

**OLWETHU MATIWANE**

**jimmyolwethu7@gmail.com**

**STATISTICAL ANALYSIS SYSTEM (SAS)**

## **PROJECT 2 – WALMART STORES SALES**

### **1 – Sorting**

**CODE:** /\* Importing \*/

```
FILENAME REFFILE '/folders/myfolders/Walmart_Store_sales.csv';
```

```
PROC IMPORT DATAFILE=REFFILE  
  DBMS=CSV  
  OUT=WORK.Walmart;  
  GETNAMES=YES;  
RUN;
```

### **2 - Which store has maximum sales**

**CODE:** /\* Maximum sales \*/

```
Proc Means Data=Walmart max;  
  Var weekly_sales;  
Run;
```

```
proc sql;  
  select store from walmart  
  where weekly_sales = 3818686.45;  
run;  
quit;
```

**SCREENSHOT:**

The MEANS Procedure	
Analysis Variable : Weekly_Sales	
Maximum	
3818686.45	
.	

---

Store
14

### 3 - Which store has maximum standard deviation

**CODE:** /\* Standard deviation \*/

```
Proc Means Data=Walmart stddev;
```

```
    Var store Weekly_Sales;
```

```
Run;
```

```
ODS Graphics on;
```

```
Proc corr Data=Walmart plots=matrix
```

```
    Plots=Matrix(Histogram);
```

```
    var weekly_sales store;
```

```
run;
```

```
ODS Graphics off;
```

#### SCREENSHOT:

The MEANS Procedure

Variable	Std Dev
Store	12.9881824
Weekly_Sales	564366.62

The CORR Procedure

2 Variables: Weekly\_Sales Store

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Weekly_Sales	6435	1046965	564367	6737218987	209986	3818686
Store	6435	23.00000	12.98818	148005	1.00000	45.00000

Pearson Correlation Coefficients, N = 6435  
 Prob > |r| under H0: Rho=0

	Weekly_Sales	Store
Weekly_Sales	1.00000	-0.33533 <.0001
Store	-0.33533 <.0001	1.00000

#### 4 - Which store/s has good quarterly growth rate in Q3'2012

**CODE:** /\* Quarterly growth rate \*/

Proc Means Data=Walmart P75;

Var Store Weekly\_Sales;

class store;

Run;

**SCREENSHOT:**

The MEANS Procedure			
Store	N Obs	Variable	75th Pctl
1	143	Store Weekly_Sales	1.0000000 1615524.71
2	143	Store Weekly_Sales	2.0000000 1957113.89
3	143	Store Weekly_Sales	3.0000000 420789.74
4	143	Store Weekly_Sales	4.0000000 2175563.69
5	143	Store Weekly_Sales	5.0000000 330063.08
6	143	Store Weekly_Sales	6.0000000 1627274.93
7	143	Store Weekly_Sales	7.0000000 621425.98
8	143	Store Weekly_Sales	8.0000000 929222.16
9	143	Store Weekly_Sales	9.0000000 561625.92
10	143	Store Weekly_Sales	10.0000000 1946875.06
11	143	Store Weekly_Sales	11.0000000 1399341.07
12	143	Store Weekly_Sales	12.0000000 1048617.17
13	143	Store	13.0000000
14	143	Store Weekly_Sales	14.0000000 2129771.13
15	143	Store Weekly_Sales	15.0000000 641965.20
16	143	Store Weekly_Sales	16.0000000 569005.06
17	143	Store Weekly_Sales	17.0000000 940554.34
18	143	Store Weekly_Sales	18.0000000 1135577.62
19	143	Store Weekly_Sales	19.0000000 1489613.32
20	143	Store Weekly_Sales	20.0000000 2161549.76
21	143	Store Weekly_Sales	21.0000000 773878.58
22	143	Store Weekly_Sales	22.0000000 1052973.28
23	143	Store Weekly_Sales	23.0000000 1447301.24
24	143	Store Weekly_Sales	24.0000000 1409705.03
25	143	Store Weekly_Sales	25.0000000 719591.13
26	143	Store Weekly_Sales	26.0000000 1056282.91
27	143	Store Weekly_Sales	27.0000000 1873812.93
28	143	Store Weekly_Sales	28.0000000 1451740.57
29	143	Store Weekly_Sales	29.0000000 552338.76
30	143	Store Weekly_Sales	30.0000000 452163.93
31	143	Store Weekly_Sales	31.0000000 1425559.02
32	143	Store Weekly_Sales	32.0000000 1187051.07
33	143	Store Weekly_Sales	33.0000000 275749.56
34	143	Store Weekly_Sales	34.0000000 979730.78
35	143	Store Weekly_Sales	35.0000000 1020651.74
36	143	Store Weekly_Sales	36.0000000 427175.03
37	143	Store Weekly_Sales	37.0000000 530367.83
38	143	Store Weekly_Sales	38.0000000 415513.97
39	143	Store Weekly_Sales	39.0000000 1522421.07
40	143	Store Weekly_Sales	40.0000000 1001943.80
41	143	Store Weekly_Sales	41.0000000 1332594.07
42	143	Store Weekly_Sales	42.0000000 593162.53
43	143	Store Weekly_Sales	43.0000000 660632.05
44	143	Store Weekly_Sales	44.0000000 320021.10
45	143	Store Weekly_Sales	45.0000000 802253.41

