	900	1570	1789	(670)	(889)
Total Gross Revenue	$90,000	$157,000	$178,900	$ (67,000)	$ (88,900)
Avg. Gross Rev./Order	$100.00	$100.00	$100.00	$0.0	$0.0
COGS @ 30%	$27,000.00	$47,100.00	$53,670.00	($20,100)	($26,670)
Total Margin $$	$63,000.00	$109,900.00	$125,230.00	($46,900)	($62,230)
Production Cost of 1 Landing Page	$1,200	$1,200	$1,200	$0	$0
Total Net Margin $$ (Net of Offer, COGS and Promotion Costs)	$61,800.00	$108,700.00	$124,030.00	($46,900)	($62,230)

Message

"Hot new merchandise products with 20% off" has the largest number of pages per visit while "on sale products with 30% off " brings the largest registrations to the website. If we have to choose between the two options, "on sale products with 30% off" is the most profitable based on the net revenues.

	Control	Test 1	Test 2	Test 3	Test 4				
	H0	H1	H2	H3	H4	Difference 1	Difference 2	Difference 3	Difference 4
What do we test?	**Current**	**Message A**	**Message B**	**Message C**	**Message D**				
Circulation/Impressions	50,000	50,000	50,000	50,000	50,000				
Number of Pages Per Visit (Average)	1	2	4	5	3	-1	-3	-4	-2
Click Rate	1.10%	1.10%	1.10%	1.10%	1.10%	0.00%	0.00%	0.00%	0.00%
Clickers	275	275	275	275	275	-	-	-	-
Response Rate	2.00%	2.05%	3.00%	2.70%	2.50%	-0.05%	-1.00%	-0.70%	-0.50%
Registrations	1000	1025	2000	1980	1250	(25)	(1,000)	(980)	(250)
Orders	900	1120	2320	2010	1789	(220)	(1,420)	(1,110)	(889)
Total Gross Revenue	$90,000	$112,000	$232,000	$201,000	$178,900	$ (22,000)	$ (142,000)	$ (111,000)	$ (88,900)
Avg. Gross Rev./Order	$100.00	$100.00	$100.00	$100.00	$100.00	$0.0	$0.0	$0.0	$0.0
COGS @ 30%	$27,000.00	$33,600.00	$69,600.00	$60,300.00	$53,670.00	($6,600)	($42,600)	($33,300)	($26,670)
Total Margin $$	$63,000.00	$78,400.00	$162,400.00	$140,700.00	$125,230.00	($15,400)	($99,400)	($77,700)	($62,230)
Production Cost of 1 Landing Page	$1,200	$1,200	$1,200	$1,200	$1,200	$0	$0	$0	$0
Total Net Margin $$ (Net of Offer, COGS and Promotion Costs)	$61,800.00	$77,200.00	$161,200.00	$139,500.00	$124,030.00	($15,400)	($99,400)	($77,700)	($62,230)

Measurement methodology

To measure the results, different statistical tools will be used based on the specific situations.

For media, I will use ***t-testing*** to see if the difference between among website, social media and email is significant.

A sample result will be as following:

Engagement	Website	Social Media	Email
Week 1	3	2	2
Week 2	3	1	1
Week 3	2	3	3
Week 4	4	2	2
Week 5	3	2	2
Mean	3	2	2

Engagement

t-Test: Two-Sample Assuming Equal Variances

	Website	*Social Media*
Mean	3	2
Variance	0.5	0.5
Observations	5	5
Pooled Variance	0.5	
Hypothesized Me	0	
df	8	
t Stat	2.236067977	
P(T<=t) one-tail	0.027883264	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.055766529	
t Critical two-tail	2.306004135	

t-Test: Two-Sample Assuming Equal Variances

	Social Media	*Email*
Mean	2	2
Variance	0.5	0.5
Observations	5	5
Pooled Variance	0.5	
Hypothesized Mean I	0	
df	8	
t Stat	0	
P(T<=t) one-tail	0.5	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	1	
t Critical two-tail	2.306004135	

1. t.stat < t critical, accept null hypothesis. There is no significant diffence between website and social media.
2. Social media and email are totally the same.

   From the measurement above, it can be concluded that all of those three channels should be used in the campaign since all of them are of equal importance to engagement.

Registration	Website	Social Media	Email
Week 1	1243	1590	1299
Week 2	1220	1409	1245
Week 3	1260	1600	1119
Week 4	1230	1500	1254
Week 5	1297	1651	1208
Mean	1250	1550	1225

Registration

t-Test: Two-Sample Assuming Equal Variances

	Website	*Social Media*
Mean	1250	1550
Variance	914.5	9170.5
Observations	5	5
Pooled Variance	5042.5	
Hypothesized Me	0	
df	8	
t Stat	-6.679874538	
P(T<=t) one-tail	7.79557E-05	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.000155911	
t Critical two-tail	2.306004135	

t-Test: Two-Sample Assuming Equal Variances

	Social Media	*Email*
Mean	1550	1225
Variance	9170.5	4560.5
Observations	5	5
Pooled Variance	6865.5	
Hypothesized Mean I	0	
df	8	
t Stat	6.201793187	
P(T<=t) one-tail	0.000129461	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.000258922	
t Critical two-tail	2.306004135	

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1250	1225
Variance	914.5	4560.5
Observations	5	5
Pooled Variance	2737.5	
Hypothesized Mea	0	
df	8	
t Stat	0.7554974	
P(T<=t) one-tail	0.2358032	
t Critical one-tail	1.859548	
P(T<=t) two-tail	0.4716064	
t Critical two-tail	2.3060041	

1. t Stat < t Critical (2.306) on the left side , reject null hypothesis. There is significant difference between website and social media in registration.
2. t Stat > t Critical (2.306) , reject null hypothesis. There is significant difference between social media and email in registration.
3. t Stat < t Critical, accept null hypothesis. There is no significant difference between website and email.

To improve registration, social media is significant better than website and email. There is no difference between website and email.

Net Revenue	Website	Social Media	Email
Week 1	88765	76899	86098
Week 2	88765	78654	85779
Week 3	96543	78754	70864
Week 4	109870	89765	66543
Week 5	65057	93428	62716
Mean	89800	83500	74400

Net Revenue

t-Test: Two-Sample Assuming Equal Variances

	Website	*Social Media*
Mean	89800	83500
Variance	265657862	56849210.5
Observations	5	5
Pooled Variance	161253536.3	
Hypothesized Me	0	
df	8	
t Stat	0.784433135	
P(T<=t) one-tail	0.227691873	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.455383745	
t Critical two-tail	2.306004135	

t-Test: Two-Sample Assuming Equal Variances

	Social Media	*Email*
Mean	83500	74400
Variance	56849210.5	119269111.5
Observations	5	5
Pooled Variance	88059161	
Hypothesized Mean I	0	
df	8	
t Stat	1.533289371	
P(T<=t) one-tail	0.081873058	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.163746116	
t Critical two-tail	2.306004135	

t-Test: Two-Sample Assuming Equal Variances

	Website	*Email*
Mean	89800	74400
Variance	265657862	119269111.5
Observations	5	5
Pooled Variance	192463487	
Hypothesized Mea	0	
df	8	
t Stat	1.7551593	
P(T<=t) one-tail	0.0586541	
t Critical one-tail	1.859548	
P(T<=t) two-tail	0.1173082	
t Critical two-tail	2.3060041	

1. t Stat < t Critical Accept null hypothesis. There is no significant difference between website and social media in increasing revenues.
2. t Stat < t Critical Accept null hypothesis. There is no significant difference between Email and social media in increasing revenues.
3. t Stat < t Critical Accept null hypothesis. There is no significant difference between website and Email in increasing revenues.

In revenues, all of the three channels are of equal importance.

Recommendation 1:

Utilize all of the three channels which are website, social media and email.

Allocate more budget to social media in increasing registrations.

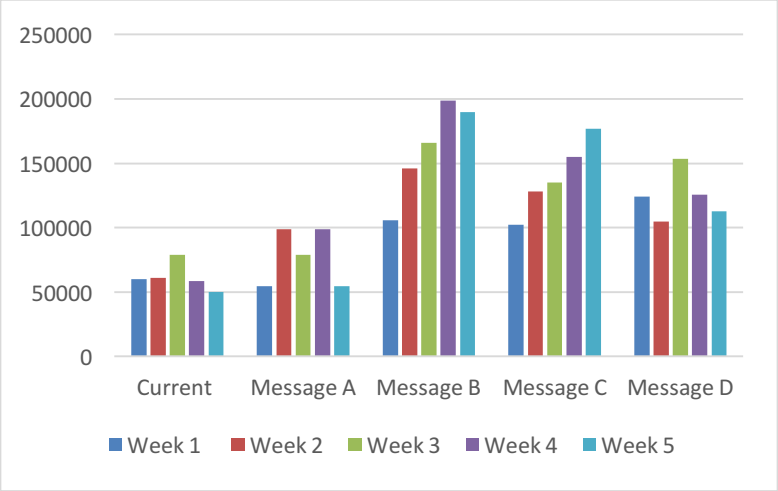
The most important goal of message is net revenue. Therefore, I will use simple comparison to measure the net revenues of Message.

Net Revenue

	Current	Message A	Message B	Message C	Message D
Week 1	59870	54739	105947	102394	123987
Week 2	61234	98760	145980	128340	104863
Week 3	78912	78903	165907	134895	153284
Week 4	58765	98765	198563	154938	125439
Week 5	50219	54833	189603	176933	112577
Mean	61800	77200	161200	139500	124030

1. Check if this is a simple comparison test.

There seems to be a pattern in the variable of “Week”.



Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Row 1	5	446937	89387.4	928078210
Row 2	5	539177	107835.4	1034584415
Row 3	5	611901	122380.2	1696499681
Row 4	5	636470	127294	2839257451
Row 5	5	584165	116833	4301755358

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4454519345	4	1113629836	0.51556101	0.725146084	2.866081402
Within Groups	4.3201E+10	20	2160035023			
Total	4.7655E+10	24				

There is no significant difference among weeks.

Multiple Comparisons

Dependent Variable: Revenue

LSD

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Week 1	Week 2	-18448.00000	29394.11521	.537	-79763.0499	42867.0499
	Week 3	-32992.80000	29394.11521	.275	-94307.8499	28322.2499
	Week 4	-37906.60000	29394.11521	.212	-99221.6499	23408.4499
	Week 5	-27445.60000	29394.11521	.362	-88760.6499	33869.4499
Week 2	Week 1	18448.00000	29394.11521	.537	-42867.0499	79763.0499
	Week 3	-14544.80000	29394.11521	.626	-75859.8499	46770.2499
	Week 4	-19458.60000	29394.11521	.516	-80773.6499	41856.4499
	Week 5	-8997.60000	29394.11521	.763	-70312.6499	52317.4499
Week 3	Week 1	32992.80000	29394.11521	.275	-28322.2499	94307.8499
	Week 2	14544.80000	29394.11521	.626	-46770.2499	75859.8499
	Week 4	-4913.80000	29394.11521	.869	-66228.8499	56401.2499
	Week 5	5547.20000	29394.11521	.852	-55767.8499	66862.2499
Week 4	Week 1	37906.60000	29394.11521	.212	-23408.4499	99221.6499
	Week 2	19458.60000	29394.11521	.516	-41856.4499	80773.6499
	Week 3	4913.80000	29394.11521	.869	-56401.2499	66228.8499
	Week 5	10461.00000	29394.11521	.726	-50854.0499	71776.0499
Week 5	Week 1	27445.60000	29394.11521	.362	-33869.4499	88760.6499
	Week 2	8997.60000	29394.11521	.763	-52317.4499	70312.6499
	Week 3	-5547.20000	29394.11521	.852	-66862.2499	55767.8499
	Week 4	-10461.00000	29394.11521	.726	-71776.0499	50854.0499

There is no significant difference between weeks. This is a simple comparison test.

2. Descriptive Analysis

Current		*Message A*		*Message B*		*Message C*		*Message D*	
Mean	61800	Mean	77200	Mean	161200	Mean	139500	Mean	124030
Standard Error	4691.46345	Standard Error	9842.66794	Standard Error	16597.32273	Standard Error	12574.11666	Standard Error	8235.0736
Median	59870	Median	78903	Median	165907	Median	134895	Median	123987
Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard Deviation	10490.43119	Standard Deviation	22008.8746	Standard Deviation	37112.74187	Standard Deviation	28116.5796	Standard Deviation	18414.1844
Sample Variance	110049146.5	Sample Variance	484390561	Sample Variance	1377355609	Sample Variance	790542048.5	Sample Variance	339082186
Kurtosis	2.71673901	Kurtosis	-3.0073939	Kurtosis	-0.110410122	Kurtosis	-0.135579953	Kurtosis	1.61674035
Skewness	1.235126949	Skewness	-0.0963718	Skewness	-0.790914581	Skewness	0.066629987	Skewness	1.10070438
Range	28693	Range	44026	Range	92616	Range	74539	Range	48421
Minimum	50219	Minimum	54739	Minimum	105947	Minimum	102394	Minimum	104863
Maximum	78912	Maximum	98765	Maximum	198563	Maximum	176933	Maximum	153284
Sum	309000	Sum	386000	Sum	806000	Sum	697500	Sum	620150
Count	5	Count	5	Count	5	Count	5	Count	5

3. One Factor Plot

	Current	*Message A*	*Message B*	*Message C*	*Message D*
Max	78912	98765	198563	176933	153284
	61234	98760	189603	154938	125439
Median	59870	78903	165907	134895	123987
	58765	54833	145980	128340	112577
Min	50219	54739	105947	105947	104863
Range	28693	44026	92616	74539	48421
AVG	61800	77200	161200	139500	124030

4. ANOVA and LSD

ANOVA

Revenue

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	42938719600.000	4	10734679900.000	14.553	.000
Within Groups	14752932446.000	20	737646622.300		
Total	57691652046.000	24			

Significance is below 5%. There is significant difference among different messages. Further conclusion will be conducted through LSD.

Multiple Comparisons

Dependent Variable: Revenue

LSD

(I) LP	(J) LP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Current	A	-15400.00000	17177.27129	.381	-51231.1600	20431.1600
	B	-99400.00000*	17177.27129	.000	-135231.1600	-63568.8400
	C	-99400.00000*	17177.27129	.000	-135231.1600	-63568.8400
	D	-62230.00000*	17177.27129	.002	-98061.1600	-26398.8400
A	Current	15400.00000	17177.27129	.381	-20431.1600	51231.1600
	B	-84000.00000*	17177.27129	.000	-119831.1600	-48168.8400
	C	-84000.00000*	17177.27129	.000	-119831.1600	-48168.8400
	D	-46830.00000*	17177.27129	.013	-82661.1600	-10998.8400
B	Current	99400.00000*	17177.27129	.000	63568.8400	135231.1600
	A	84000.00000*	17177.27129	.000	48168.8400	119831.1600
	C	.00000	17177.27129	1.000	-35831.1600	35831.1600
	D	37170.00000*	17177.27129	.043	1338.8400	73001.1600
C	Current	99400.00000*	17177.27129	.000	63568.8400	135231.1600
	A	84000.00000*	17177.27129	.000	48168.8400	119831.1600
	B	.00000	17177.27129	1.000	-35831.1600	35831.1600
	D	37170.00000*	17177.27129	.043	1338.8400	73001.1600
D	Current	62230.00000*	17177.27129	.002	26398.8400	98061.1600
	A	46830.00000*	17177.27129	.013	10998.8400	82661.1600
	B	-37170.00000*	17177.27129	.043	-73001.1600	-1338.8400
	C	-37170.00000*	17177.27129	.043	-73001.1600	-1338.8400

*. The mean difference is significant at the 0.05 level.

Conclusion: The differences between messages are significant except for Current and Message A and Message B with Message C. Message B and Message C are recognized as the same while Message A does not make a change from the current version in net revenues.

Recommendation 2:

Among all of the landing pages, Message A should be abandoned. Message B is the most ideal option in terms of net revenues.

For **Audience** and **Offer**, I will conduct a **2-level Full Factorial Design**.

Although there are three factors in the Offer Category, I would like to segment it into two: discount offer and no discount offer to fit in the 2-level full factorial design.

I select two effects which are registration and number of pages per visit.

Here is the test process.

Factors	Name	Measurement Units	Level 1 (-)	Level 2 (+)	
1	Offer	Categorical	No Discount	Discount	
2	Audience	Categorical	Returning Customer	New Customer	
Dependent Variables:		Registrations	Registration	Y1	
		Engagement	Number of Pages per Visit	Y2	

Stand. Order	Experiment Order	Offer	Audience	Registrations	Engagement
2	1	+	-	1306	3.25
3	2	-	-	1265	3
1	3	-	+	1113	1
4	4	+	+	1294	2.25
Effect Y1		111	82		
Effect Y2		0.75	1.5		

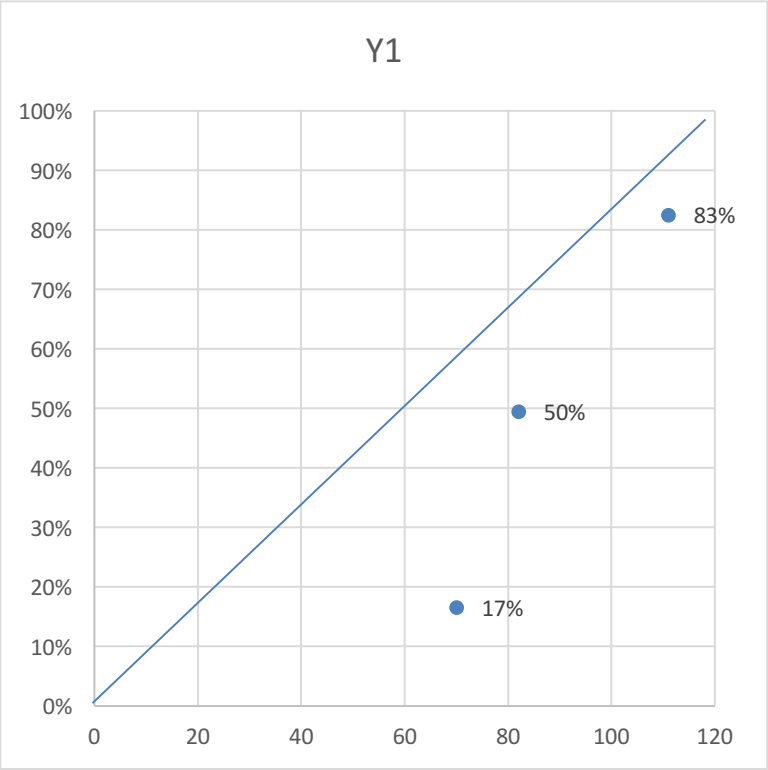
Stand. Order	Experiment Order	Offer	Audience	Registrations	Engagement
1	3	-	+	1113	1
2	1	+	-	1306	3.25
3	2	-	-	1265	3
4	4	+	+	1294	2.25
Effect Y1		-111	82		
Effect Y2		-0.75	1.5		

		Main Effect		Interaction Effect		
Stand. Order	Experiment Order	Offer	Audience	OxA	Registration	Engagement
1	3	-	+	-	1113	1
2	1	+	-	-	1306	3.25
3	2	-	-	+	1265	3
4	4	+	+	+	1294	2.25
Effect Y1		-111	82	-70		
Effect Y2		-0.75	1.5	-0.5		

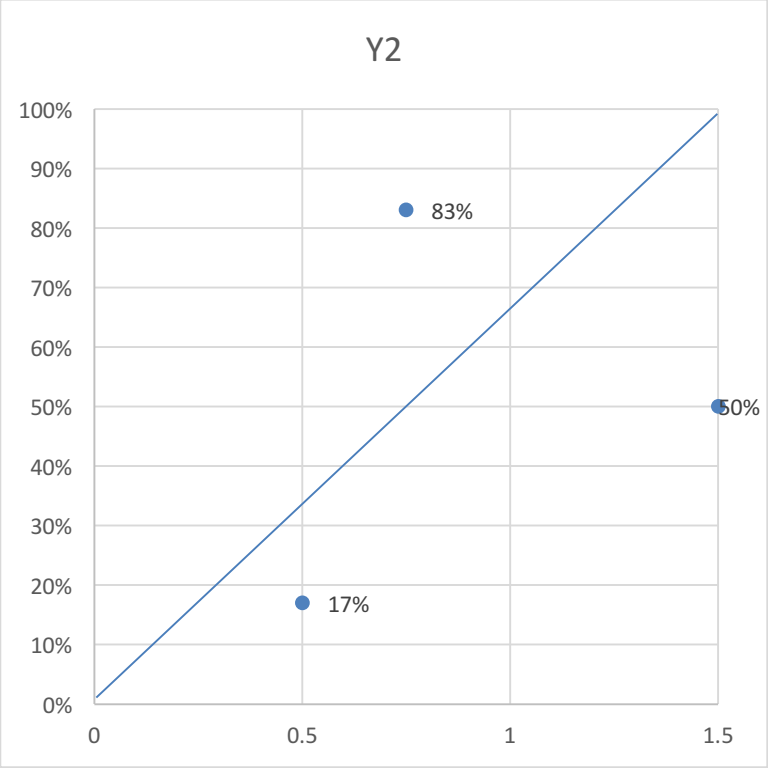
		Registration (Y1)	Engagement (Y2)
Offer	No Discount (-)	1189	2
	Discount Offer (+)	1300	2.75
Audience	Returning Customers (-)	1285.5	3.125
	New Customer (+)	1203.5	1.625
	no discount to new customers or discount offer to returning customers (-)	1209.5	2.125
	no discount to returning customers or discount offer to new customers (+)	1279.5	2.625
Conclusion		Discout offer outperformed no discount option in registrations	Disount offer outperformed no discount option in engagement
		Targeting returning customers is more efficient in increasing registrations	Targeting returning customers is more efficient in increasing engagement.
		Providing returning customers with no discount offer or providing discount offer to new custmers work better.	

Graph

Point	Effect	Y1 Value	Y1 ABS	Cum (P)	Point	Effect	Y2 Value	Y2 ABS	Cum (P)
1	OxA	-70	70	17%	1	OxA	-0.5	0.5	17%
2	Audience	82	82	50%	2	Audience	1.5	1.5	50%
3	Offer	-111	111	83%	3	Offer	-0.75	0.75	83%



In terms of registration, the difference of all of the three factors are significant.



The difference of Offer is not significant in engagement but audience and the interaction effect are all significant.

F-Test

ANOVA Significance Testing and Validation - Y1:						
	SS	Df	MS	F Value	Sign (P) Value @ 1%	
Model	38090	1	38090	7.8	insignificant	13.2
Audience	13448	1	13448	2.7	insignificant	16.3
Offer	24642	1	24642	5.0	insignificant	16.3
Residual	9800	2	4900			
Total	47890	3				

ANOVA Significance Testing and Validation - Y2:						
	SS	Df	MS	F Value	Sign (P) Value @ 1%	
Model	5.625	1	5.625	22.5	significant	13.2
Audience	4.5	1	4.5	18.0	significant	16.3
Offer	1.125	1	1.125	4.5	insignificant	16.3
Residual	0.5	2	0.25			
Total	6.125	3				

Here is the next treatment.

Registration	Audience(F1)	Insignificant	Simple Comparison
	Offer(F2)	Insignificant	
Engagement	Audience(F1)	Significant	Randomized Block
	Offer(F2)	Significant	

From the F test result, we learn that for Effect 1 – Registration – the two factors are not significant in difference. They do not have an influence on registration. For Effect 2 – Engagement – the difference in levels of each of the factors is significant compared with other effects.