5 - Holidays which have higher sales than the mean sales in non-holiday season for all stores together

```
CODE: Proc Means Data=Walmart;  
BY Store;  
Var weekly_sales;  
class Holiday_Flag;  
RUN;
```

```
Proc glm Data=Walmart;  
Model weekly_sales=holiday_flag;  
run;
```

SCREENSHOT:

The GLM Procedure					
Dependent Variable: Weekly_Sales					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.7889697E12	2.7889697E12	8.77	0.0031
Error	6433	2.0465023E15	318125654857		
Corrected Total	6434	2.0492913E15			

R-Square	Coeff Var	Root MSE	Weekly_Sales Mean
0.001361	53.87251	564026.3	1046965

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Holiday_Flag	1	2.7889697E12	2.7889697E12	8.77	0.0031

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Holiday_Flag	1	2.7889697E12	2.7889697E12	8.77	0.0031

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	1041256.380	7290.66712	142.82	<.0001
Holiday_Flag	81631.512	27569.90618	2.96	0.0031

## 6 - Statistical Model – Linear Regression Model

### CODE:

```
/*Ho sales affect the unemployment,fule price and CPI*/  
/*H1 Sale doesnt affect the above mentioned*/  
/* Alpha= 0.05*/ /*IF pvalue >0.05 Reject H0*/
```

```
Proc Reg Data=Walmart_Store_sales;  
    Model weekly_sales=unemployment fuel_price cpi;  
Run;
```

### SCREENSHOT:

The REG Procedure					
Model: MODEL1					
Dependent Variable: Weekly_Sales					
Number of Observations Read			6435		
Number of Observations Used			6435		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4.830684E13	1.610228E13	51.75	<.0001
Error	6431	2.000984E15	3.111467E11		
Corrected Total	6434	2.049291E15			

Root MSE		557805	R-Square	0.0236
Dependent Mean		1046965	Adj R-Sq	0.0231
Coeff Var		53.27832		