Audience

Registration	level (-)	level(+)
offer(-)	1265	1113
offer(+)	1306	1294

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	2	2571	1285.5	840.5
Column 2	2	2407	1203.5	16380.5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6724	1	6724	0.780907032	0.4700847	18.512821
Within Groups	17221	2	8610.5			
Total	23945	3				

Offer	Registration	level (-)	level(+)
	Audience(-)	1265	1306
	Audience(+)	1113	1294

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	2	2378	1189	11552
Column 2	2	2600	1300	72

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12321	1	12321	2.119924295	0.2826757	18.512821
Within Groups	11624	2	5812			
Total	23945	3				

From the simple comparison result above for the first effect, we learn that the difference of each factor is not significant enough.

Recommendation 3:

In increasing registrations, there is no difference in discount offer or no discount offer. Also, there is no difference in selecting between return customers and new customers.

Recommendation 4:

In increasing engagement, there is significant difference in discount offer and no discount offer. Also, there is difference in selecting between return customers and new customers. The data above shows that we should choose the following tactics:

1. **Using discount offer**
2. **Targeting returning customers**
3. **Providing discount offer to new customers**
4. **Providing no discount offer to returning customers**

Summary and Recommendation:

From the test above, I tested four factors which are : audience, media, message and offer. Suggested by the results, I have the following recommendations:

1. **To improve engagement, all of the three channels (social media, email and website) should be considered as equal important. Discount offer and using discount offer to new customers work better. Targeting returning customers is more efficient. For returning customers, it is also feasible if we send them no discount offer.**

2. **To improve registrations, social media is the most effective media channel. There is no difference in terms of which offer is provided or to which segment of customers.**

3. **To improve net revenues, all of the three media channels should be utilized to have the best result. "On Sale Product without an Offer" is the message which we should abandon. The message of " On Sale Product with a Strong Offer: 30% Off" is the most effective message.**

Part 2:

Simple Comparative Experiments:
Objective: Determine overall the "best" of LP of maximized registrations in x.com
Data Set

Registrations		On Sale	OnSale 30% off	New Arrivals 20% off	Hot New Merch	Control
Social sites	Facebook	254	324	282	325	329
	Twitter	263	301	354	325	288
	LinkedIn	459	548	496	288	265
	Retellity	256	325	302	289	299
	Reddit	247	312	287	309	327
	Total	1479	1810	1721	1536	1508

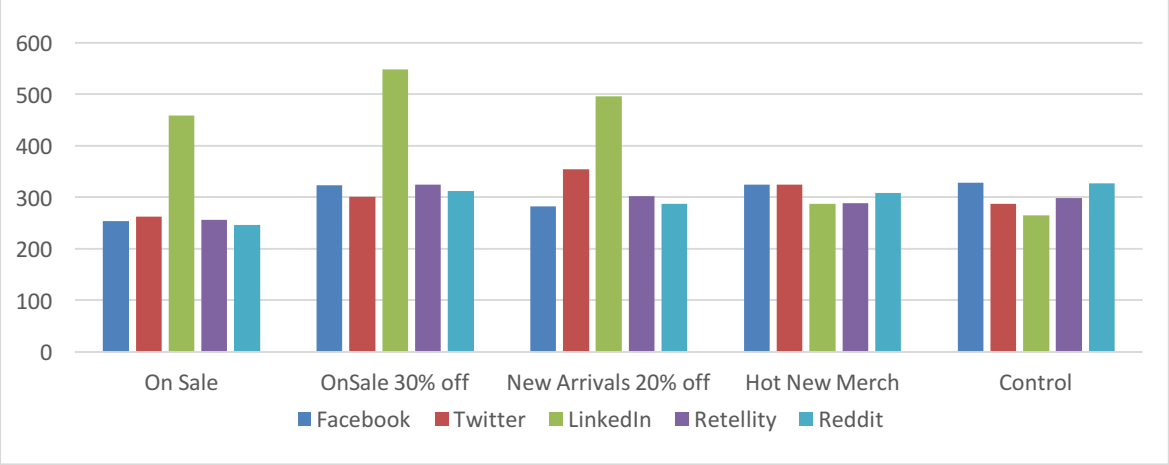
Questions: Is is Simple Comparison or Randomized Blocks?

1. Calculate Averages, Medians and SD for registrations by LP
2. Do the LPs have a different or the same impact on registrations? (use ANOVA/Excel)
3. Draw a one factor plot
4. Calculate LSD
5. Run a t-test to show if there is significant difference between specific LPs AT 95% confidence level.

Output 1: It is a Randomized Block. All the calculations below are based on the dataset which are neutralized.

- **Elementary Observation - Graph**

Linkedin is the highest in the first three while the smallest in the last two. There seems to be a pattern in social sites` performance.



- **ANOVA Based on Social Sites**

Significance (0.038) < 0.05, Reject null hypothesis, there is significant difference among social sites.

ANOVA					
Registration					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	50014.96	4	12503.74	3.109	0.038
Within Groups	80434.4	20	4021.72		
Total	130449.36	24			

- **A Deeper Exploration with LSD**

There is a significant difference between Linkedin with other variables.

LSD						
Multiple Comparisons						
Dependent Variable: Registration						
LSD						
(I) Social	(J) Social	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Facebook	Twitter	-3.4	40.10845	0.933	-87.0648	80.2648
	Linkedin	-108.40000*	40.10845	0.014	-192.0648	-24.7352
	Retellity	8.6	40.10845	0.832	-75.0648	92.2648
	Reddit	6.4	40.10845	0.875	-77.2648	90.0648
Twitter	Facebook	3.4	40.10845	0.933	-80.2648	87.0648
	Linkedin	-105.00000*	40.10845	0.016	-188.6648	-21.3352
	Retellity	12	40.10845	0.768	-71.6648	95.6648
	Reddit	9.8	40.10845	0.809	-73.8648	93.4648
Linkedin	Facebook	108.40000*	40.10845	0.014	24.7352	192.0648
	Twitter	105.00000*	40.10845	0.016	21.3352	188.6648
	Retellity	117.00000*	40.10845	0.009	33.3352	200.6648
	Reddit	114.80000*	40.10845	0.01	31.1352	198.4648
Retellity	Facebook	-8.6	40.10845	0.832	-92.2648	75.0648
	Twitter	-12	40.10845	0.768	-95.6648	71.6648
	Linkedin	-117.00000*	40.10845	0.009	-200.6648	-33.3352
	Reddit	-2.2	40.10845	0.957	-85.8648	81.4648
Reddit	Facebook	-6.4	40.10845	0.875	-90.0648	77.2648
	Twitter	-9.8	40.10845	0.809	-93.4648	73.8648
	Linkedin	-114.80000*	40.10845	0.01	-198.4648	-31.1352
	Retellity	2.2	40.10845	0.957	-81.4648	85.8648
* The mean difference is significant at the 0.05 level.						

- **Neutralization**

	Facebook	Twitter	LinkedIn	Retellity	Reddit	Mean of Single LP
On Sale	254	263	459	256	247	295.8
OnSale 30% off	324	301	548	325	312	362
New Arrivals 20% off	282	354	496	302	287	344.2
Hot New Merch	325	325	288	289	309	307.2
Control	329	288	265	299	327	301.6
Total	1514	1531	2056	1471	1482	1610.8
Mean	302.8	306.2	411.2	294.2	296.4	322.16
Difference From Overall Mean	-19.36	-15.96	89.04	-27.96	-25.76	0

-

Neutralization

	Facebook	Twitter	LinkedIn	Retellity	Reddit	Mean of Single LP
On Sale	273.36	278.96	369.96	283.96	272.76	295.8
OnSale 30% off	343.36	316.96	458.96	352.96	337.76	362
New Arrivals 20% off	301.36	369.96	406.96	329.96	312.76	344.2
Hot New Merch	344.36	340.96	198.96	316.96	334.76	307.2
Control	348.36	303.96	175.96	326.96	352.76	301.6
Total	1610.8	1610.8	1610.8	1610.8	1610.8	1610.8
Mean	322.16	322.16	322.16	322.16	322.16	322.16
Difference From Overall Mean	0	0	0	0	0	0

-

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	5	1610.8	322.16	1110.7		
Column 2	5	1610.8	322.16	1215.7		
Column 3	5	1610.8	322.16	16185.7		
Column 4	5	1610.8	322.16	629.7		
Column 5	5	1610.8	322.16	966.8		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0	4	0	0	1	2.866081402
Within Groups	80434.4	20	4021.72			
Total	80434.4	24				

Output 2: Calculate Averages, Medians and SD for registrations by LP

On Sale		**OnSale 30% off**		**New Arrivals 20% off**		**Hot New Merch**		**Control**	
Mean	**295.8**	**Mean**	**362**	**Mean**	**344.2**	**Mean**	**307.2**	**Mean**	**301.6**
Standard Error	18.65190607	Standard Error	24.94653483	Standard Error	19.530632	Standard Error	27.4693575	Standard Error	32.58677032
Median	**278.96**	**Median**	**343.36**	**Median**	**329.96**	**Median**	**334.76**	**Median**	**326.96**
Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard Deviation	**41.70692988**	**Standard Deviation**	**55.78214768**	**Standard Deviation**	**43.671822**	**Standard Deviation**	**61.4233506**	**Standard Deviation**	**72.86623361**
Sample Variance	1739.468	Sample Variance	3111.648	Sample Variance	1907.228	Sample Variance	3772.828	Sample Variance	5309.488
Kurtosis	4.758078372	Kurtosis	4.002662538	Kurtosis	-1.0099953	Kurtosis	4.38686891	Kurtosis	3.582916405
Skewness	2.170362683	Skewness	1.917328694	Skewness	0.7589054	Skewness	-2.0793356	Skewness	-1.864059072
Range	97.2	Range	142	Range	105.6	Range	145.4	Range	176.8
Minimum	272.76	Minimum	316.96	Minimum	301.36	Minimum	198.96	Minimum	175.96
Maximum	369.96	Maximum	458.96	Maximum	406.96	Maximum	344.36	Maximum	352.76
Sum	1479	Sum	1810	Sum	1721	Sum	1536	Sum	1508
Count	5	Count	5	Count	5	Count	5	Count	5

Output 3: One-Factor Plot

	On Sale	OnSale 30% Off	New Arrivals 20% Off	Hot New Merch	Control
Max	369.96	458.96	406.96	344.36	352.76
	283.96	★ 352.96	★ 369.96	340.96	348.36
Median	★ 278.96	343.36	329.96	★ 334.76	★ 326.96
	273.36	337.76	312.76	316.96	303.96
Min	272.76	316.96	301.36	198.96	175.96
Range	97.2	142	105.6	145.4	176.8
AVG	296	362	344	307	302
Var	1739	3112	1907	3773	5309

Least Significant Difference =			69.77		
LSD =	t(critical @ 95%)	x	SD (pooled)	x	SQR(2/n)
	1.96	x	56		0.632455532

Output 4: ANOVA and LSD

There is no significant difference between different landing pages after neutralization.

ANOVA

Registration

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	17071.760	4	4267.940	1.347	.287
Within Groups	63362.640	20	3168.132		
Total	80434.400	24			

Multiple Comparisons

Dependent Variable: Registration

LSD

(I) LandingPage	(J) LandingPage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
On Sale	OnSale 30% Off	-66.20000	35.59849	.078	-140.4572	8.0572
	NewArrivals 20% Off	-48.40000	35.59849	.189	-122.6572	25.8572
	Hot New Merchandise	-11.40000	35.59849	.752	-85.6572	62.8572
	Control	-5.80000	35.59849	.872	-80.0572	68.4572
OnSale 30% Off	On Sale	66.20000	35.59849	.078	-8.0572	140.4572
	NewArrivals 20% Off	17.80000	35.59849	.623	-56.4572	92.0572
	Hot New Merchandise	54.80000	35.59849	.139	-19.4572	129.0572
	Control	60.40000	35.59849	.105	-13.8572	134.6572
NewArrivals 20% Off	On Sale	48.40000	35.59849	.189	-25.8572	122.6572
	OnSale 30% Off	-17.80000	35.59849	.623	-92.0572	56.4572
	Hot New Merchandise	37.00000	35.59849	.311	-37.2572	111.2572
	Control	42.60000	35.59849	.245	-31.6572	116.8572
Hot New Merchandise	On Sale	11.40000	35.59849	.752	-62.8572	85.6572
	OnSale 30% Off	-54.80000	35.59849	.139	-129.0572	19.4572
	NewArrivals 20% Off	-37.00000	35.59849	.311	-111.2572	37.2572
	Control	5.60000	35.59849	.877	-68.6572	79.8572
Control	On Sale	5.80000	35.59849	.872	-68.4572	80.0572
	OnSale 30% Off	-60.40000	35.59849	.105	-134.6572	13.8572
	NewArrivals 20% Off	-42.60000	35.59849	.245	-116.8572	31.6572
	Hot New Merchandise	-5.60000	35.59849	.877	-79.8572	68.6572

Output 5: t-test

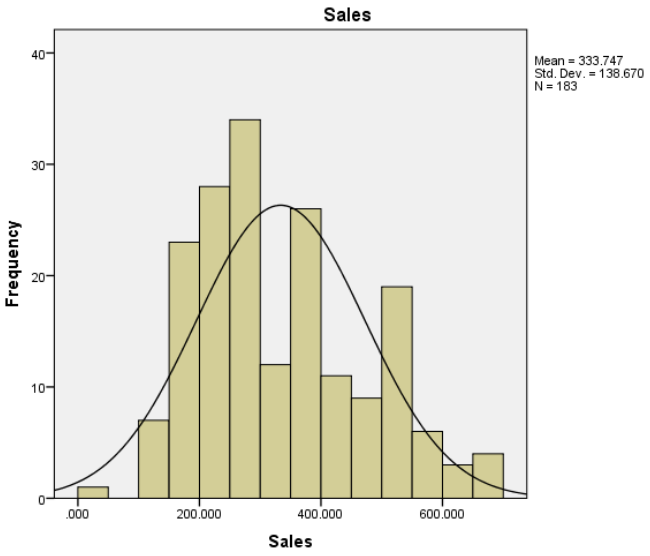
Since there is no significant difference between different landing pages, I would like to do a t-test of control group with on-sale 30% off to affirm my conclusion.

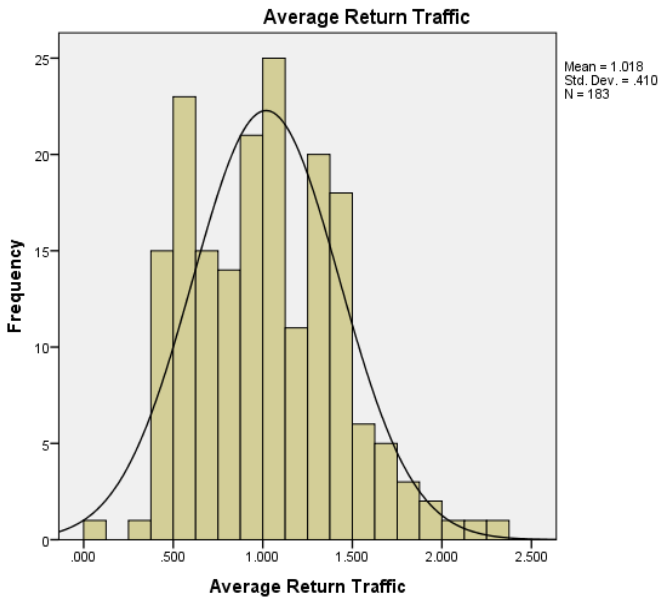
34.22% > 5%, accept null hypothesis which is the two variables are significantly the same. There is no significant difference between "control" group with " on-sale 30% off".

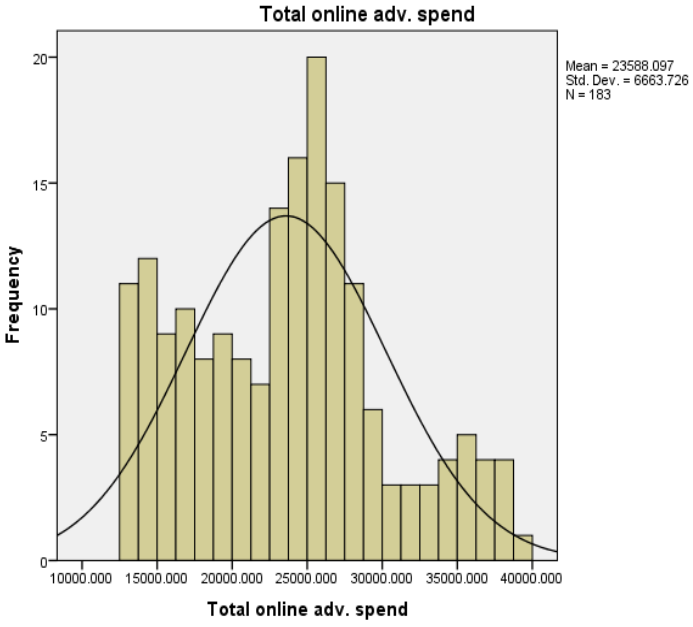
t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	362	301.6
Variance	3111.648	5309.488
Observations	5	5
Pearson Correlation	-0.89979346	
Hypothesized Mean Difference	0	
df	4	
t Stat	1.076657394	
P(T<=t) one-tail	0.17111879	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.34223758	
t Critical two-tail	2.776445105	

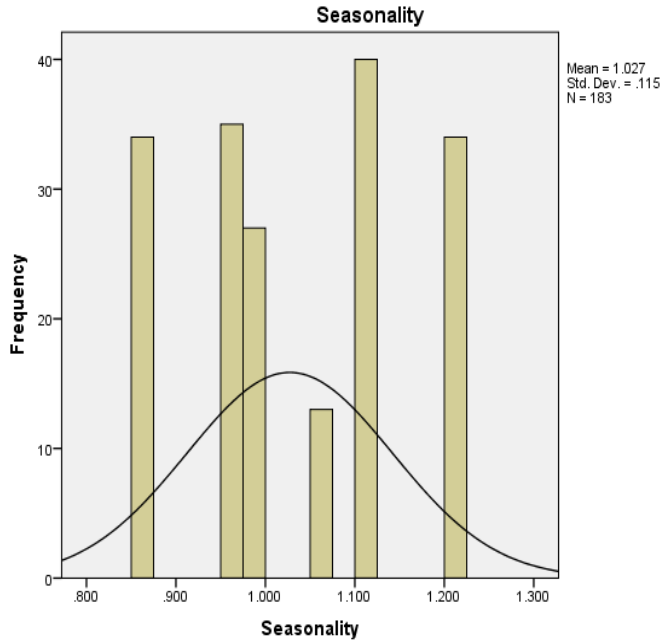
Part 3:

Step 1: Histograms:



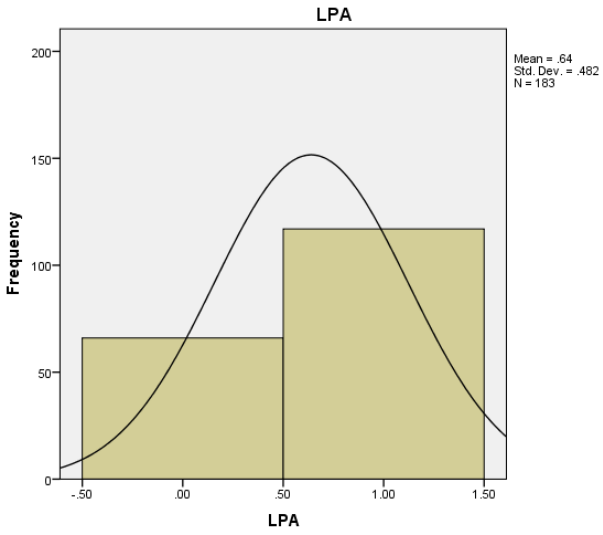


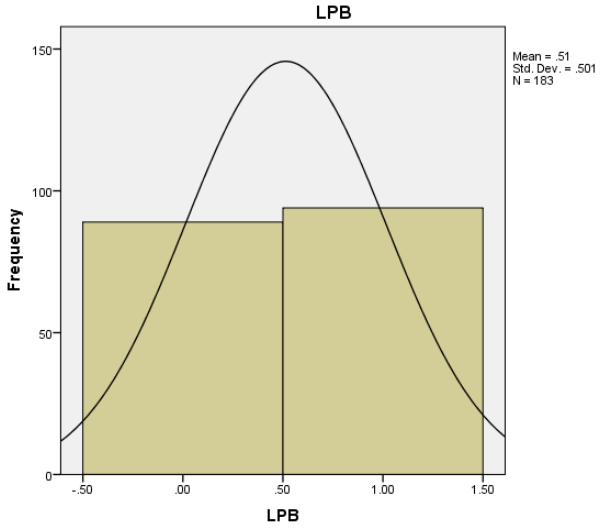


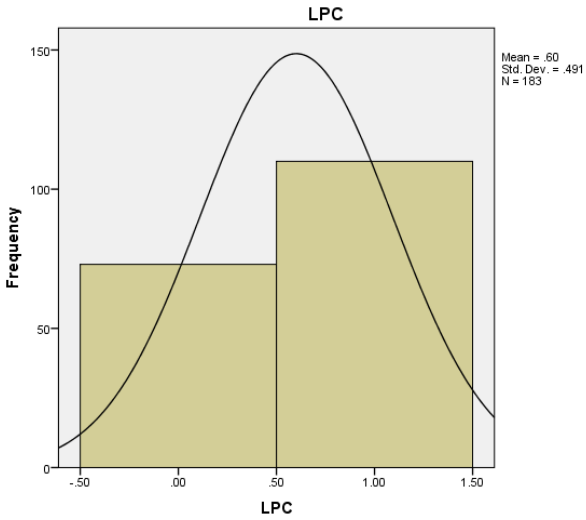


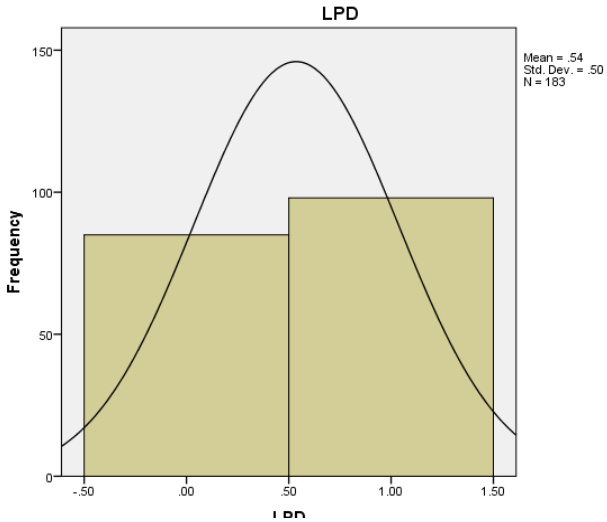
Observation 1:

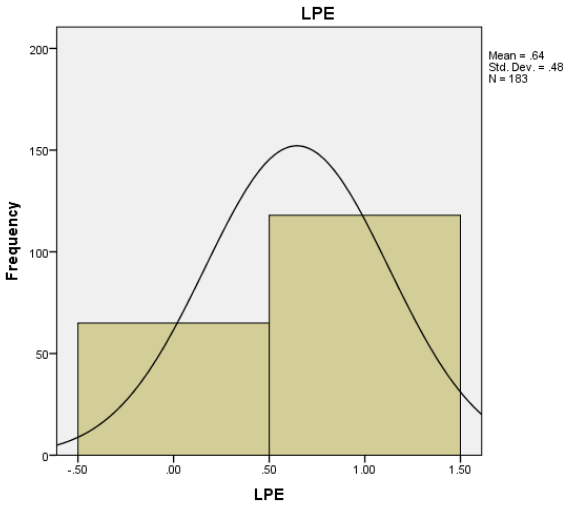
Sales is normally distributed. Seasonality may be transformed into a binary variable while total online adv. Spend also needs to be transformed.

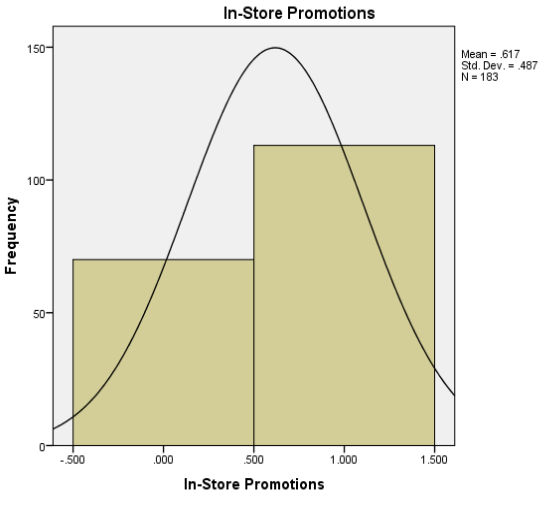




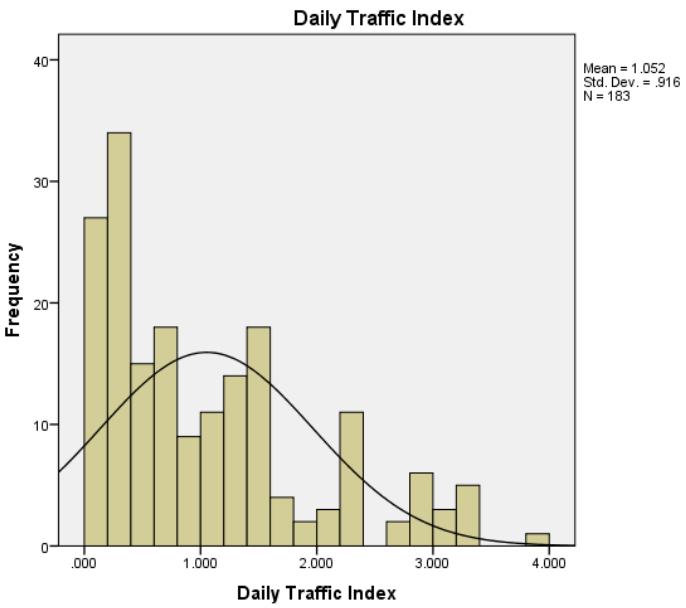


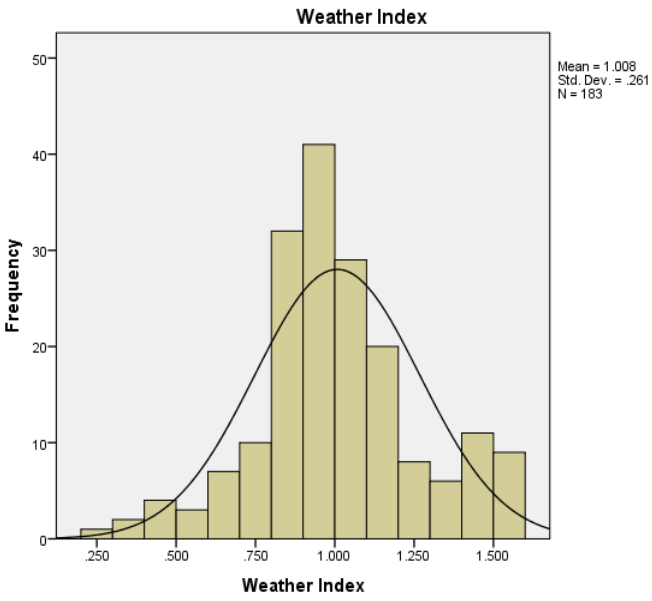






Observation 2: LPA/LPB/LPC/LPD/LPE and In-Store Promotions are all binary variables.



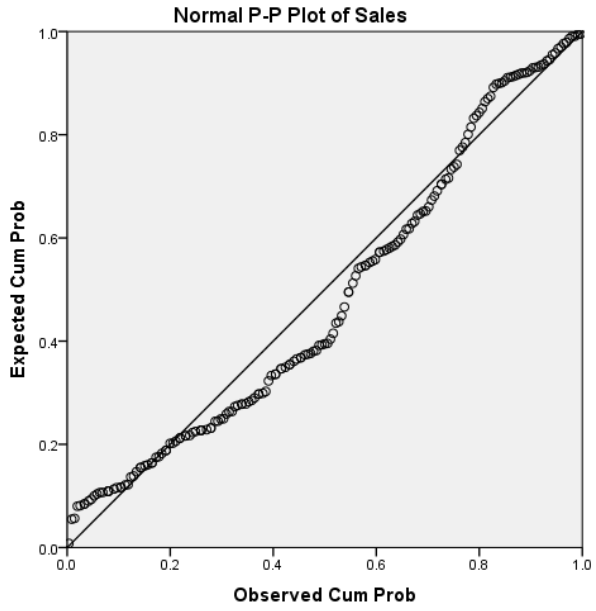


Observation 3: Weather Index is good to use while Daily Traffic Index needs to be transformed.

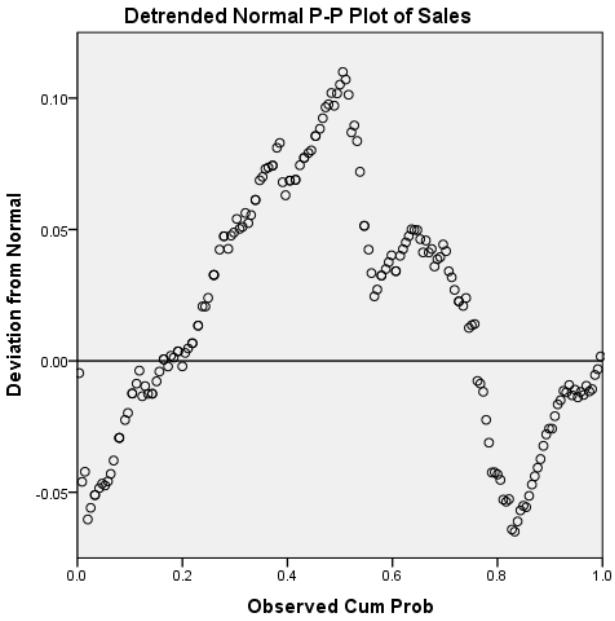
Conclusion:

1. Variables that can be used are
    a. Normalized Variables: Sales (DV) , Weather Index
    b. Binarized Variables: LPA, LPB,LPC,LPD,LPE , In-Store Promotions
2. Variables that should be transformed to be better used are: Average Return Traffic, Total Online Adv. Spend, Seasonality, Daily Traffic Index

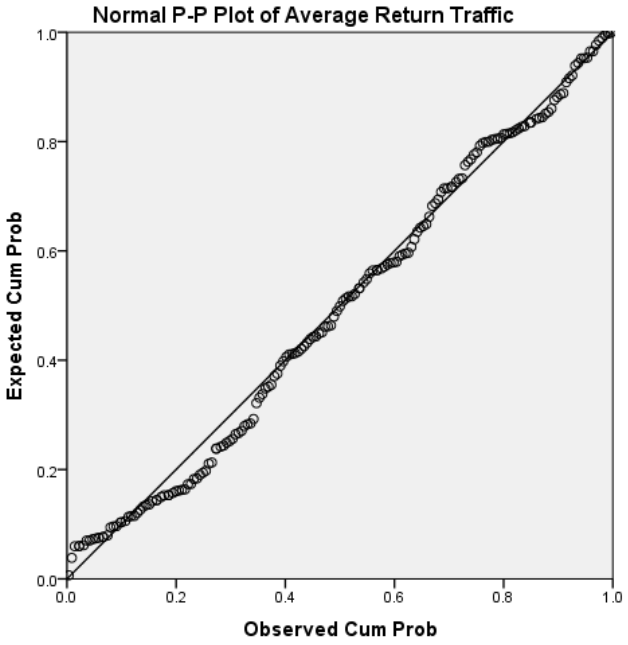
Step 2: Normality Plots and Descriptive Analysis



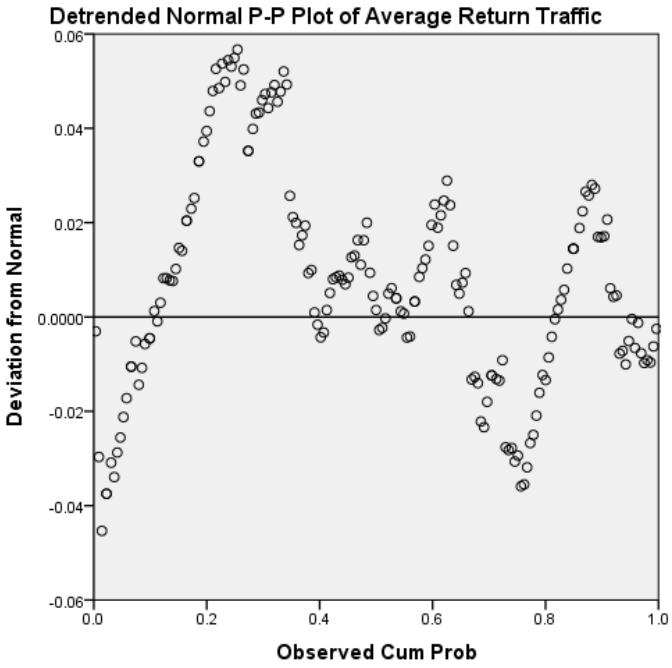
The observed cumulative probability is almost around the diagonal line indicating that it may be normally distributed.



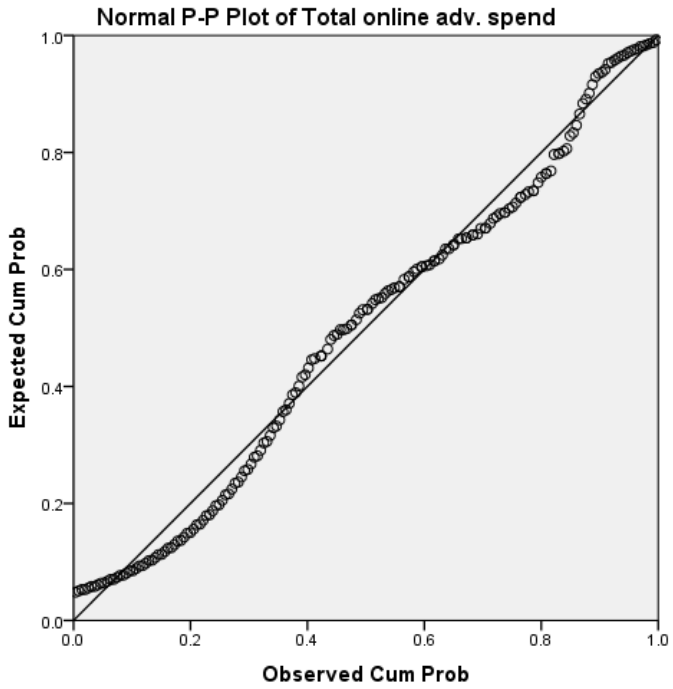
However, the detrended normal P-P plot shows that there are deviations of the observed values from normalized values.



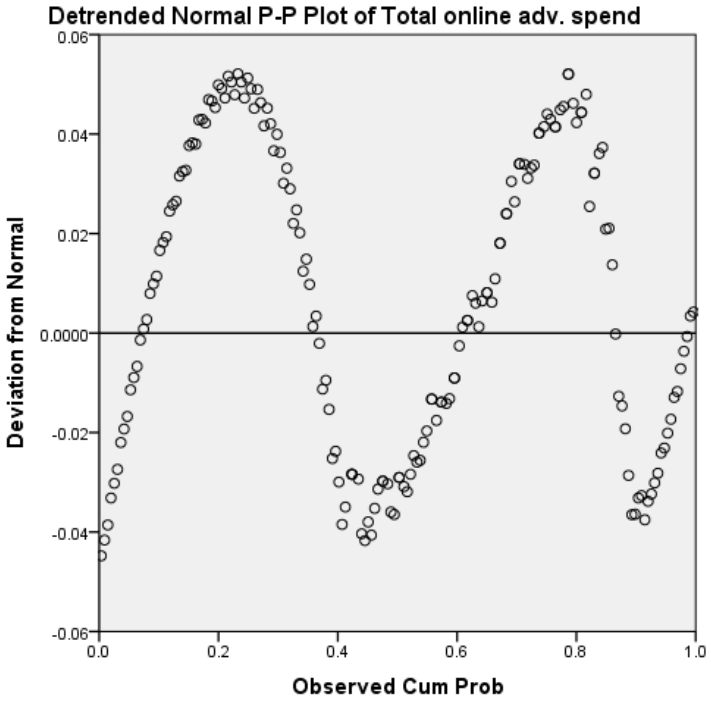
Observed values are close to expected normalized values.



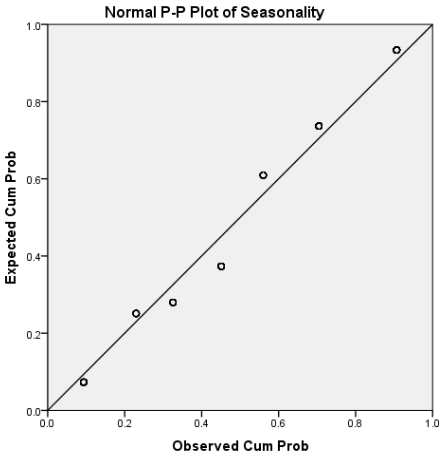
The total probabilities of not deviating from o is almost 0.

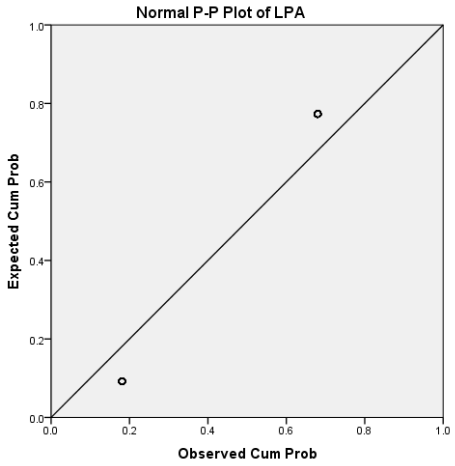


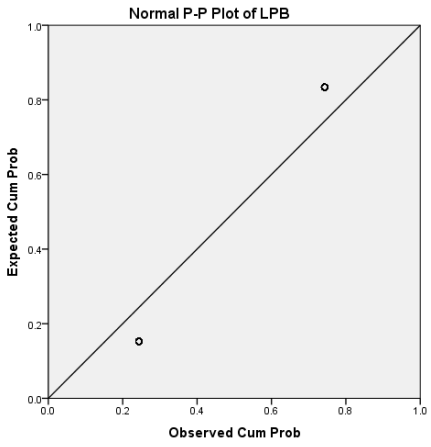
Almost all of the dots are around the 45 degree line.

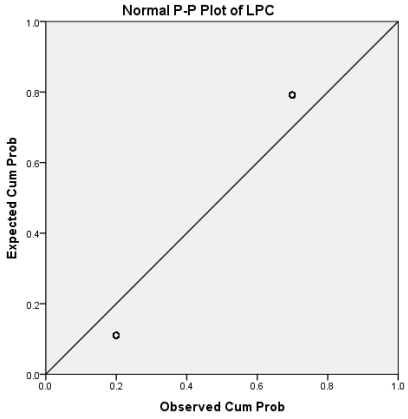


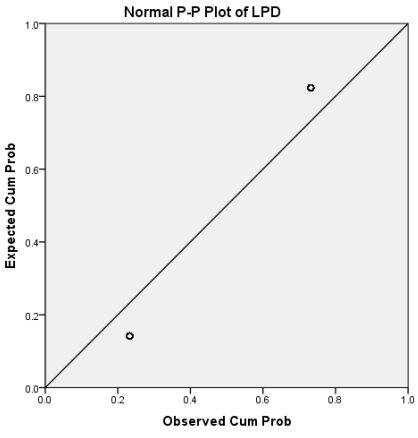
There is deviation but not large enough.

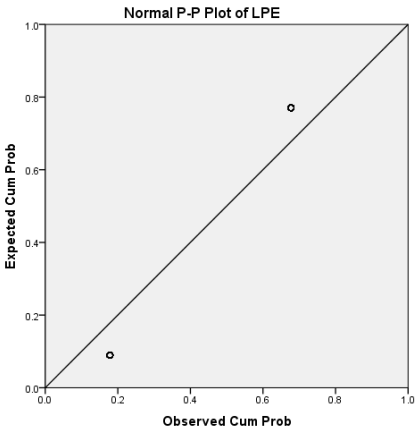

Normal P-P Plot of Seasonality
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

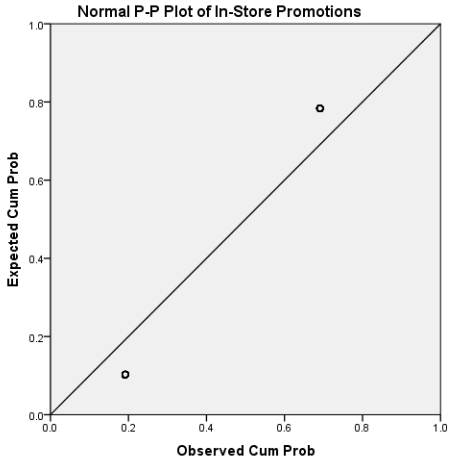

Normal P-P Plot of LPA
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

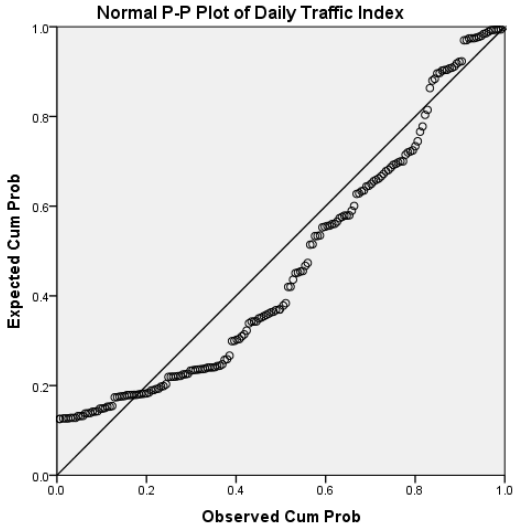

Normal P-P Plot of LPB
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of LPC
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

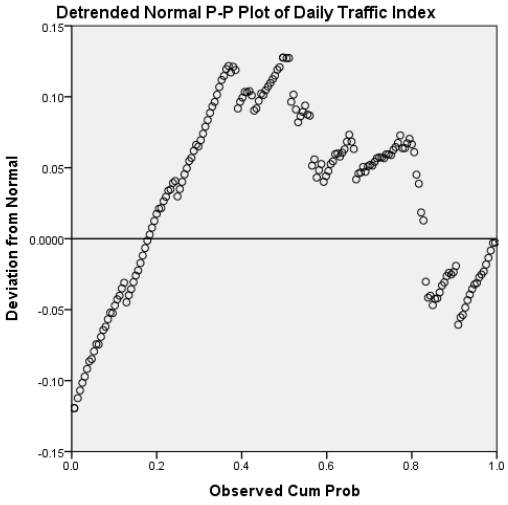

Normal P-P Plot of LPD
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

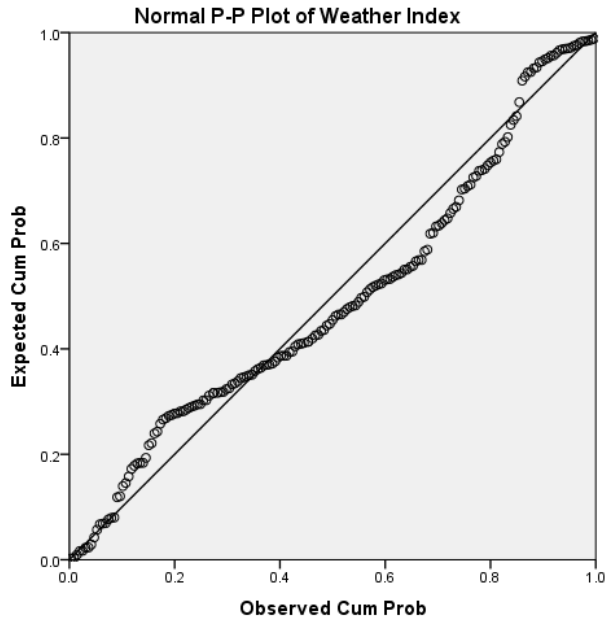

Normal P-P Plot of LPE
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0



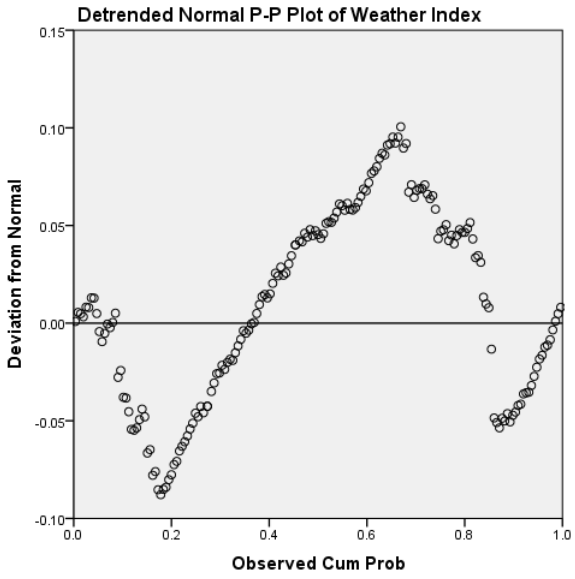


There is large deviation from the equation line.





The fitness is just fine but not perfect.



There is deviation from the normally distributed line.

Conclusion:

Average return traffic performed the best in the round. Online adv. Spend is also close to the 45 degree line. Combined the conclusion from the previous part, I conclude that except for the binary variables, most of the variables are normally distributed to some extent but not perfect.

Removing Outliers:

By calculating cooking values, I found three outliers and would like to remove them for further analysis.

Sales/(orders)	Average Return Traffic	Total online adv. spend	Seasonality	LPA	LPB	LPC	LPD	LPE	Daily traffic index	Weather Index	In-Store Promotions	Cook Value
662	1.1	16855	0.99	0	0	0	0	0	1.23	1.42	0	0.06743
252	1.8	17678.7	0.99	1	1	1	1	1	2.25	0.49	1	0.07753
546	0.9	19682.3	1.06	1	1	0	1	1	0.57	0.8	0	0.05808

Descriptives:

(Conducted at the same time with removing outliers)

Statistics

		Sales	Average Return Traffic	Total online adv. spend	Seasonality	LPA	LPB	LPC	LPD	LPE	Daily Traffic Index	Weather Index	In-Store Promotions
N	Valid	183	183	183	183	183	183	183	183	183	183	183	183
	Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean		333.74736	1.01835	23588.09743	1.02721	.6393	.5137	.6011	.5355	.6448	1.05182	1.00779	.61749
Median		296.75800	1.01684	24119.23303	.99000	1.0000	1.0000	1.0000	1.0000	1.0000	.74691	.97810	1.00000
Mode		163.000[a]	1.417	22784.187	1.100	1.00	1.00	1.00	1.00	1.00	.000[a]	.883[a]	1.000
Std. Deviation		138.669736	.409572	6663.726289	.115061	.48151	.50118	.49102	.50011	.47988	.916480	.260517	.487334
Skewness		.526	.359	.284	.065	-.585	-.055	-.416	-.144	-.610	.999	.148	-.487
Std. Error of Skewness		.180	.180	.180	.180	.180	.180	.180	.180	.180	.180	.180	.180
Kurtosis		-.437	-.081	-.504	-1.138	-1.676	-2.019	-1.847	-2.001	-1.646	.265	.355	-1.782
Std. Error of Kurtosis		.357	.357	.357	.357	.357	.357	.357	.357	.357	.357	.357	.357
Range		690.000	2.297	27241.720	.340	1.00	1.00	1.00	1.00	1.00	3.985	1.322	1.000
Minimum		.000	.000	12508.804	.860	.00	.00	.00	.00	.00	.000	.278	.000
Maximum		690.000	2.297	39750.523	1.200	1.00	1.00	1.00	1.00	1.00	3.985	1.600	1.000
Sum		61075.766	186.359	4316621.829	187.979	117.00	94.00	110.00	98.00	118.00	192.483	184.426	113.000

a. Multiple modes exist. The smallest value is shown

1st Rule: Standard Deviation is 45% of Mean, for binary variables: it cannot have the same value for more than 90% of the rows.