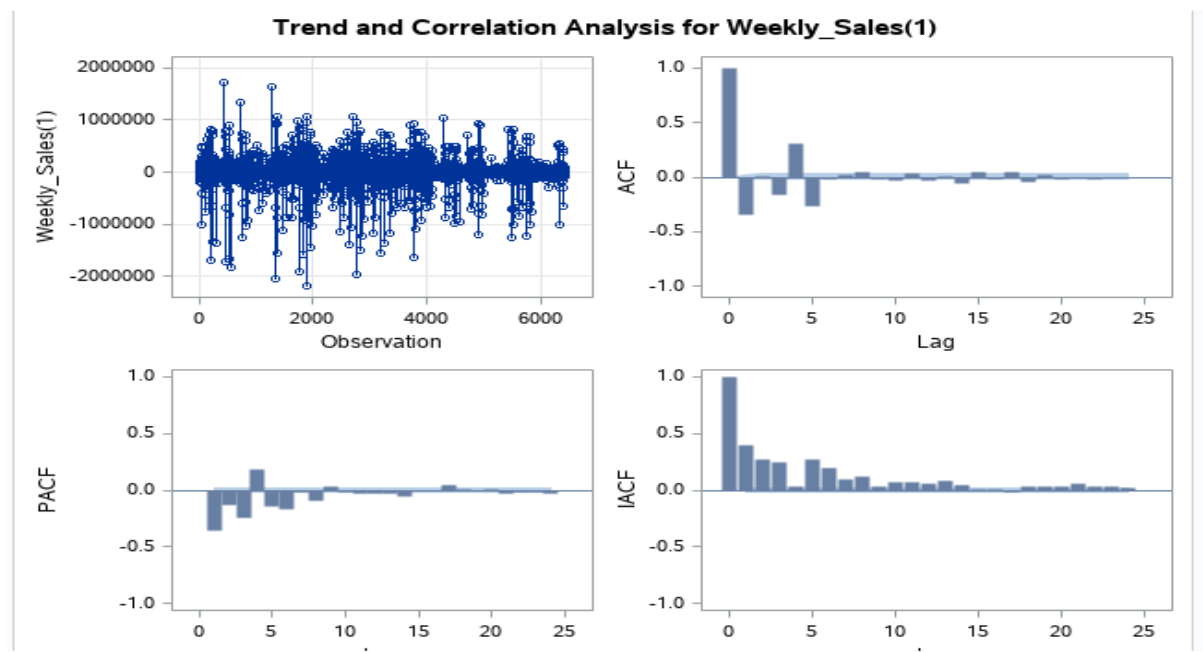
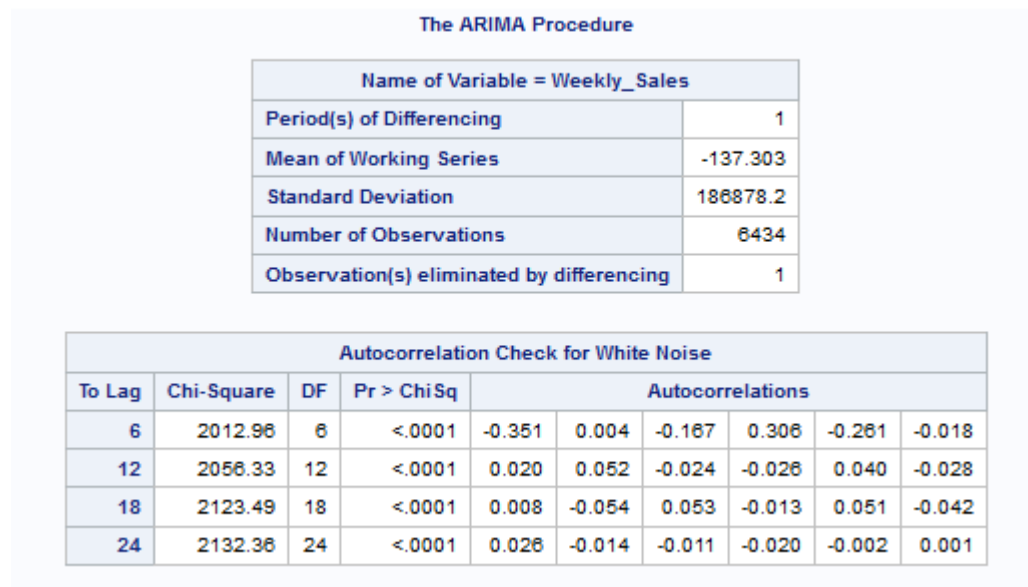
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	1745657	79573	21.94	<.0001
Unemployment	1	-42859	3905.19701	-10.97	<.0001
Fuel_Price	1	-19266	15441	-1.25	0.2122
CPI	1	-1696.87596	188.79298	-8.99	<.0001

## 7 - ARIMA MODEL and Differencing

### CODE:

```
proc arima data = Walmart_Store_sales;  
    identify var = weekly_sales(1) nlag = 24 ;  
    estimate p = 1 q = 1;  
  
run;  
  
forecast lead=12 interval=month id=Date out=Walmart_Store_sales;  
  
quit
```

### SCREENSHOT:



Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-138.96211	1615.8	-0.09	0.9315	0
AR1,1	-0.35060	0.01168	-30.02	<.0001	1

Constant Estimate	-187.682
Variance Estimate	3.064E10
Std Error Estimate	175043.6
AIC	173613.6
SBC	173627.1
Number of Residuals	6434

\* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	0.000
AR1,1	0.000	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1043.36	5	<.0001	-0.048	-0.202	-0.090	0.220	-0.219	-0.119
12	1095.57	11	<.0001	0.039	0.066	-0.020	-0.027	0.029	-0.017
18	1142.95	17	<.0001	-0.023	-0.045	0.041	0.025	0.044	-0.023
24	1156.56	23	<.0001	0.011	-0.012	-0.028	-0.030	-0.010	0.009
30	1193.46	29	<.0001	0.034	0.027	-0.004	0.034	0.036	-0.037
36	1238.34	35	<.0001	-0.066	-0.008	-0.006	-0.005	0.050	-0.008
42	1282.76	41	<.0001	0.028	0.014	-0.047	-0.040	0.043	0.016
48	1680.01	47	<.0001	-0.029	0.075	0.062	-0.018	-0.202	0.100