Sales: 42% ; Average Return Traffic: 40% ; Total Online Adv. Spend: 28% ; Seasonality: 11% ;

Daily Traffic Index: 87% ; Weather Index: 26%

Sales and average return traffic are stabler than other variables. But variables with the percentages much lower or higher do not necessarily to be not good.

2nd Rule: Skewness close to 1

All of the skewness is less than 1 which means that most of them are right hand side skewness distribution. Average Return Traffic is almost 1 and it is the most balanced one.

3rd Rule: Kurtosis is a small and positive number.

Only Daily Traffic Index and Weather Index meet the requirements. All of the other variables have negative kurtosis which means that the distribution is choppy and wide.

List of Assumptions of Linear Regression Model:

1. Each IV should be NORMALLY distributed
    a. The sum of the residuals should be ~ Zero
    b. There should be no specific pattern in the Residual plot
2. The variance for each of the IVs needs to be more or less equal
3. The IVs should be independent of each other

Variable Transformation:

Based on the descriptive analysis above, I decided to transform Sales, Total Online Adv. Spend, Average Return Traffic and Daily Traffic Index.

	s.d deviation	0.4149745
	mean	1.020765
Square Root	**Standarization**	**EXP**
1	-0.050039289	2.7182818
0.948683298	-0.291017972	2.4596031
1.140175425	0.672896758	3.6692967
1.183215957	0.913875441	4.0552
1.183215957	0.913875441	4.0552
1.140175425	0.672896758	3.6692967
1.095445115	0.431918076	3.3201169
1.048808848	0.190939393	3.004166
0.948683298	-0.291017972	2.4596031
1.095445115	0.431918076	3.3201169
1.048808848	0.190939393	3.004166
1	-0.050039289	2.7182818
0.836660027	-0.772975337	2.0137527
1.048808848	0.190939393	3.004166
1.183215957	0.913875441	4.0552
1.303840481	1.636811488	5.4739474
1.183215957	0.913875441	4.0552
1.183215957	0.913875441	4.0552
1.183215957	0.913875441	4.0552
0.774596669	-1.013954019	1.8221188
1.140175425	0.672896758	3.6692967
1.224744871	1.154854123	4.4816891
1.264911064	1.395832806	4.9530324
1.183215957	0.913875441	4.0552
1.183215957	0.913875441	4.0552
1.183215957	0.913875441	4.0552
1.341640786	1.877790171	6.0496475
1.140175425	0.672896758	3.6692967
1.183215957	0.913875441	4.0552
1	-0.050039289	2.7182818
1.095445115	0.431918076	3.3201169

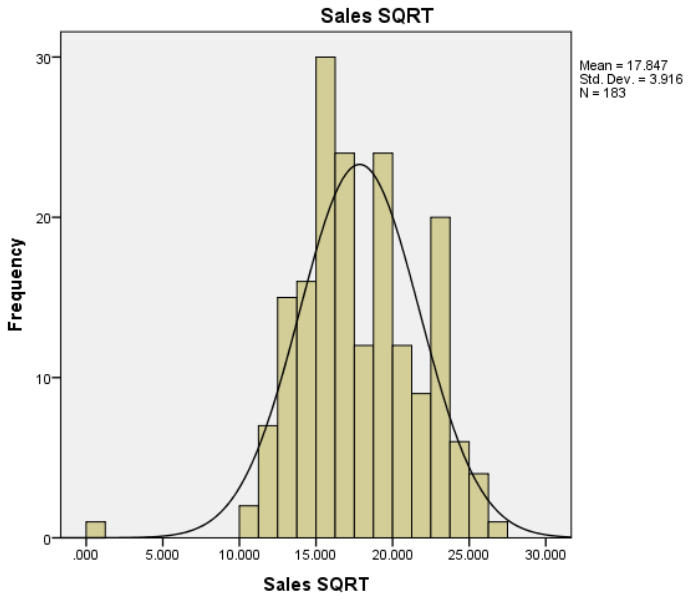
		s.d deviation	6663.726655
		mean	23588.09563
Total online adv. spend	**Log**	**Square Root**	**Standarization**
23402	4.369252975	152.9771225	-0.02792666
24572.1	4.390442274	156.7549042	0.147665777
22784.2	4.357633784	150.9443606	-0.120637546
23675.2	4.374293656	153.8674754	0.01307142
24773.2	4.393982109	157.3950444	0.177844085
26011.9	4.415172076	161.2820511	0.363731062
24119.2	4.382362899	155.3035737	0.079700804
25062.5	4.39902439	158.311402	0.221258231
26224.8	4.418712185	161.9407299	0.395680151
27536.1	4.43990243	165.9400494	0.592461933
25532.5	4.407093341	159.7889233	0.291789335
26531	4.42375362	162.8833939	0.441630416
27761.5	4.443442928	166.6178262	0.626286849
29149.6	4.4646326	170.7325394	0.834593713
27028.6	4.431823551	164.4037712	0.516303346
28380	4.453012391	168.4636459	0.719102781
26315	4.420203374	162.2189878	0.409216121
27344.1	4.436863634	165.3605152	0.563649226
25354.5	4.404055051	159.2309643	0.265077555
26346.1	4.420716336	162.3148176	0.413883179
24429.1	4.387907467	156.2981126	0.126206313
25384.5	4.404568614	159.3251393	0.269579541
23537.4	4.371758488	153.419034	-0.007607699
24457.9	4.388419165	156.3902171	0.130528219
22678.3	4.355610496	150.5931605	-0.136529554
23565.2	4.37227113	153.5096088	-0.003435859
24658.1	4.39195961	157.0289782	0.160571468
25891	4.413148825	160.9068053	0.345588061
24007.2	4.380341511	154.94257	0.062893392
25207.5	4.401529776	158.7686997	0.243017827
23373.4	4.368721891	152.8836159	-0.032218553

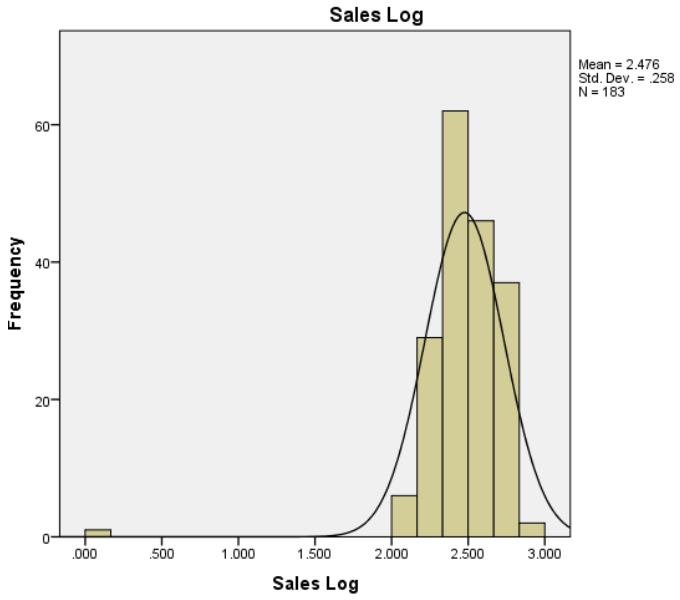
		s.d deviation	0.916435494	
		mean	1.051912568	
Daily traffic index	**Log**	**Square Root**	**Standarization**	**EXP**
0.25	-0.60206	0.5	-0.875034385	1.284025417
0.01	-2	0.1	-1.136918611	1.010050167
0.11	-0.958607	0.3316625	-1.027800183	1.11627807
0.07	-1.154902	0.2645751	-1.071447554	1.072508181
0.2	-0.69897	0.4472136	-0.929593598	1.221402758
0.21	-0.677781	0.4582576	-0.918681756	1.23367806
0.34	-0.468521	0.5830952	-0.7768278	1.404947591
0.35	-0.455932	0.591608	-0.765915957	1.419067549
0.73	-0.136677	0.8544004	-0.351265933	2.075080608
0.19	-0.721246	0.4358899	-0.940505441	1.209249598
0.21	-0.677781	0.4582576	-0.918681756	1.23367806
1.13	0.0530784	1.0630146	0.085207778	3.0956565
0.24	-0.619789	0.4898979	-0.885946227	1.27124915
0.03	-1.522879	0.1732051	-1.115094925	1.030454534
0.87	-0.060481	0.9327379	-0.198500134	2.386910854
1.43	0.155336	1.1958261	0.41256306	4.178699192
1.24	0.0934217	1.1135529	0.205238048	3.455613465
0.95	-0.022276	0.9746794	-0.111205392	2.585709659
0.1	-1	0.3162278	-1.038712026	1.105170918
0.68	-0.167491	0.8246211	-0.405825146	1.973877732
2.27	0.3560259	1.5066519	1.329157851	9.679400814
3.36	0.5263393	1.8330303	2.518548711	28.78919088
2.89	0.4608978	1.7	2.005692102	17.9933096
2.83	0.4517864	1.6822604	1.940221045	16.94546082
2.24	0.350248	1.496663	1.296422323	9.393331287
1.19	0.075547	1.0908712	0.150678834	3.287081207
1.2	0.0791812	1.0954451	0.161590677	3.320116923
1.48	0.1702617	1.2165525	0.467122274	4.392945681
1.53	0.1846914	1.2369317	0.521681488	4.618176822
0.57	-0.244125	0.7549834	-0.525855417	1.768267051
0.7	-0.154902	0.83666	-0.384001461	2.013752707

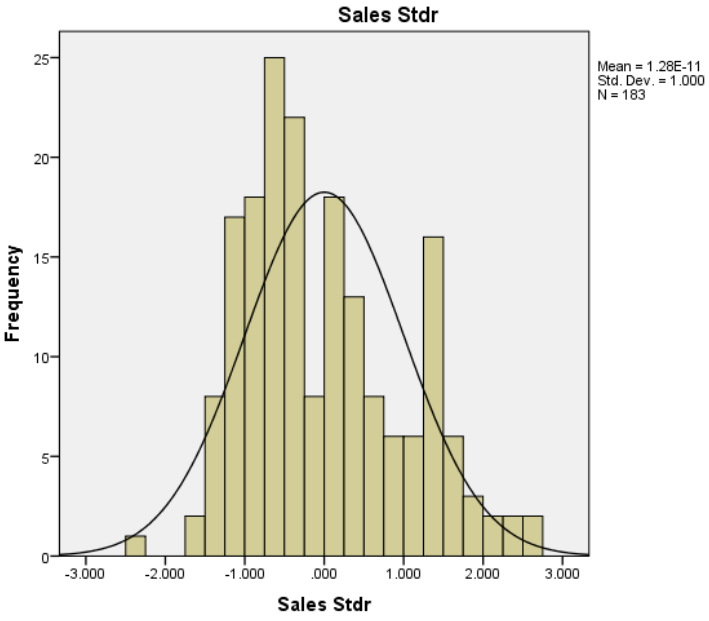
1- Seasonality	1. 2=1
2- Seasonality	1. 1=1
3-Seasonality	0. 96=1
4-Seasonality	0. 95=1
5-Seasonality	0. 86=1
6-Seasonality	0. 99=1
7-Seasonality	1. 06=1

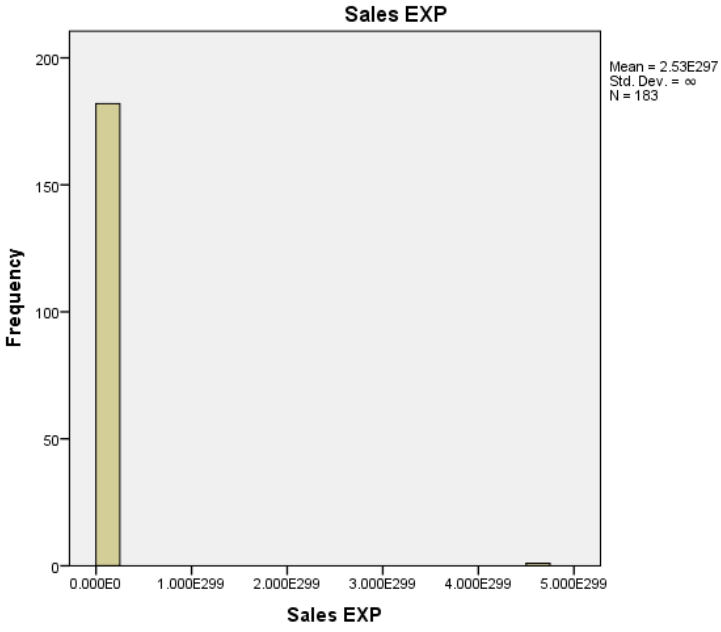
Seasonality	1-Seasonality	2- Seasonality	3-Seasonality	4-Seasonality	5-Seasonality	6-Seasonality	7-Seasonality
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.2	1	0	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0
1.1	0	1	0	0	0	0	0

Histograms of All

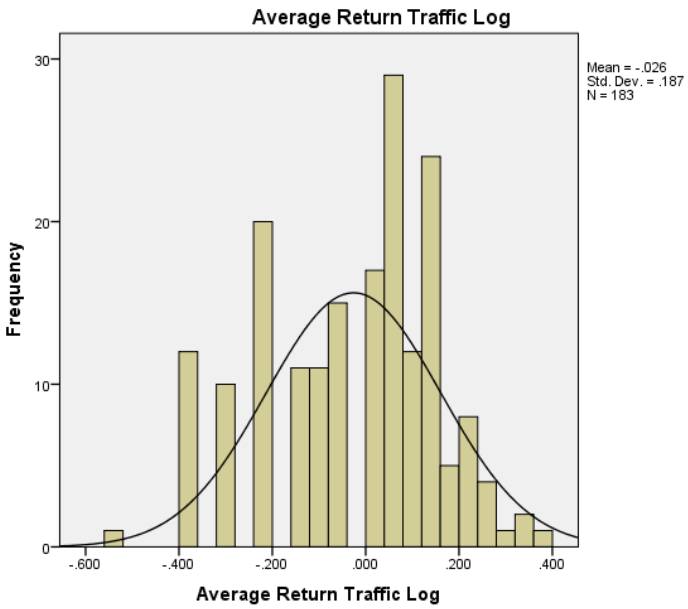


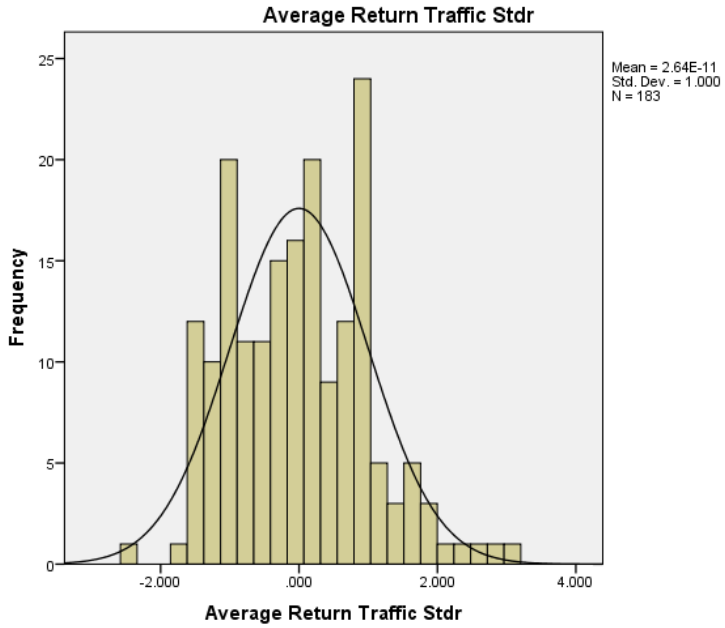


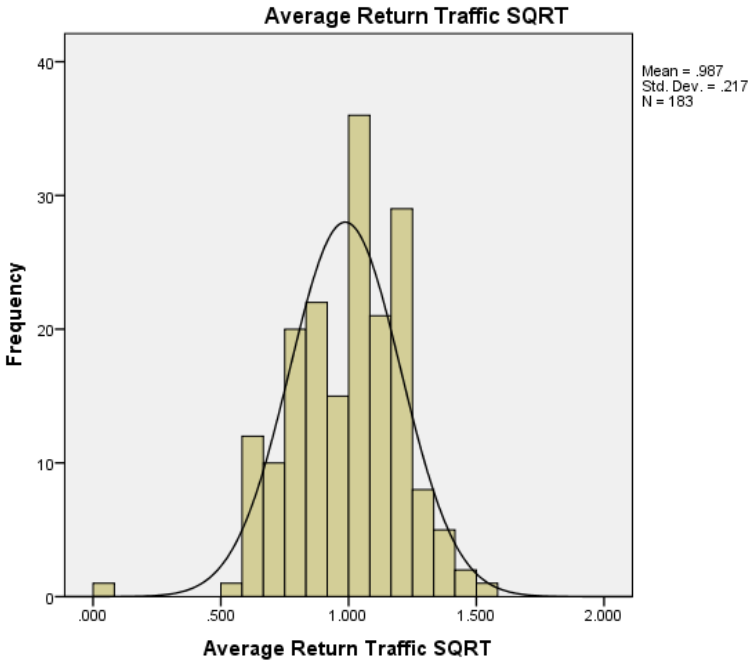


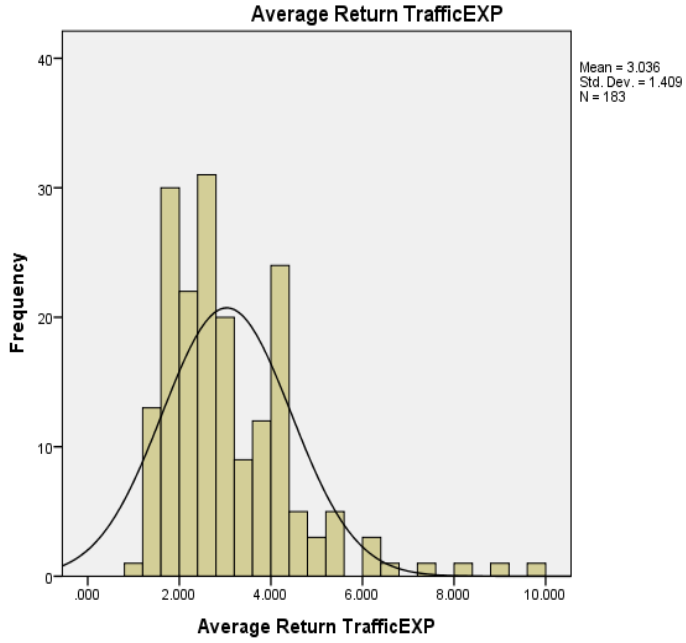


Good to use variables: Sales SQRT/Sales Stdr

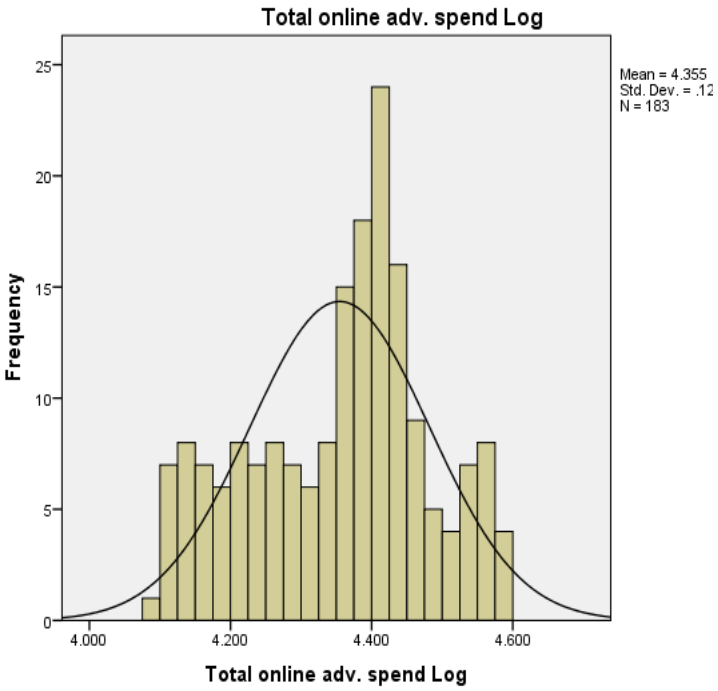


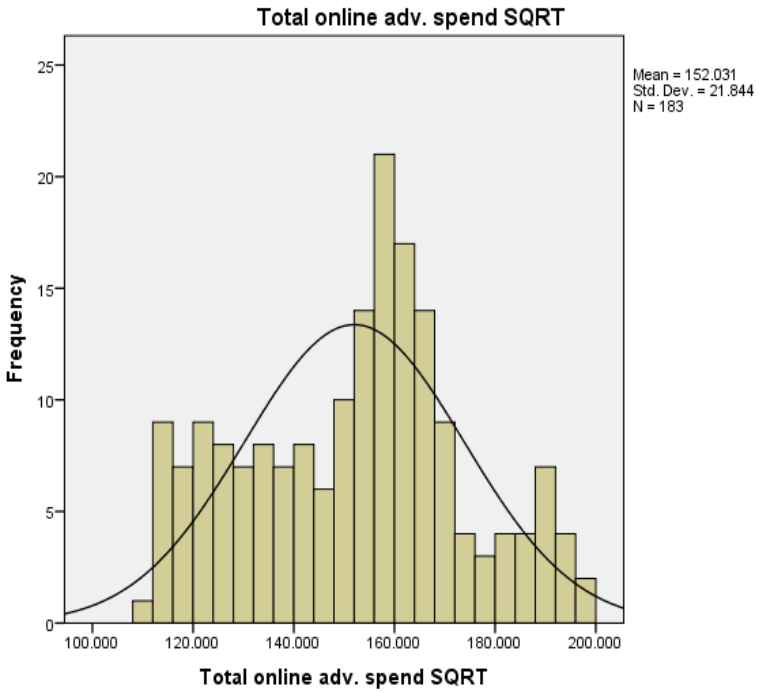


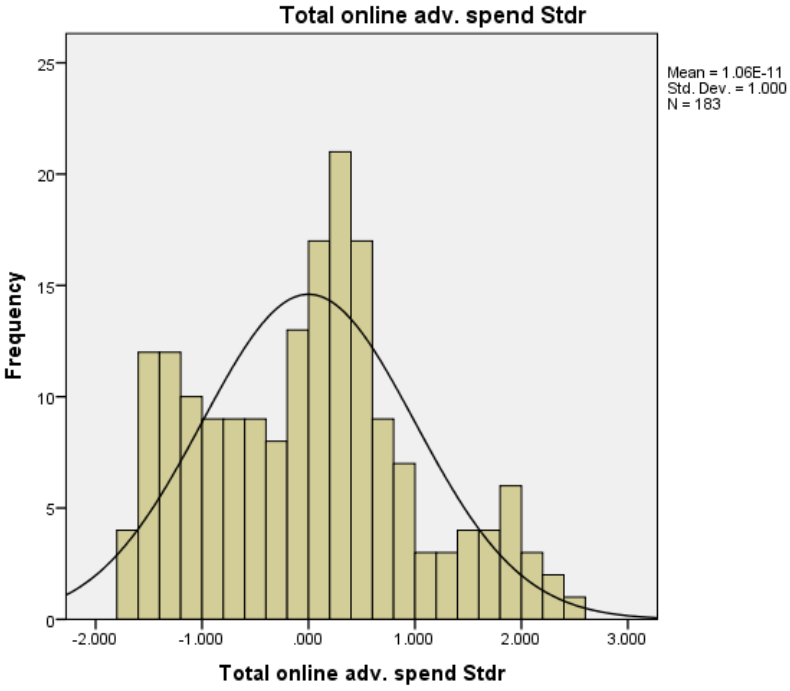


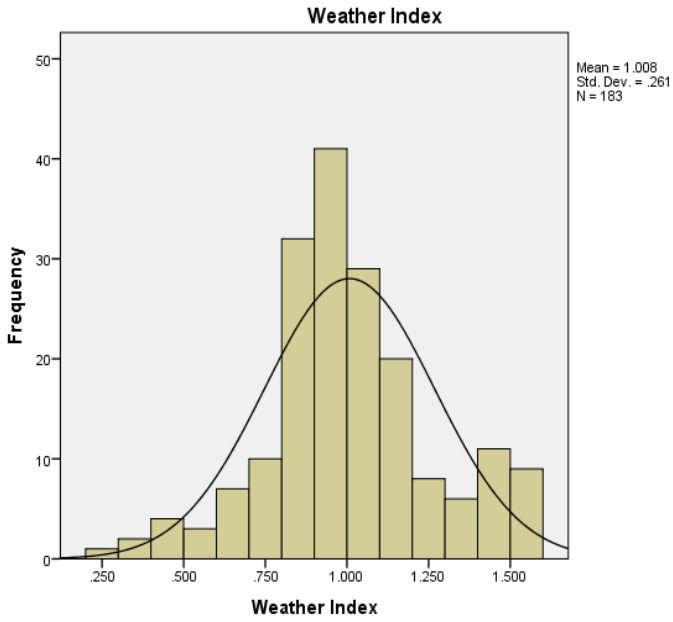


Good to use Variables: Average Return Traffic SQRT

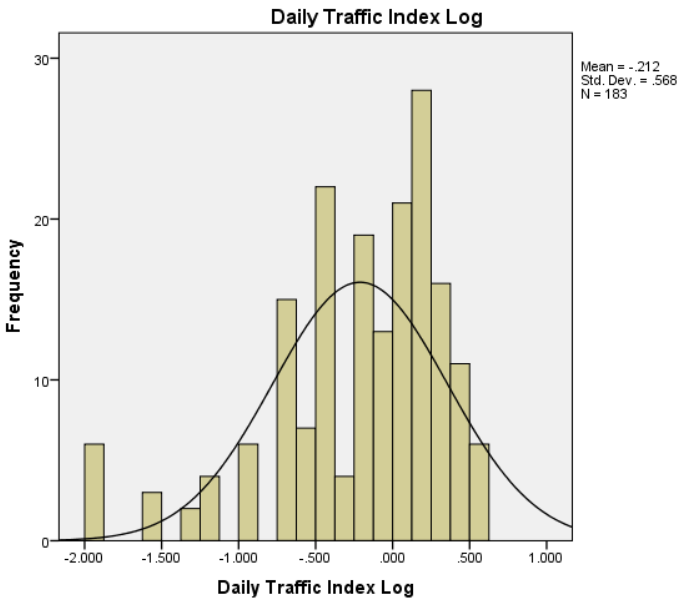


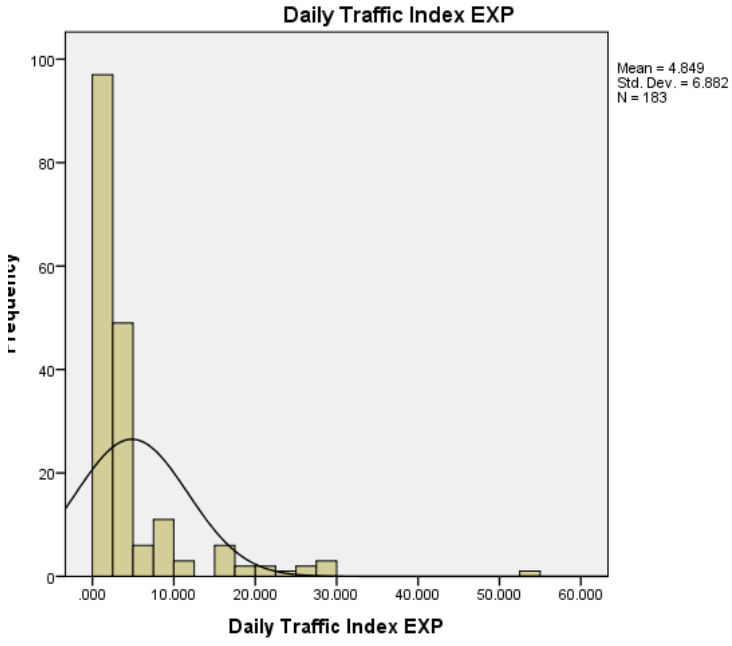






Good to use variables: total online adv. Spend Log/SQRT/Stdr and Weather Indez





Good to use variables: Daily Traffic Index SQRT

Descriptives:

Statistics

		Weather Index	Sales Log	Sales SQRT	Sales Stdr	Sales EXP
N	Valid	180	180	180	180	180
	Missing	0	0	0	0	0
Mean		1.00953	2.47283	17.78360	-.01838	.
Median		.98055	2.47202	17.21917	-.26877	6.6142E+128
Std. Deviation		.257492	.257636	3.879350	.984933	.000000
Skewness		.162	-4.956	-.270	.518	
Std. Error of Skewness		.181	.181	.181	.181	.181
Kurtosis		.437	46.517	1.295	-.416	
Std. Error of Kurtosis		.360	.360	.360	.360	.360
Range		1.322	2.839	26.268	4.976	4.605E+299
Minimum		.278	.000	.000	-2.407	1.000
Maximum		1.600	2.839	26.268	2.569	4.605E+299
Sum		181.715	445.109	3201.048	-3.308	.

Statistics

		Average Return Traffic Log	Average Return Traffic SQRT	Average Return Traffic Stdr	Average Return TrafficEXP
N	Valid	180	180	180	180
	Missing	0	0	0	0
Mean		-.02789	.98473	-.00988	3.02241
Median		.00000	1.00000	-.05004	2.71828
Std. Deviation		.187219	.217326	.998141	1.401741
Skewness		-.429	-.470	.362	1.806
Std. Error of Skewness		.181	.181	.181	.181
Kurtosis		-.556	1.294	-.071	5.186
Std. Error of Kurtosis		.360	.360	.360	.360
Range		.885	1.517	5.543	8.974
Minimum		-.523	.000	-2.460	1.000
Maximum		.362	1.517	3.083	9.974
Sum		-5.021	177.252	-1.778	544.034

Statistics

		Total online adv. spend Log	Total online adv. spend SQRT	Total online adv. spend Stdr	Daily Traffic Index Log
N	Valid	180	180	180	180
	Missing	0	0	0	0
Mean		4.35638	152.32590	.01380	-.21650
Median		4.38387	155.57386	.09232	-.11071
Std. Deviation		.127552	21.897819	1.002274	.570607
Skewness		-.301	-.031	.255	-1.235
Std. Error of Skewness		.181	.181	.181	.181
Kurtosis		-.720	-.692	-.510	1.580
Std. Error of Kurtosis		.360	.360	.360	.360
Range		.502	87.533	4.088	2.600
Minimum		4.097	111.843	-1.663	-2.000
Maximum		4.599	199.375	2.425	.600
Sum		784.149	27418.662	2.483	-38.970

Statistics

		Daily Traffic Index SQRT	Daily Traffic Index Stdr	Daily Traffic Index EXP
N	Valid	180	180	180
	Missing	0	0	0
Mean		.91111	-.00542	4.84812
Median		.86603	-.32944	2.11700
Std. Deviation		.466948	1.002709	6.926603
Skewness		.168	1.012	3.468
Std. Error of Skewness		.181	.181	.181
Kurtosis		-.748	.287	15.761
Std. Error of Kurtosis		.360	.360	.360
Range		1.995	4.343	52.517
Minimum		.000	-1.148	1.000
Maximum		1.995	3.195	53.517
Sum		163.999	-.976	872.661

Based on the analysis above, I decided to use the following variables for further testing:

Sales SQRT, Average Return Traffic SQRT , Weather Index , In-Store Promotions, Daily Traffic Index Log LPA, LPB, LPC, LPD, LPE, Seasonality 0.86, Seasonality 0.96, Seasonality 0.99, Seasonality 1.06, Seasonality 1.1, Seasonality 1.2, Seasonality 0.95

Univariate Regression model for the following independent variables

Variable – Weather Index

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Weather[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.136[a]	.018	.013	3.85419

a. Predictors: (Constant), Weather

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49.685	1	49.685	3.345	.069[b]
	Residual	2644.150	178	14.855		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), Weather

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	15.716	1.166		13.476	.000
	Weather	2.048	1.120	.136	1.829	.069

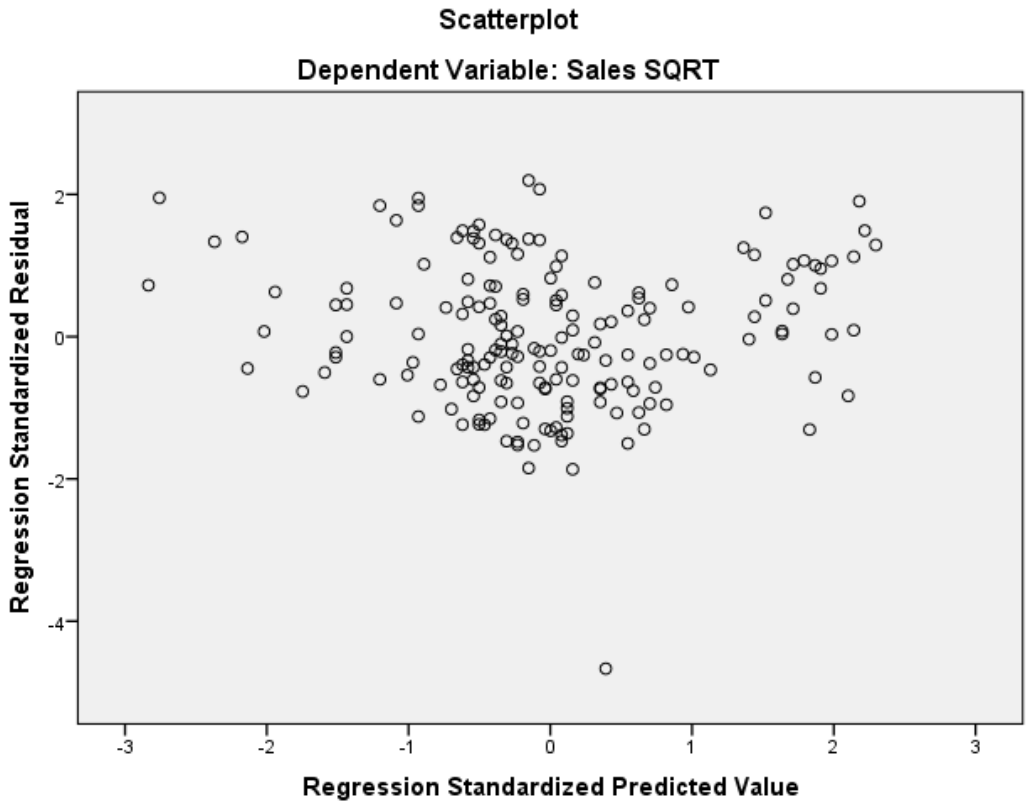
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

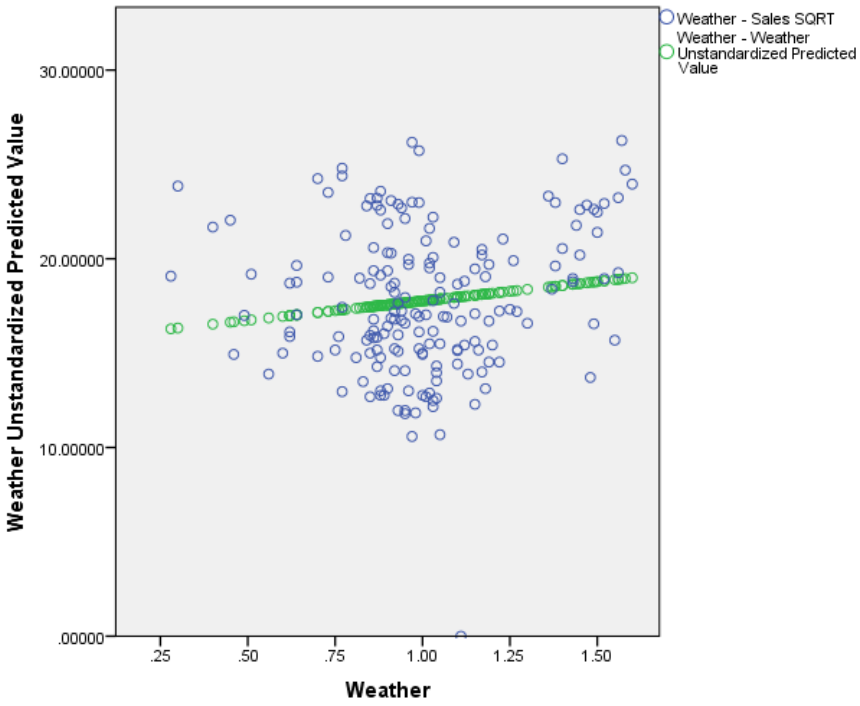
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	16.2898	18.9931	17.7836	.52685	180
Residual	-17.98965	8.46957	.00000	3.84341	180
Std. Predicted Value	-2.835	2.296	.000	1.000	180
Std. Residual	-4.668	2.197	.000	.997	180

a. Dependent Variable: Sales SQRT

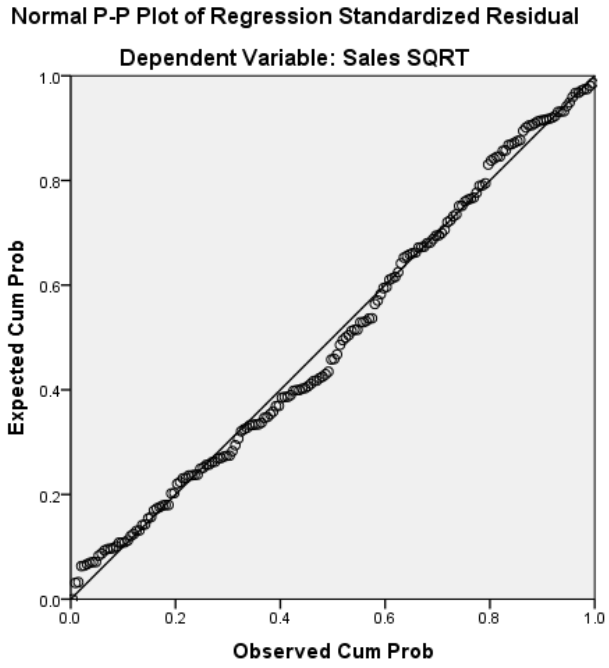
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 15.716 + 2.048 \times V12(Weather\ Index)$$

Observation:

1. R Square = 0.018 weather Index can explain Sales (SQRT) 1.8%
2. F – Value = 3.345 Significance = 0.069 which is higher than 0.05, it is not significant
3. Coefficients of Weather Index is not significant with a Significance Value of 0.069
4. There is a pattern of radiation in the residual plot.
5. The authentic values of Sales (SQRT) are distributed randomly above and below the predicted value line in the Line Fit Gragh which may bring us the sum of residuals to be zero.
6. In the Normal Probability Plot, the observation values and predicted values fit well with each other.

Variable – LPA

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPA[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.083[a]	.007	.001	3.87667

a. Predictors: (Constant), LPA

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.751	1	18.751	1.248	.265[b]
	Residual	2675.083	178	15.029		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), LPA

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.354	.481		36.091	.000
	LPA	.672	.602	.083	1.117	.265

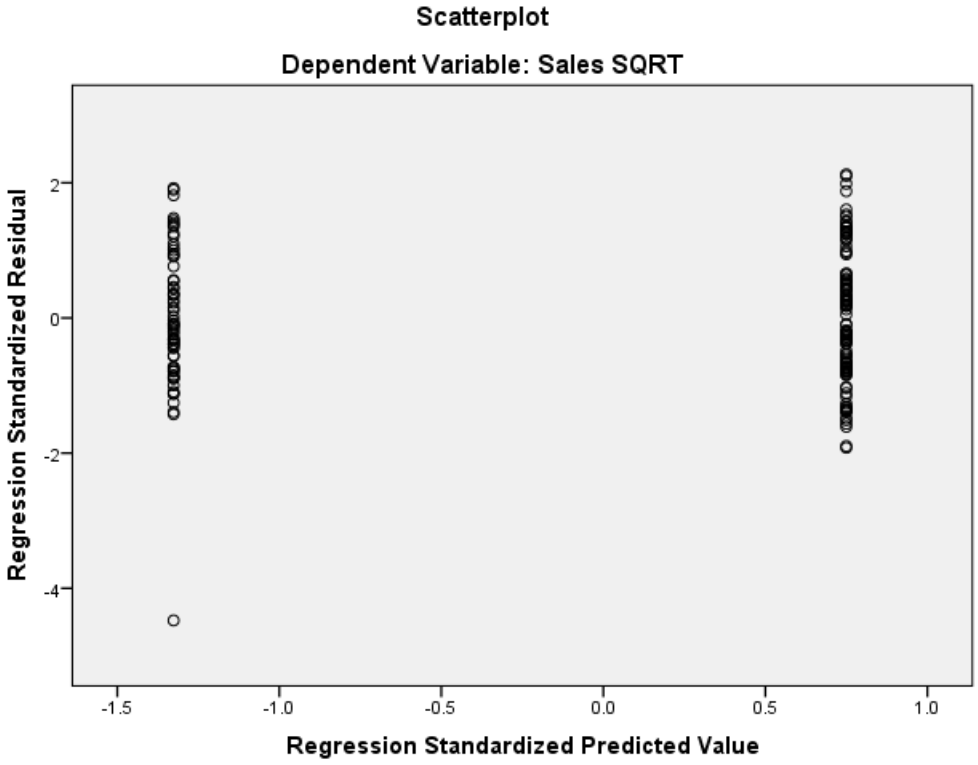
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

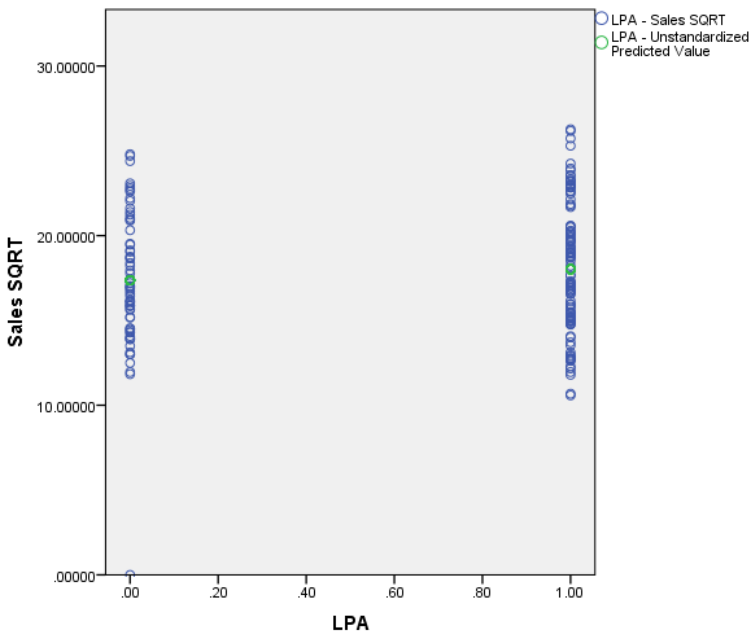
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.3543	18.0263	17.7836	.32366	180
Residual	-17.35429	8.24160	.00000	3.86582	180
Std. Predicted Value	-1.326	.750	.000	1.000	180
Std. Residual	-4.477	2.126	.000	.997	180

a. Dependent Variable: Sales SQRT

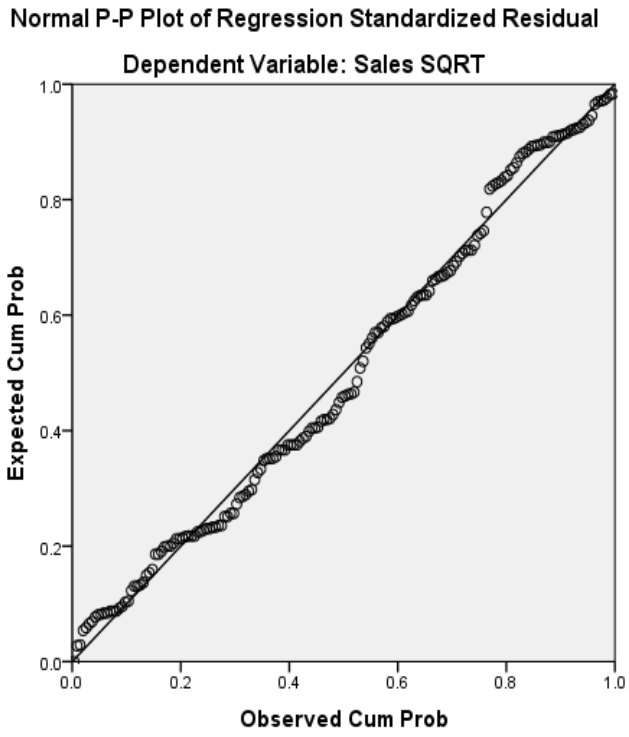
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.354 + 0.672\ LPA$$

Observation:

1. R Square = 0.7%
2. F = 1.248, Significance = 0.265 > 0.05. Not significant.
3. Coefficient of LPA is not significant with a value of 0.265
4. There is no pattern in either residual plot, line fit or normal probability plot.

Variable – LPB

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPB[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.147[a]	.022	.016	3.84777

a. Predictors: (Constant), LPB

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58.482	1	58.482	3.950	.048[b]
	Residual	2635.353	178	14.805		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), LPB

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.201	.410		41.935	.000
	LPB	1.140	.574	.147	1.987	.048

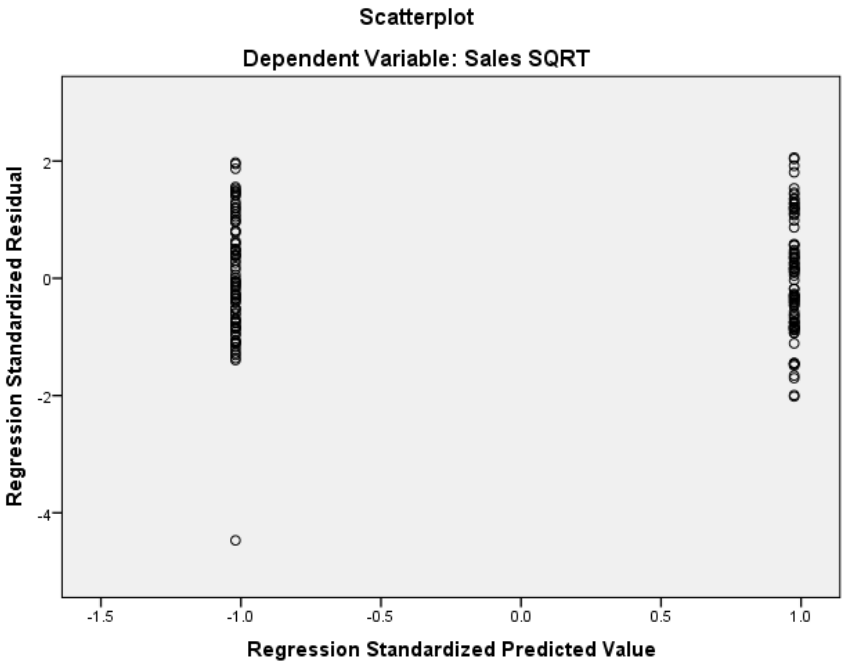
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

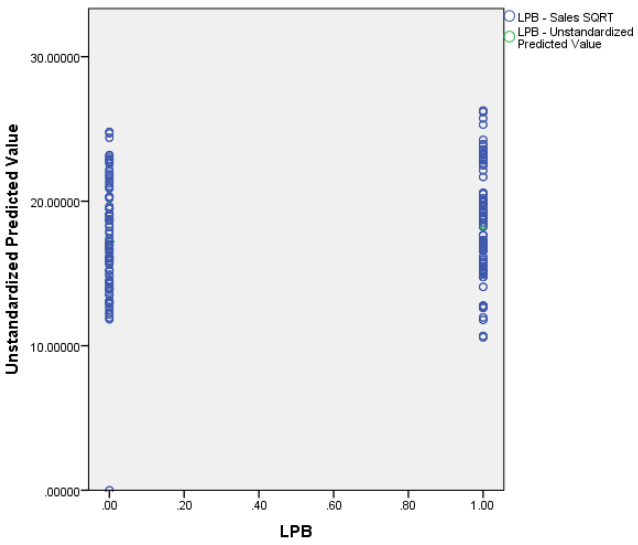
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.2008	18.3411	17.7836	.57159	180
Residual	-17.20079	7.92678	.00000	3.83701	180
Std. Predicted Value	-1.020	.975	.000	1.000	180
Std. Residual	-4.470	2.060	.000	.997	180

a. Dependent Variable: Sales SQRT

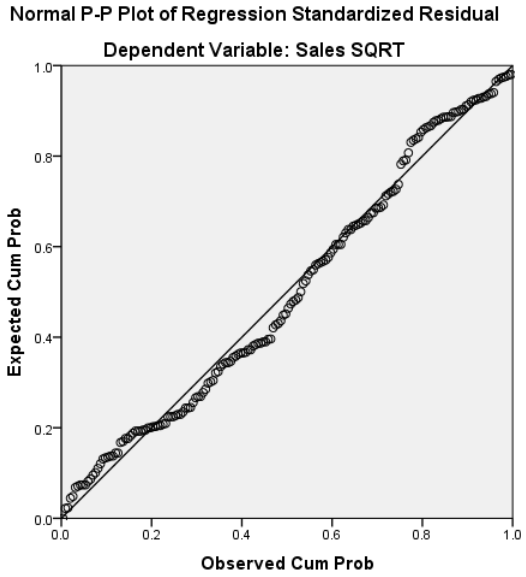
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.201 + 1.140 \times LPB$$

Observation:

1. R Square = 2.2%
2. F Value = 3.95 Significance = 0.048 < 0.05 It is a significant variable for Sales SQRT
3. Coefficient = 1.140, Significance = 0.048 < 0.05 . It is a significant variable.
4. The normal probability plot fits well. This is a binary variable so there is no pattern in line fit and residual plot.

Variable - LPC

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPC[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.094[a]	.009	.003	3.87313

a. Predictors: (Constant), LPC

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.629	1	23.629	1.575	.211[b]
	Residual	2670.206	178	15.001		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), LPC

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.335	.460		37.712	.000
	LPC	.741	.591	.094	1.255	.211

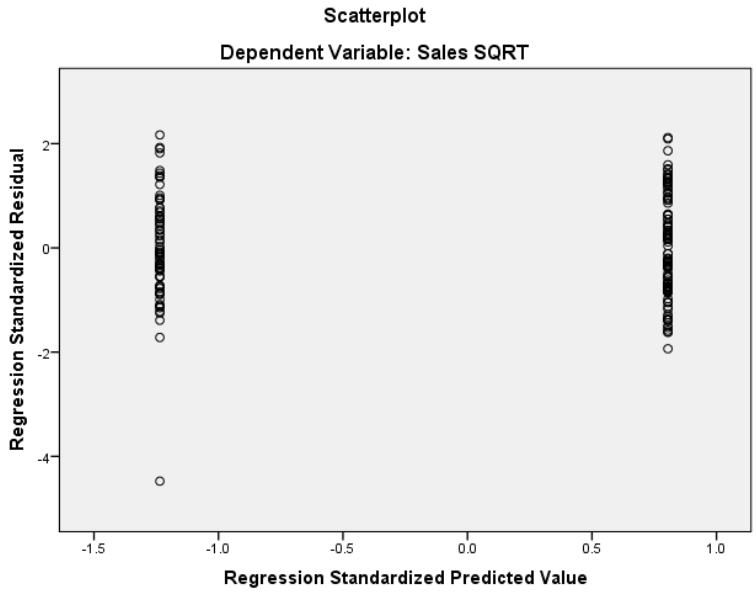
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

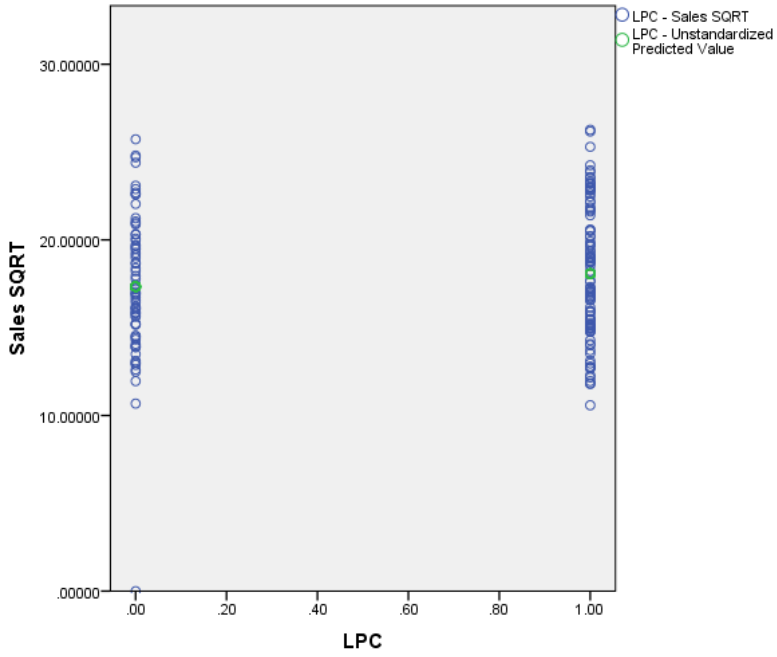
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.3347	18.0760	17.7836	.36332	180
Residual	-17.33468	8.39468	.00000	3.86230	180
Std. Predicted Value	-1.236	.805	.000	1.000	180
Std. Residual	-4.476	2.167	.000	.997	180

a. Dependent Variable: Sales SQRT

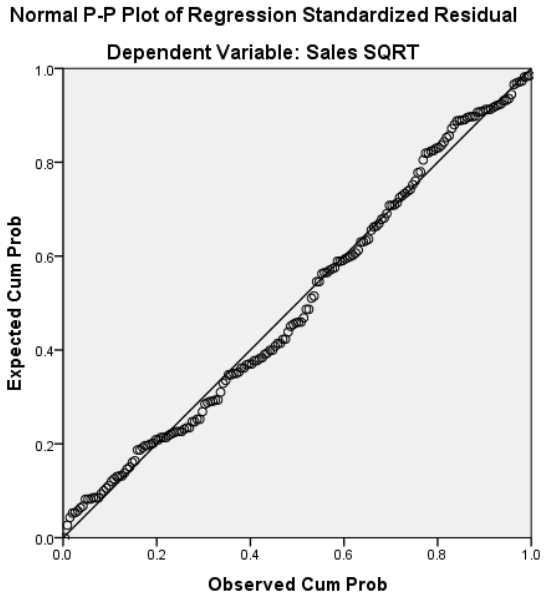
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 17.335 + 0.741 \times LPC}$$

Observation:

1. R Square = 0.9%
2. F Value = 1.575 Significance = 0.211 > 0.05. LPC is not a significant variable.
3. Coefficient = 0.741 with a significance of 0.211 which is not significant.
4. The normal probability plot fits well but this is a binary variable. No pattern can be distinguished in Line Fit or Residual Plot.

Variable – LPD

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPD[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.050[a]	.002	-.003	3.88543

a. Predictors: (Constant), LPD

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.645	1	6.645	.440	.508[b]
	Residual	2687.189	178	15.097		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), LPD

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.578	.424		41.464	.000
	LPD	.385	.580	.050	.663	.508

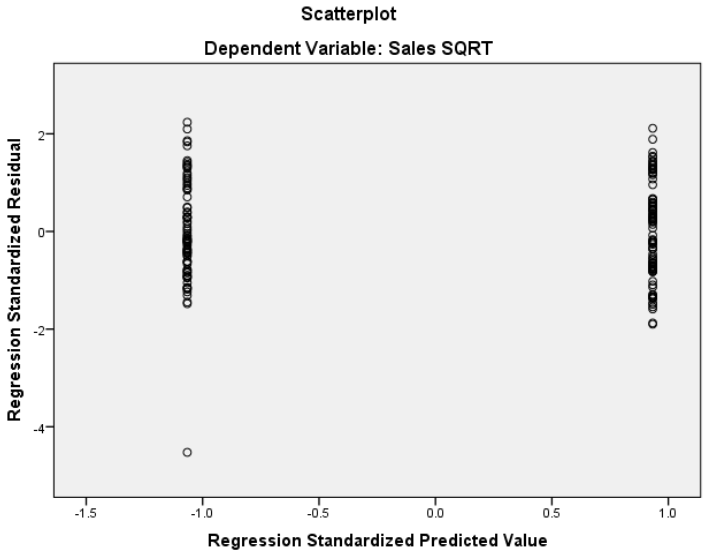
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

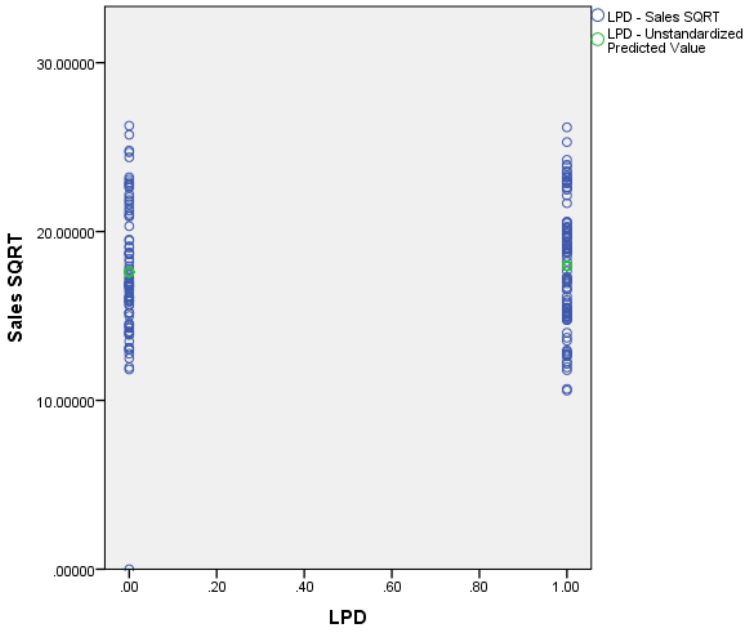
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.5782	17.9633	17.7836	.19268	180
Residual	-17.57819	8.68966	.00000	3.87456	180
Std. Predicted Value	-1.066	.933	.000	1.000	180
Std. Residual	-4.524	2.236	.000	.997	180

a. Dependent Variable: Sales SQRT

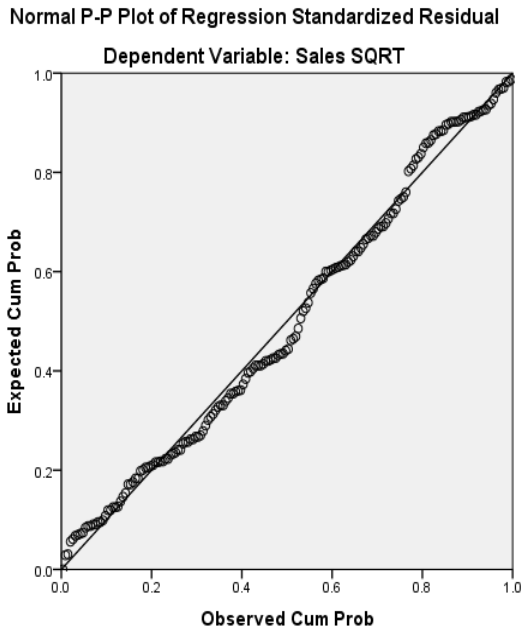
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.578 + 0.385 \times LPD$$

Observation:

1. R Square = 0.2%
2. F Value = 0.440 with a significance of 0.508. LPD is not a significant variable for the dependent variable.
3. Coefficient is 0.385 and the significance is 0.508. To the second point, it is not significant.
4. The normal probability plot fits well. This is a binary variable. No pattern is recognized in residual plot or line fit.

Variable – LPE

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPE[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.044[a]	.002	-.004	3.88642

a. Predictors: (Constant), LPE

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.272	1	5.272	.349	.555[b]
	Residual	2688.562	178	15.104		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), LPE

Coefficients[a]

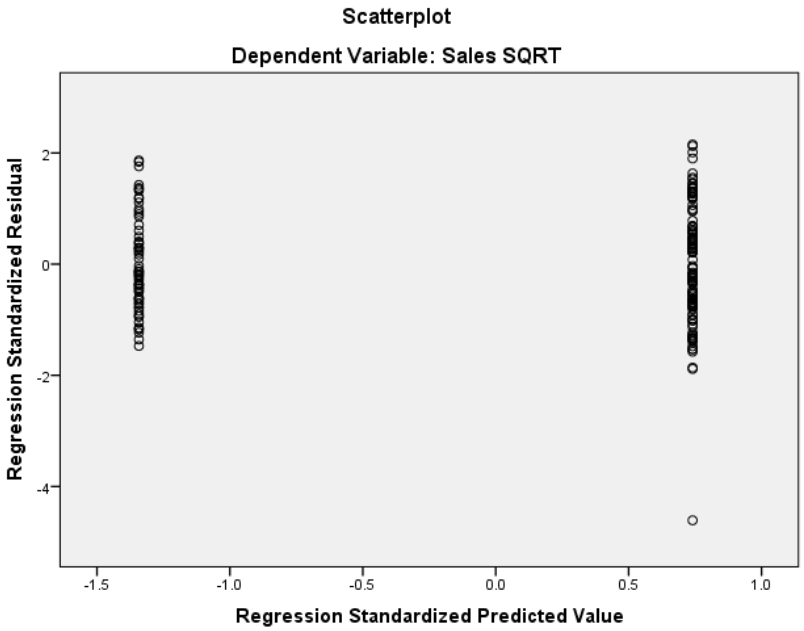
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.553	.486		36.132	.000
	LPE	.358	.605	.044	.591	.555

a. Dependent Variable: Sales SQRT

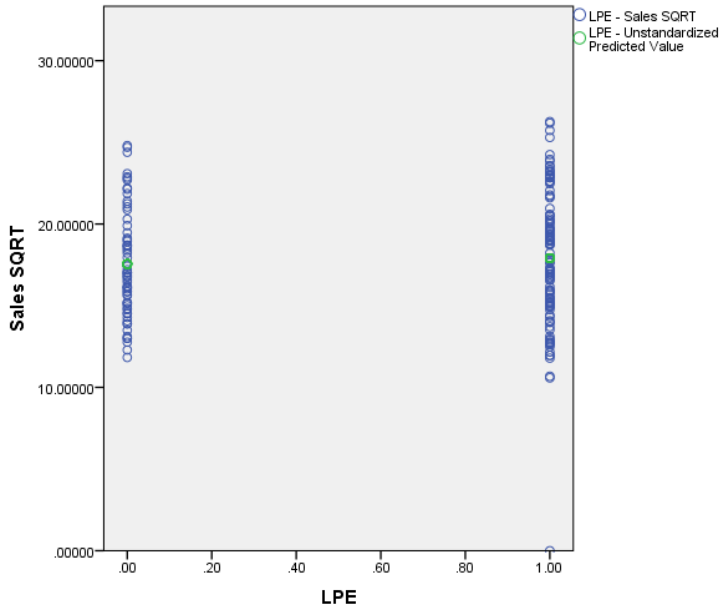
Residuals Statisticsa

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.5532	17.9107	17.7836	.17162	180
Residual	-17.91072	8.35713	.00000	3.87555	180
Std. Predicted Value	-1.343	.741	.000	1.000	180
Std. Residual	-4.609	2.150	.000	.997	180

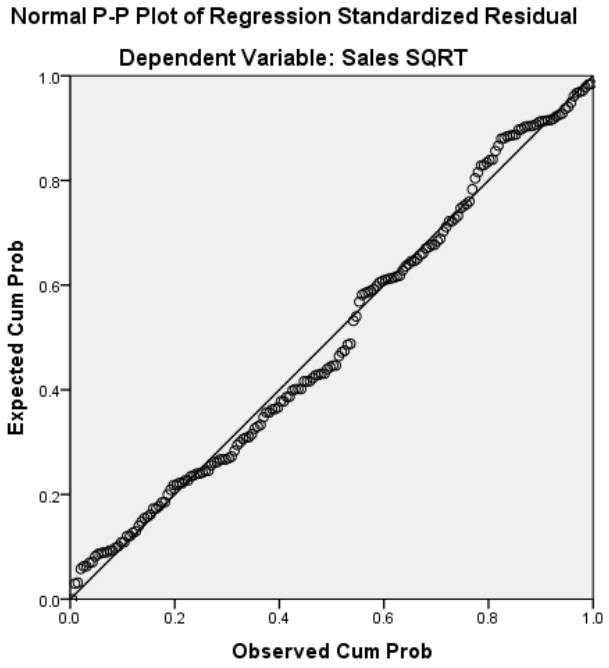
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 17.553 + 0.358 \times LPE}$$

Observation:

1. R Square = 0.2%
2. F Value = 0.349 with a significance of 0.555 LPE is not a significant variable for Sales SQRT
3. Coefficient = 0.358 Significance = 0.555 > 0.05 Not Significant
4. LPE is a binary variable. There is no pattern recognized in the plots.

Variable – Average Return Traffic SQRT

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Average Return Traffic SQRT[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.805[a]	.648	.646	2.30956

a. Predictors: (Constant), Average Return Traffic SQRT

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1744.372	1	1744.372	327.025	.000[b]
	Residual	949.462	178	5.334		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), Average Return Traffic SQRT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.639	.801		4.543	.000
	Average Return Traffic SQRT	14.364	.794	.805	18.084	.000

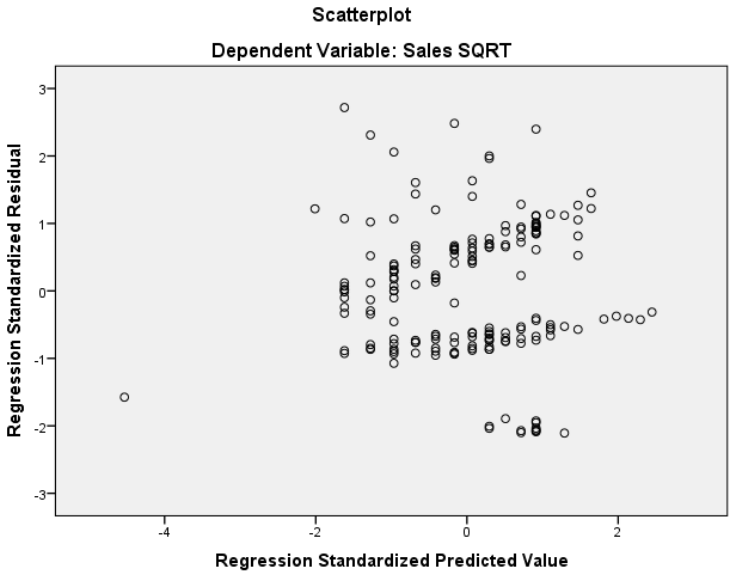
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

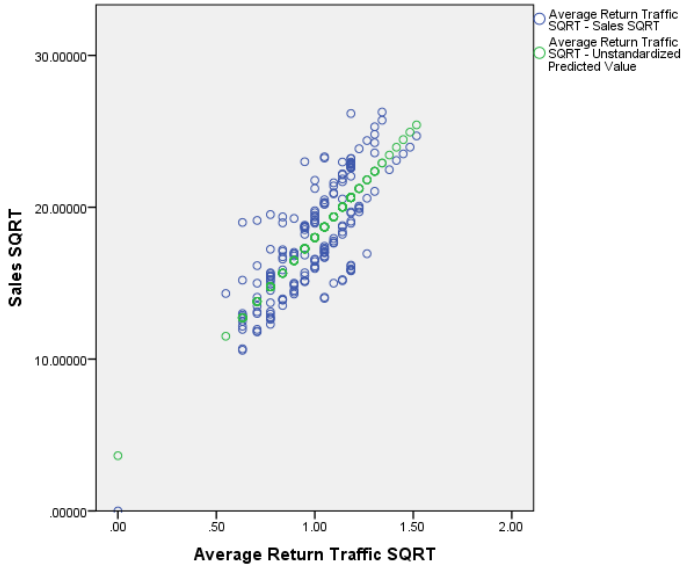
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3.6387	25.4231	17.7836	3.12171	180
Residual	-4.86707	6.27658	.00000	2.30310	180
Std. Predicted Value	-4.531	2.447	.000	1.000	180
Std. Residual	-2.107	2.718	.000	.997	180

a. Dependent Variable: Sales SQRT

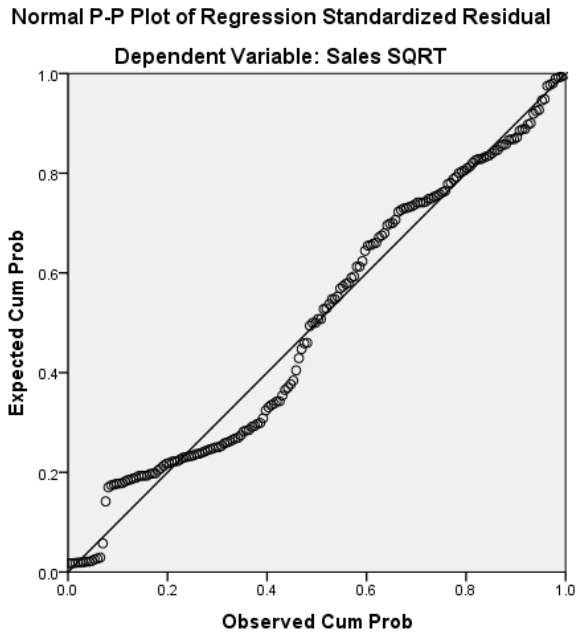
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 3.639 + 14.364 \times Average\ Return\ Traffic\ SQRT$$

Observation:

1. R Square is 64.8% which is much higher than 5%. Average Return Traffic SQRT explains much of the Sales SQRT.
2. F Value = 327.025 Significance is almost 0. It is very significant.
3. Coefficient = 14.364 with a significance close to zero.
4. Dots are distributed randomly in the Residual Plot. The observed values and predicted values fit well with each other. Although there is a little bit deviation in the normal probability plot. All in all, it is a good variable.

Variable – In Store Promotions

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	InstorePromotions[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.110[a]	.012	.006	3.86677

a. Predictors: (Constant), InstorePromotions

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	32.390	1	32.390	2.166	.143[b]
	Residual	2661.445	178	14.952		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), InstorePromotions

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.239	.469		36.764	.000
	InstorePromotions	.875	.594	.110	1.472	.143

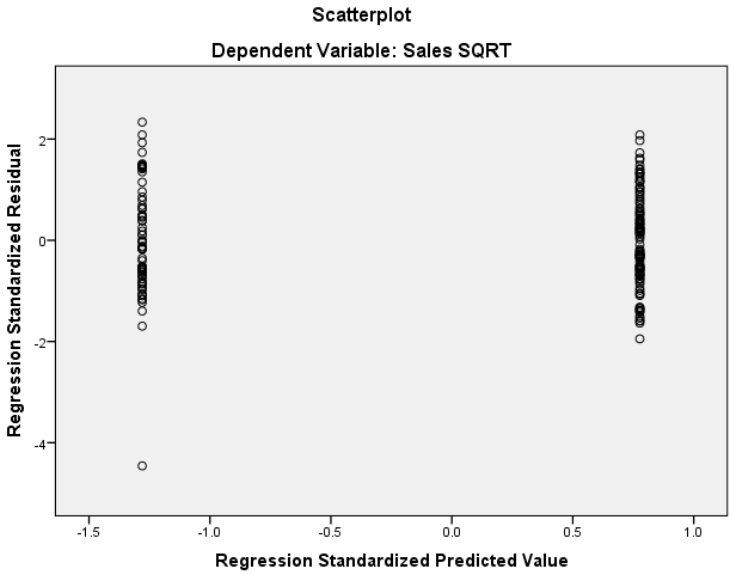
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

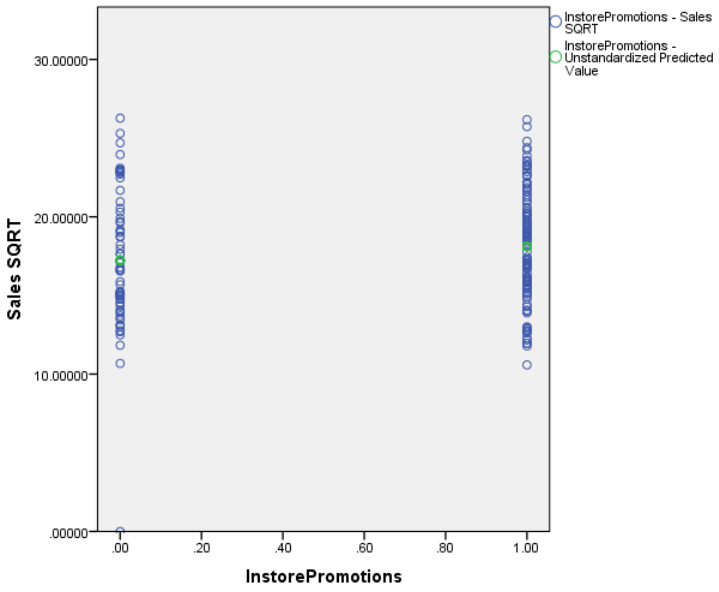
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.2392	18.1141	17.7836	.42538	180
Residual	-17.23920	9.02865	.00000	3.85596	180
Std. Predicted Value	-1.280	.777	.000	1.000	180
Std. Residual	-4.458	2.335	.000	.997	180

a. Dependent Variable: Sales SQRT

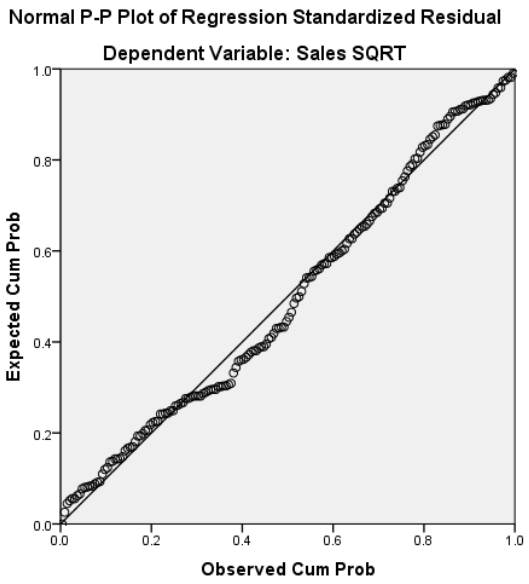
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 17.239 + 0875 \times In\ Store\ Promotions}$$

Observation:

1. R Square = 1.2%
2. F Value = 2.166 Significance = 0.143 > 0.05. Not significant .
3. Coefficient = 0.875. Significance = 0.143 > 0.05. Not significant.
4. In Store Promotions is a binary variable. There is no pattern recognized in the chart.

Variable - Daily Traffic Index Log

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Log[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.134[a]	.018	.013	3.85496

a. Predictors: (Constant), Log

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	48.628	1	48.628	3.272	.072[b]
	Residual	2645.206	178	14.861		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

b. Predictors: (Constant), Log

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.981	.307		58.490	.000
	Log	.913	.505	.134	1.809	.072

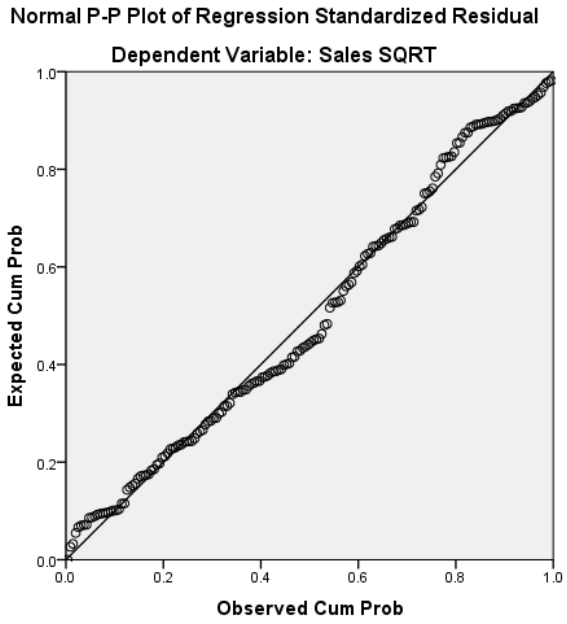
a. Dependent Variable: Sales SQRT

Residuals Statistics[a]

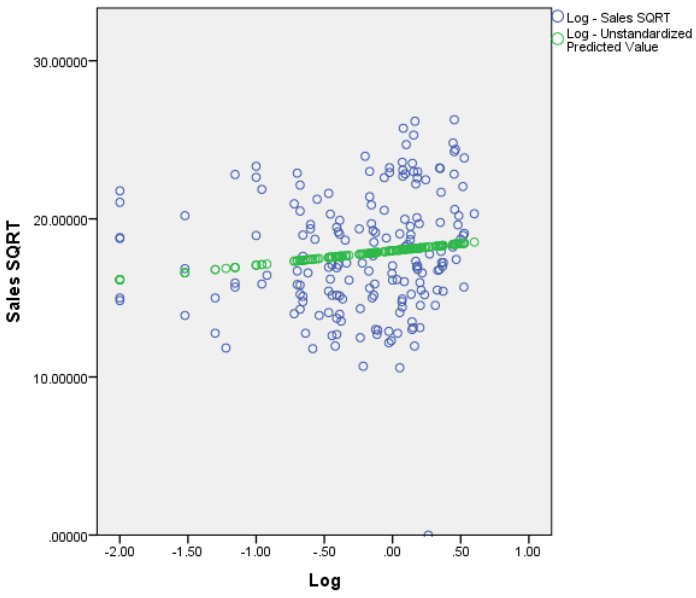
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	16.1545	18.5293	17.7836	.52122	180
Residual	-18.22109	8.04102	.00000	3.84418	180
Std. Predicted Value	-3.126	1.431	.000	1.000	180
Std. Residual	-4.727	2.086	.000	.997	180

a. Dependent Variable: Sales SQRT

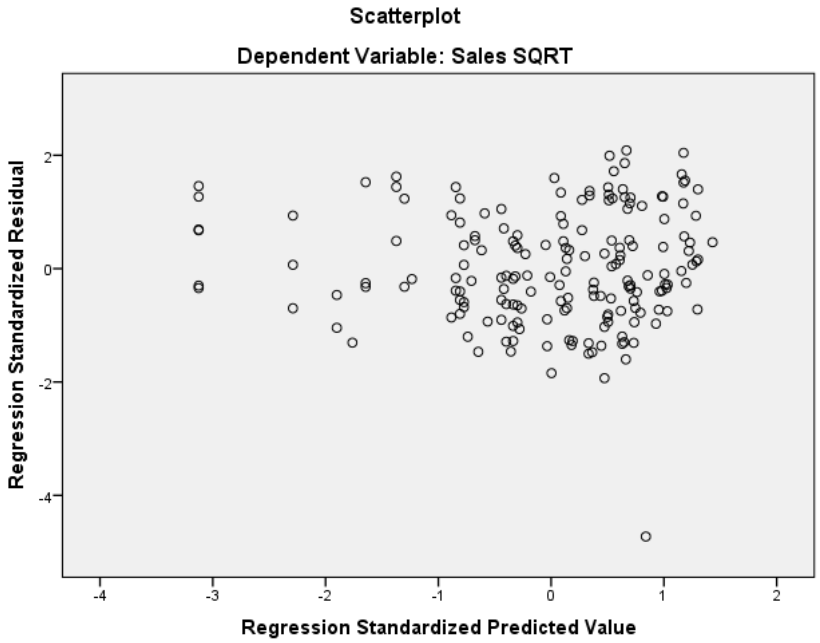
Normal Probability Plot



Line Fit



Residual Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.981 + 0.913 \times Daily\ Traffic\ Index\ Log$$

Observation:

1. R Square = 1.8%
2. F Value = 3.272 with a significance of 0.072 > 0.05. Not significant.
3. The coefficient is 0.913 with a significance of 0.072 which is not a significant variable.
4. There is no pattern in residual plot and line fit. The normal probability plot fits well.

Variable – Seasonality 1.2

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	1.2[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.086[a]	.007	.002	3.87565

a. Predictors: (Constant), 1.2

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.154	1	20.154	1.342	.248[b]
	Residual	2673.680	178	15.021		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

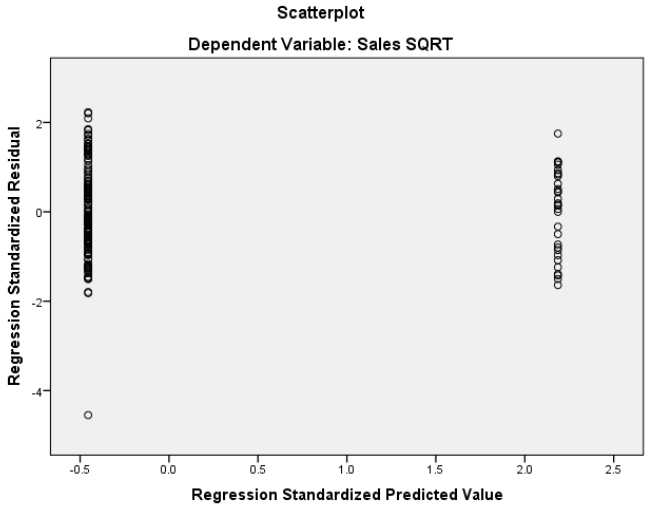
b. Predictors: (Constant), 1.2

Coefficients[a]

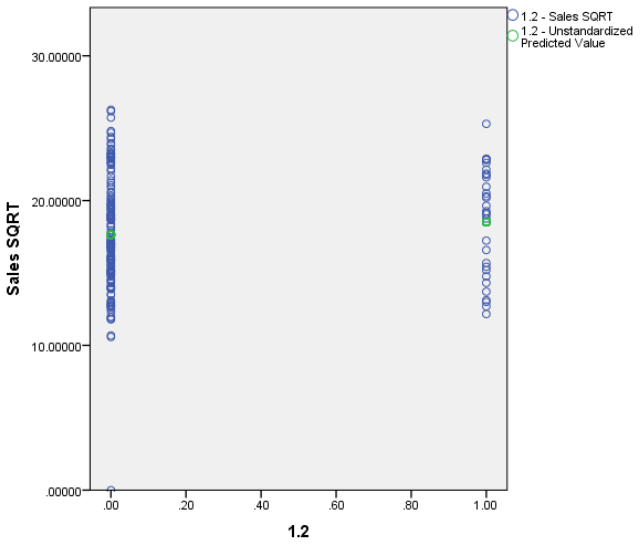
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.631	.318		55.530	.000
	1.2	.886	.765	.086	1.158	.248

a. Dependent Variable: Sales SQRT

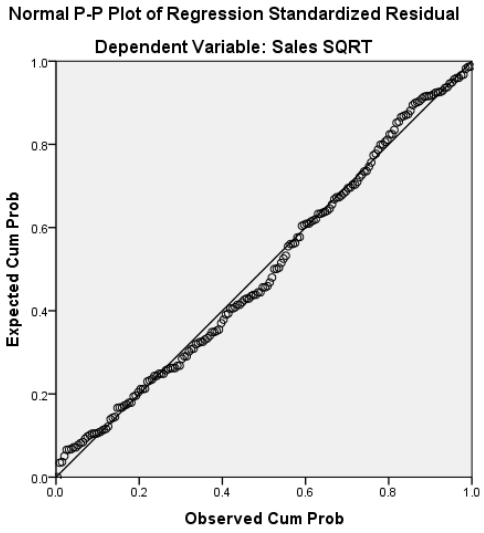
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.631 + 0.886 \times Seasonality\ 1.2$$

Observation:

1. R Square = 0.7%
2. F Value = 1.342 with a significance of 0.248 > 0.05. It is not significant.
3. Coefficient = 0.886. Significance = 0.248 > 0.05. Not significant.
4. This is a binary variable. There is no pattern recognized in the plots. The Normal Probability Plot also fits well.

Variable – Seasonality 1.1

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	1.1[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.055[a]	.003	-.003	3.88430

a. Predictors: (Constant), 1.1

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.205	1	8.205	.544	.462[b]
	Residual	2685.630	178	15.088		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

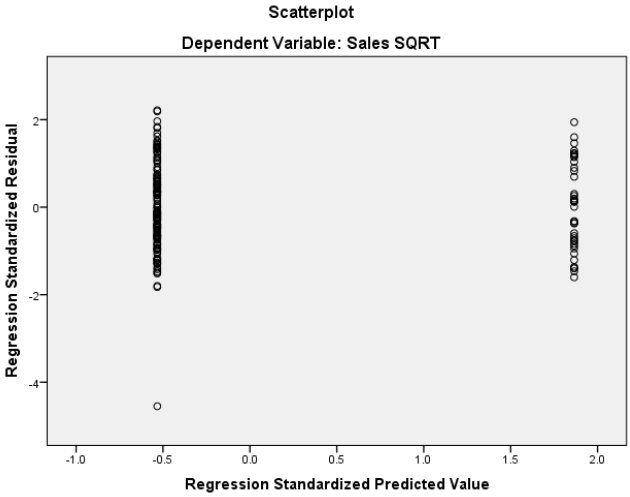
b. Predictors: (Constant), 1.1

Coefficients[a]

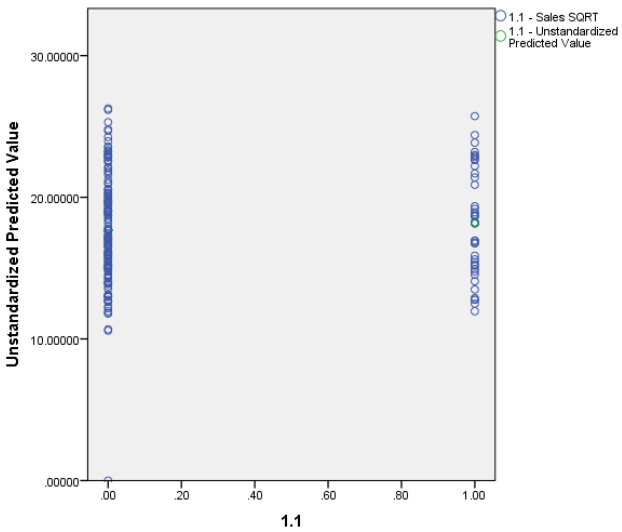
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.669	.328		53.824	.000
	1.1	.514	.696	.055	.737	.462

a. Dependent Variable: Sales SQRT

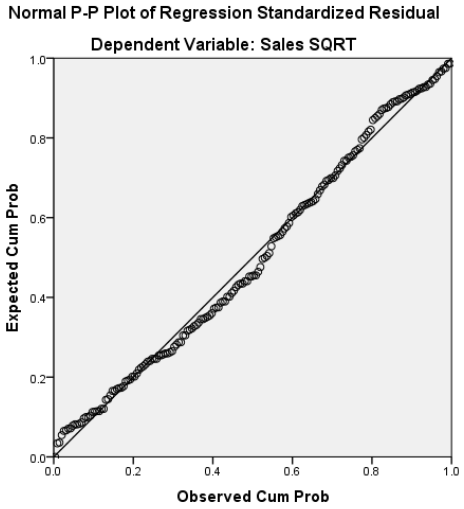
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.669 + 0.514 \times Seasonality\ 1.1$$

Observation:

1. R Square = 0.3%
2. F Value = 0.544 with significance of 0.462. It means that the variable is not significant.
3. Coefficient= 0.514 with a significance of 0.462 which is not significant.
4. It is a binary variable. There is no pattern recognized in the plots.

Variable – Seasonality 0.96

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	0.96[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.154[a]	.024	.018	3.84367

a. Predictors: (Constant), 0.96

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64.096	1	64.096	4.339	.039[b]
	Residual	2629.738	178	14.774		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

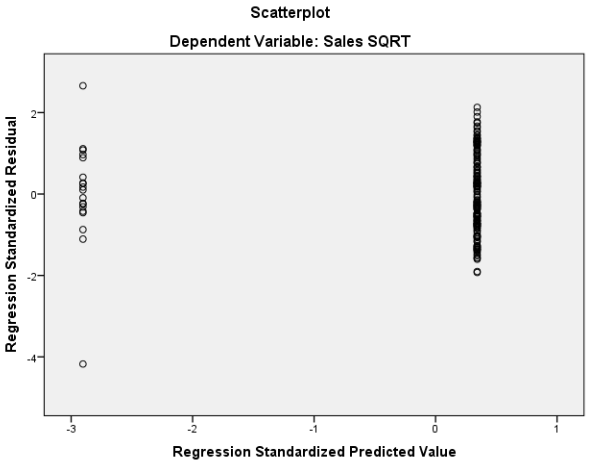
b. Predictors: (Constant), 0.96

Coefficients[a]

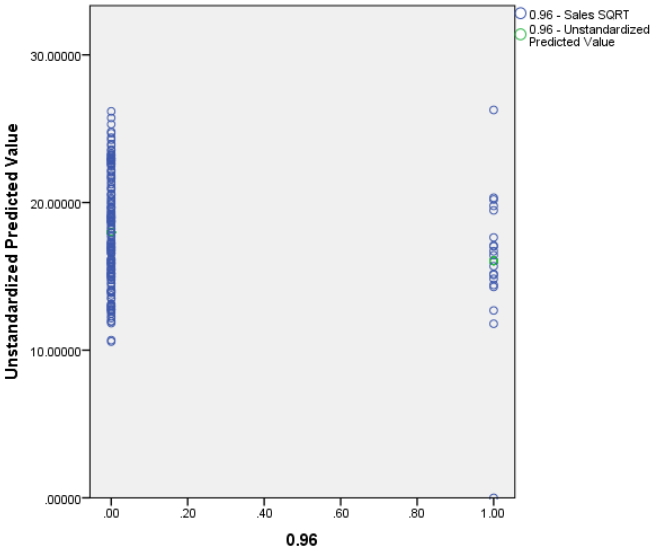
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.989	.303		59.383	.000
	0.96	-1.942	.932	-.154	-2.083	.039

a. Dependent Variable: Sales SQRT

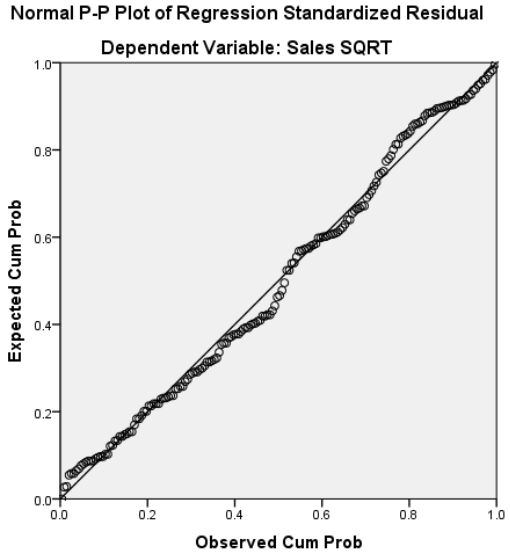
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.989 - 1.942 \times Seasonality\ 0.96$$

Observation:

1. R Square = 2.4%
2. F Value = 4.339 with a significance of 0.039 < 0.05. It is significant in explaining Sales SQRT.
3. Coefficient = -1.942 It is significant in explaining the variable.
4. It is a binary variable. There is no pattern discovered in the plots.

Variable – Seasonality 0.95

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	0.95[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.018[a]	.000	-.005	3.88957

a. Predictors: (Constant), 0.95

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.917	1	.917	.061	.806[b]
	Residual	2692.917	178	15.129		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

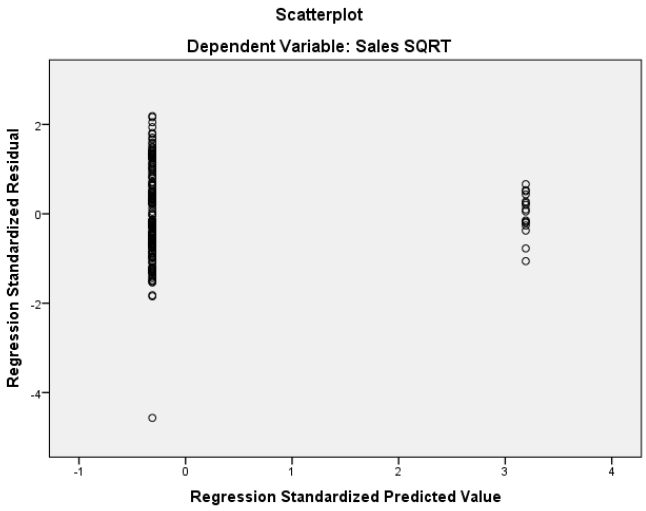
b. Predictors: (Constant), 0.95

Coefficients[a]

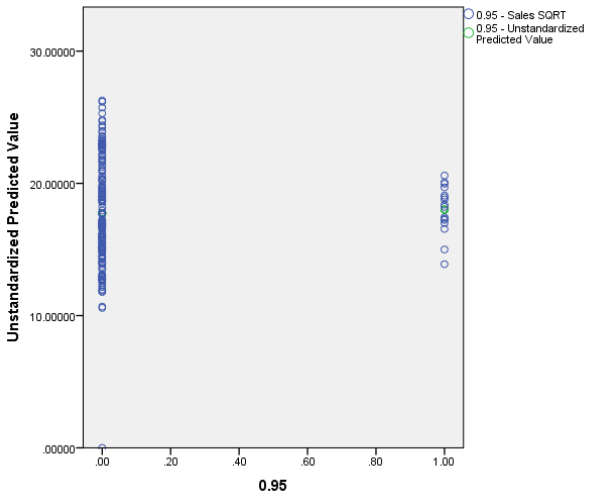
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.761	.304		58.478	.000
	0.95	.251	1.019	.018	.246	.806

a. Dependent Variable: Sales SQRT

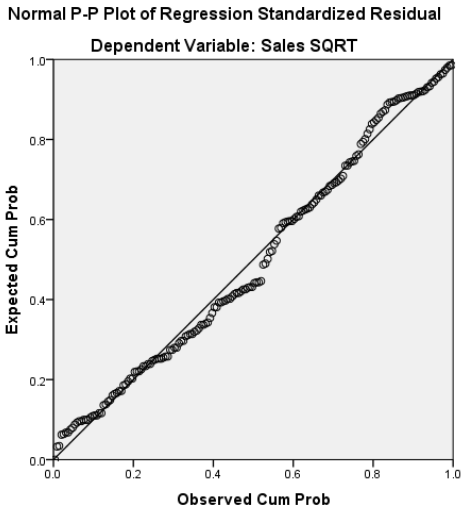
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.761 + 0.251 \times Seasonality\ 0.95$$

Observation:

1. R Square = 0
2. F Value = 0.061 Significance = 0.806 > 0.05. It is not significant.
3. Coefficient = 0.251 with a significance of 0.806 which is much higher than 0.05
4. It is binary variable. There is no pattern recognized in the plots.

Variable – Seasonality 0.86

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	0.86[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.190[a]	.036	.031	3.81969

a. Predictors: (Constant), 0.86

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	96.803	1	96.803	6.635	.011[b]
	Residual	2597.032	178	14.590		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

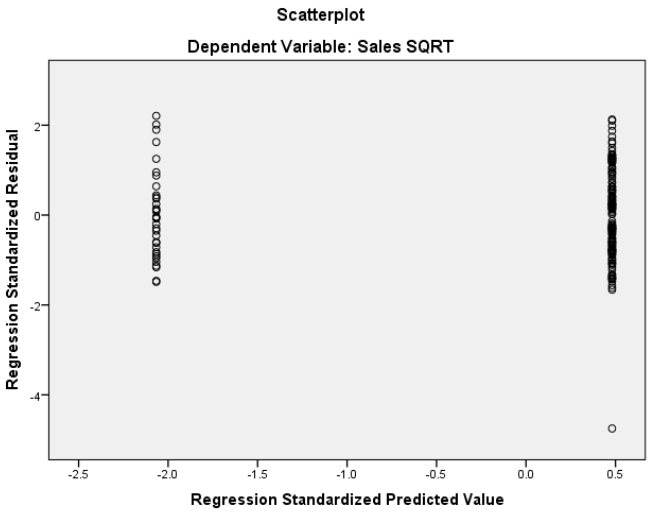
b. Predictors: (Constant), 0.86

Coefficients[a]

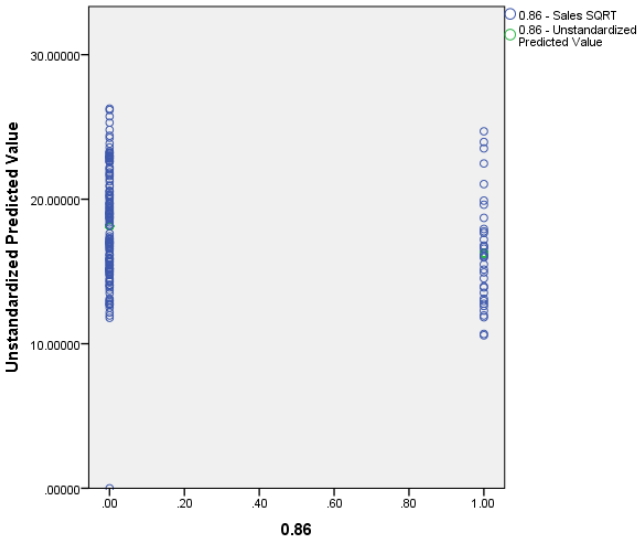
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.137	.316		57.375	.000
	0.86	-1.874	.727	-.190	-2.576	.011

a. Dependent Variable: Sales SQRT

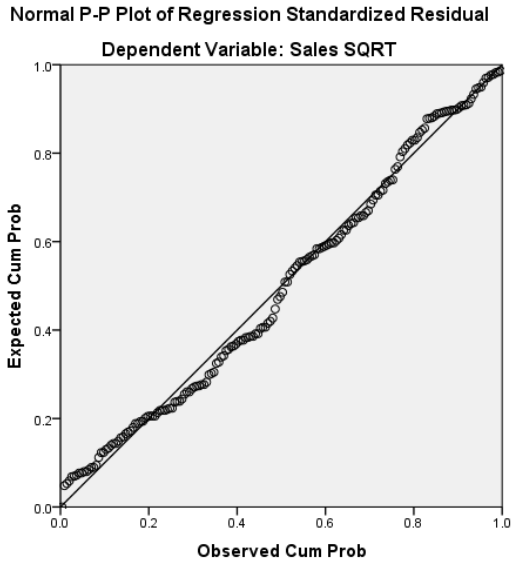
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 18.137 - 1.874 \times Seasonality\ 0.86$$

Observation:

1. R Square = 3.6%
2. F Value = 6.635 with a significance of 0.011 < 0.05
3. Coefficient = -1.874 Signifiance = 0.011 < 0.05. It is significant.
4. It is a binary variable. All of the plots fit well.

Variable – Seasonality 0.99

Model	Variables Entered	Variables Removed	Method
1	0.99[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.181[a]	.033	.027	3.82581

a. Predictors: (Constant), 0.99

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	88.475	1	88.475	6.045	.015[b]
	Residual	2605.359	178	14.637		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

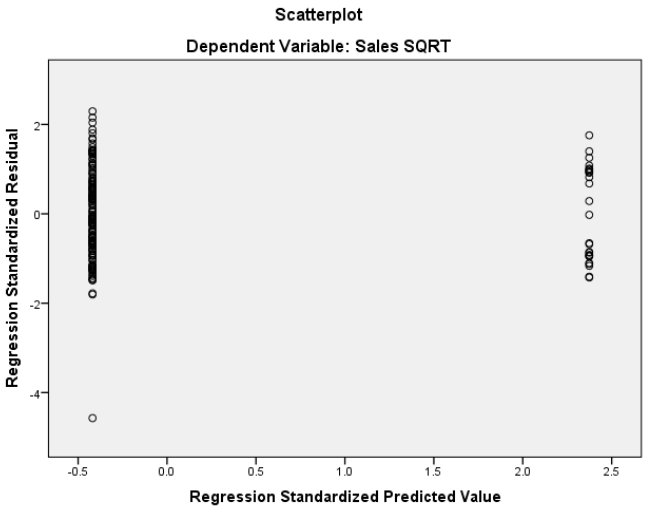
b. Predictors: (Constant), 0.99

Coefficients[a]

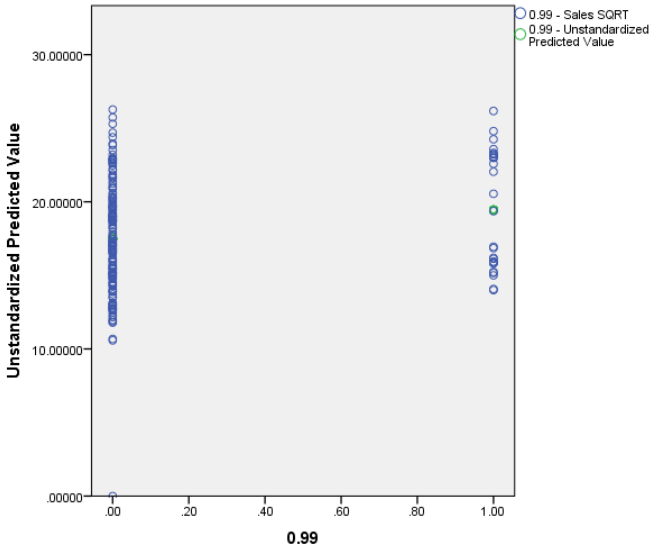
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.489	.309		56.544	.000
	0.99	1.963	.799	.181	2.459	.015

a. Dependent Variable: Sales SQRT

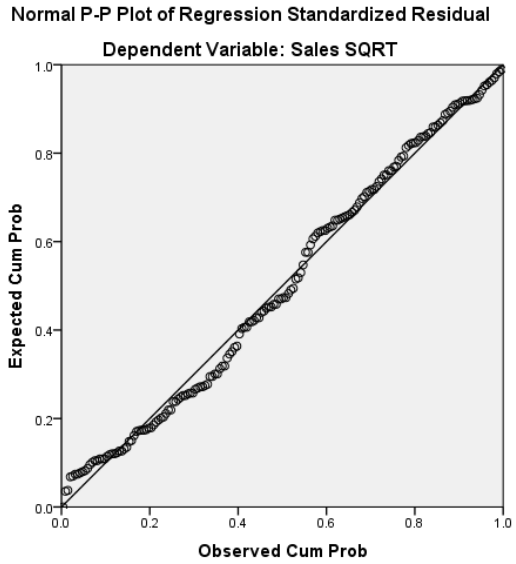
Residual Plot



Line Fit



Normal Probability Plot



Regression Model:

$$Y(Sales\ SQRT) = 17.489 + 1.963 \times Seasonality\ 0.99$$

Observation:

1. R Square = 3.3%
2. F Value = 6.045 with a significance of 0.015 < 0.05. It is significant.
3. Coefficient is 1.963 and the significance if 0.015 which is much lower than 0.05.
4. There is no pattern in the Residual Plot and Line Fit. The Normal Probability Plot fits well.

Variable – Seasonality 1.06

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	1.06[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.015[a]	.000	-.005	3.88977

a. Predictors: (Constant), 1.06

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.633	1	.633	.042	.838[b]
	Residual	2693.201	178	15.130		
	Total	2693.835	179			

a. Dependent Variable: Sales SQRT

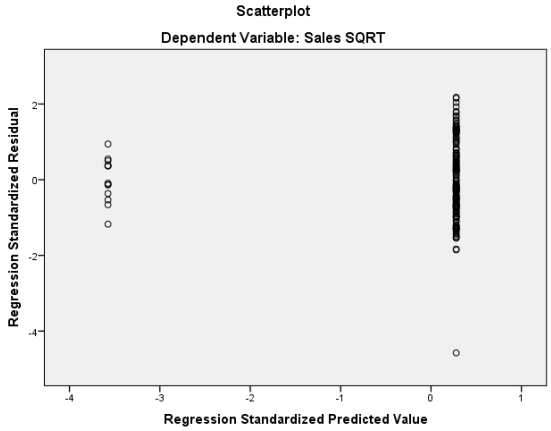
b. Predictors: (Constant), 1.06

Coefficients[a]

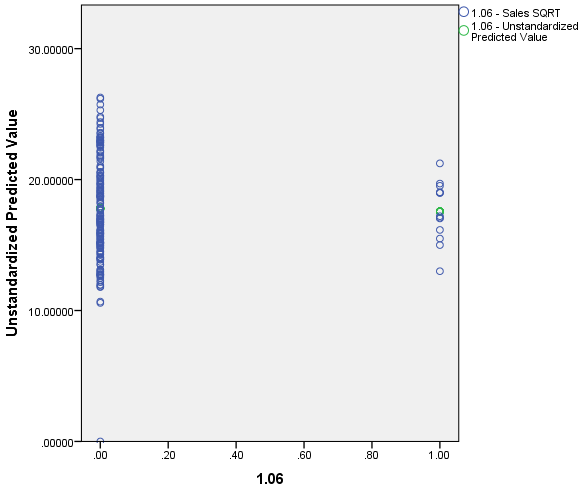
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.800	.301		59.137	.000
	1.06	-.229	1.120	-.015	-.205	.838

a. Dependent Variable: Sales SQRT

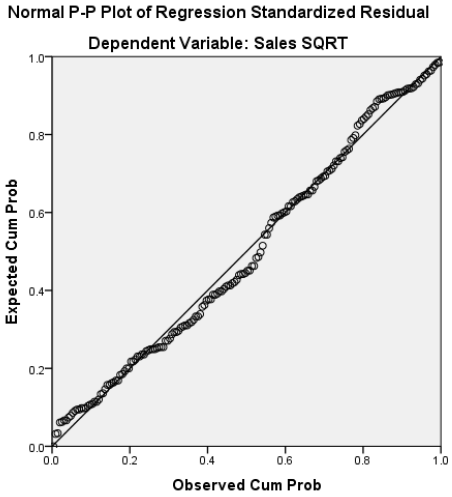
Residual Plot



Line Fi



Normal Probability Plot



Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 17.8 - 0.229 \times Seasonality\ 1.06}$$

Observation:

1. R Square = 0
2. F Value = 0.042 Significance = 0.838 > 0.05 It is not significant.
3. Coefficient = -0.229 Significance > 0.05. It is not significant.
4. It is a binary variable. There is no pattern discovered.

Comparison

Comparison of All of the Variables

Average Return Traffic SQRT, Seasonality 0.86, Seasonality 0.99, Seasonality 0.96 and LPB have higher R Squares than other variables. Their P Values indicate that they are significant variables.

DV: Sales (SQRT)						
Variable	**R Square**	**P Value(Significant or Not)**	**F Value**	**Coefficient**	**Residual Plot**	**Line Fit**
Average Return Traffic SQRT	64.80%	Significant	327.025	14.364	No Pattern	No Pattern
Seasonality 0.86	3.60%	Significant	6.635	-1.874	Binary	Binary
Seasonality 0.99	3.30%	Significant	6.045	1.963	Binary	Binary
Seasonality 0.96	2.40%	Significant	4.339	-1.942	Binary	Binary
LPB	2.20%	Significant	3.95	1.14	Binary	Binary
Weather Index	1.80%	Not Significant	3.345	2.048	Radiation	No Pattern
Daily Traffic Index Log	1.80%	Not Significant	3.272	0.913	No Pattern	No Pattern
Seasonality 1.06	1.30%	Not Significant	2.3	1.752	Binary	Binary
In-Store Promotions	1.20%	Not Significant	2.166	0.875	Binary	Binary
LPC	0.90%	Not Significant	1.575	0.741	Binary	Binary
LPA	0.70%	Not Significant	1.248	0.672	Binary	Binary
Seasonality 1.1	0.30%	Not Significant	0.544	0.514	Binary	Binary
LPD	0.20%	Not Significant	0.44	0.385	Binary	Binary
LPE	0.20%	Not Significant	0.349	0.358	Binary	Binary
Seasonality 1.2	0.70%	Not Significant	1.342	0.886	Binary	Binary
Seasonality 0.95	0.00%	Not Significant	0.061	0.251	Binary	Binary

Comparison of All of the Binary Variables

Seasonality 0.86, Seasonality 0.99, Seasonality 0.96, LPB have higher R Squares than other variables.

DV: Sales (SQRT)						
Variable	**R Square**	**P Value(Significant or Not)**	**F Value**	**Coefficient**	**Residual Plot**	**Line Fit**
Seasonality 0.86	3.60%	Significant	6.635	-1.874	Binary	Binary
Seasonality 0.99	3.30%	Significant	6.045	1.963	Binary	Binary
Seasonality 0.96	2.40%	Significant	4.339	-1.942	Binary	Binary
LPB	2.20%	Significant	3.95	1.14	Binary	Binary
Seasonality 1.06	1.30%	Not Significant	2.3	1.752	Binary	Binary
In-Store Promotions	1.20%	Not Significant	2.166	0.875	Binary	Binary
LPC	0.90%	Not Significant	1.575	0.741	Binary	Binary
LPA	0.70%	Not Significant	1.248	0.672	Binary	Binary
Seasonality 1.1	0.30%	Not Significant	0.544	0.514	Binary	Binary
LPD	0.20%	Not Significant	0.44	0.385	Binary	Binary
LPE	0.20%	Not Significant	0.349	0.358	Binary	Binary
Seasonality 1.2	0.70%	Not Significant	1.342	0.886	Binary	Binary
Seasonality 0.95	0.00%	Not Significant	0.061	0.251	Binary	Binary

Comparison of All Variables Except for Binary Variables

DV: Sales (SQRT)						
Variable	**R Square**	**P Value(Significant or Not)**	**F Value**	**Coefficient**	**Residual Plot**	**Line Fit**
Average Return Traffic SQRT	64.80%	Significant	327.025	14.364	No Pattern	No Pattern
Weather Index	1.80%	Not Significant	3.345	2.048	Radiation	No Pattern
Daily Traffic Index Log	1.80%	Not Significant	3.272	0.913	No Pattern	No Pattern

The Model

1st Model:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	LPB, 0.96, Average Return Traffic SQRT, 0.86, 0.99[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.856[a]	.732	.724	2.03626

a. Predictors: (Constant), LPB, 0.96, Average Return Traffic SQRT, 0.86, 0.99

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1972.369	5	394.474	95.138	.000[b]
	Residual	721.465	174	4.146		
	Total	2693.835	179			

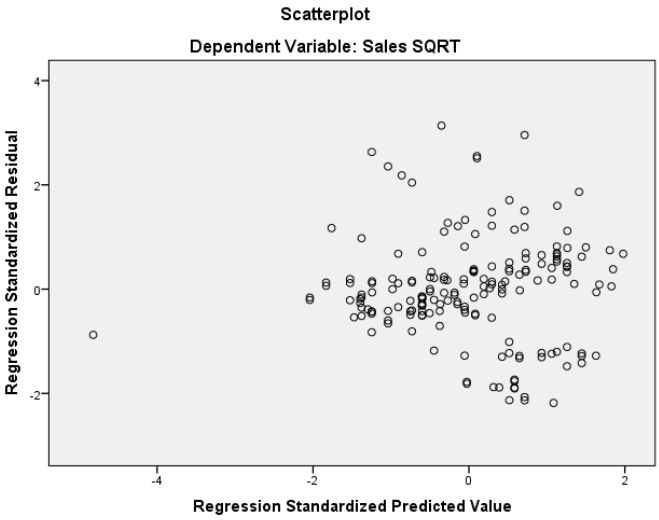
a. Dependent Variable: Sales SQRT

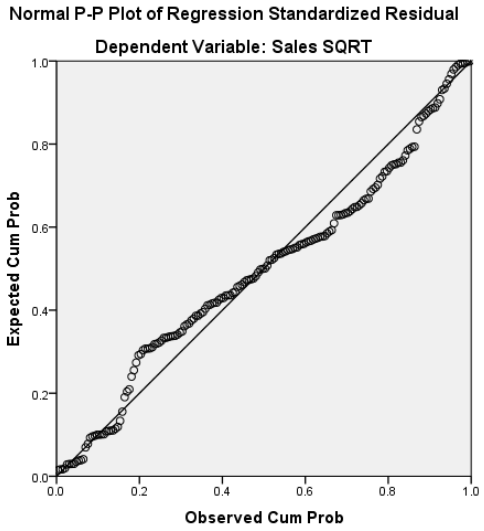
b. Predictors: (Constant), LPB, 0.96, Average Return Traffic SQRT, 0.86, 0.99

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.665	.737		4.970	.000
	Average Return Traffic SQRT	15.096	.745	.846	20.258	.000
	0.86	-2.638	.406	-.267	-6.495	.000
	0.96	-1.880	.510	-.149	-3.687	.000
	0.99	-1.805	.466	-.167	-3.871	.000
	LPB	.432	.306	.056	1.412	.160

a. Dependent Variable: Sales SQRT





Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 3.665 + 15.096 \times Average\ Return\ Traffic\ SQRT - 2.638 \times Seasonality\ 0.86 - 1.88 \times Seasonality\ 0.96 - 1.805 \times Seasonality\ 0.99 + 0.432 \times LPB}$$

Conclusion:

R Square is 73.2% which is very good. Significance of the whole model is almost zero. However, the significance of the coefficient of LPB is 0.16 which is above 0.05. LPB should be eliminated.

2nd Model:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	0.99, 0.96, 0.86, Average Return Traffic SQRT[b]	.	Enter

a. Dependent Variable: Sales SQRT

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.854[a]	.729	.723	2.04204

a. Predictors: (Constant), 0.99, 0.96, 0.86, Average Return Traffic SQRT

b. Dependent Variable: Sales SQRT

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1964.099	4	491.025	117.754	.000[b]
	Residual	729.735	175	4.170		
	Total	2693.835	179			

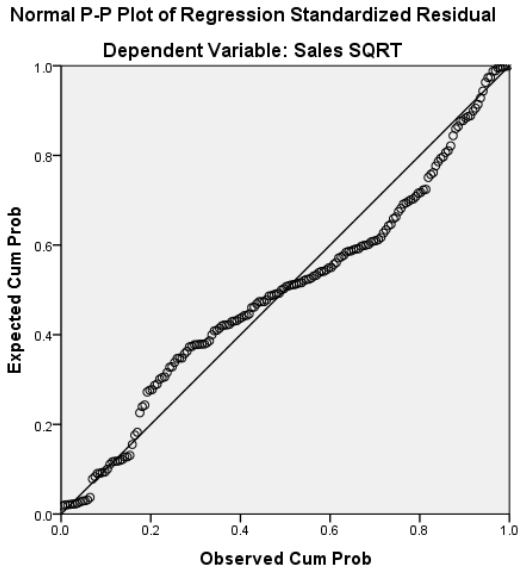
a. Dependent Variable: Sales SQRT

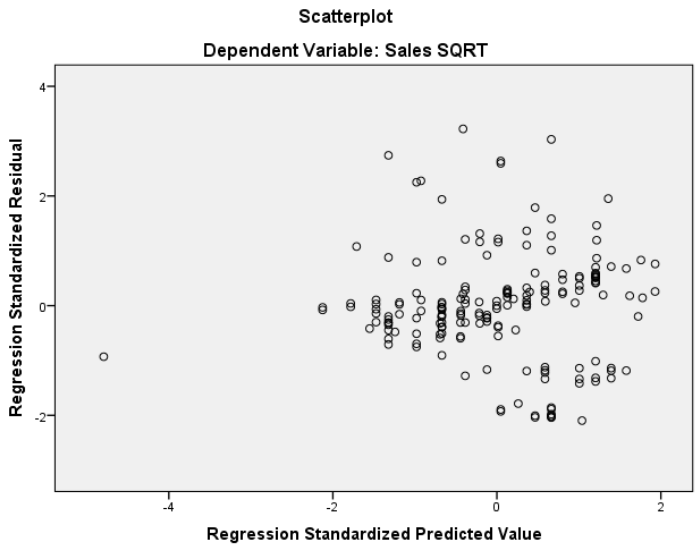
b. Predictors: (Constant), 0.99, 0.96, 0.86, Average Return Traffic SQRT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.795	.734		5.173	.000
	Average Return Traffic SQRT	15.191	.744	.851	20.412	.000
	0.86	-2.660	.407	-.269	-6.535	.000
	0.96	-1.894	.511	-.150	-3.705	.000
	0.99	-1.789	.468	-.165	-3.826	.000

a. Dependent Variable: Sales SQRT





Regression Model:

$$\boldsymbol{Y(Sales\ SQRT) = 3.795 + 15.191 \times Average\ Return\ Traffic\ SQRT\text{-}2.66 \times Seasonality\ 0.86 - 1.894 \times Seasonality\ 0.96 - 1.789 \times Seasonality\ 0.99}$$

1. R Square = 72.9%
2. 72.9% of the dependent variables can be explained by Average Traffic SQRT, Seasonality 0.86, Seasonality 0.96 and Seasonality 0.99. All of the variables combined together have a strong power in explaining the dependent variable.
3. The overall significance is close to zero which means that the combined effect of all factors is significant enough for the model.
4. Significance of each factor is close to zero. All of the variables selected are significant.
5. There is no pattern in residual plot. The Normal Probability Plot also indicates that observed values fit well with expected values.