The CONTENTS Procedure

Data Set Name	BANPROJS.WALMART_TRAIN	**Observations**	421570
Member Type	DATA	**Variables**	5
Engine	V9	**Indexes**	0
Created	07/14/2021 09:39:38	**Observation Length**	40
Last Modified	07/14/2021 09:39:38	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	130
First Data Page	1
Max Obs per Page	3265
Obs in First Data Page	3204
Number of Data Set Repairs	0
Filename	/home/u54770142/walmart_train.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	17273876427
Access Permission	rw-r--r--
Owner Name	u54770142
File Size	16MB
File Size (bytes)	17170432

Variables in Creation Order					
#	**Variable**	**Type**	**Len**	**Format**	**Informat**
1	Store	Num	8	BEST12.	BEST32.
2	Dept	Num	8	BEST12.	BEST32.
3	Date	Num	8	YYMMDD10.	YYMMDD10.
4	Weekly_Sales	Num	8	BEST12.	BEST32.
5	IsHoliday	Char	5	$5.	$5.

Listing the first 50 Observations of WALMART_FEATURES Data Set

Obs	Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	1	2010-02-05	24924.5	FALSE
2	1	1	2010-02-12	46039.49	TRUE
3	1	1	2010-02-19	41595.55	FALSE
4	1	1	2010-02-26	19403.54	FALSE
5	1	1	2010-03-05	21827.9	FALSE
6	1	1	2010-03-12	21043.39	FALSE
7	1	1	2010-03-19	22136.64	FALSE
8	1	1	2010-03-26	26229.21	FALSE
9	1	1	2010-04-02	57258.43	FALSE
10	1	1	2010-04-09	42960.91	FALSE

Obs	Store	Dept	Date	Weekly_Sales	IsHoliday
11	1	1	2010-04-16	17596.96	FALSE
12	1	1	2010-04-23	16145.35	FALSE
13	1	1	2010-04-30	16555.11	FALSE
14	1	1	2010-05-07	17413.94	FALSE
15	1	1	2010-05-14	18926.74	FALSE
16	1	1	2010-05-21	14773.04	FALSE
17	1	1	2010-05-28	15580.43	FALSE
18	1	1	2010-06-04	17558.09	FALSE
19	1	1	2010-06-11	16637.62	FALSE
20	1	1	2010-06-18	16216.27	FALSE
21	1	1	2010-06-25	16328.72	FALSE
22	1	1	2010-07-02	16333.14	FALSE
23	1	1	2010-07-09	17688.76	FALSE
24	1	1	2010-07-16	17150.84	FALSE
25	1	1	2010-07-23	15360.45	FALSE
26	1	1	2010-07-30	15381.82	FALSE
27	1	1	2010-08-06	17508.41	FALSE
28	1	1	2010-08-13	15536.4	FALSE
29	1	1	2010-08-20	15740.13	FALSE
30	1	1	2010-08-27	15793.87	FALSE
31	1	1	2010-09-03	16241.78	FALSE
32	1	1	2010-09-10	18194.74	TRUE
33	1	1	2010-09-17	19354.23	FALSE
34	1	1	2010-09-24	18122.52	FALSE
35	1	1	2010-10-01	20094.19	FALSE
36	1	1	2010-10-08	23388.03	FALSE
37	1	1	2010-10-15	26978.34	FALSE
38	1	1	2010-10-22	25543.04	FALSE
39	1	1	2010-10-29	38640.93	FALSE
40	1	1	2010-11-05	34238.88	FALSE
41	1	1	2010-11-12	19549.39	FALSE
42	1	1	2010-11-19	19552.84	FALSE
43	1	1	2010-11-26	18820.29	TRUE
44	1	1	2010-12-03	22517.56	FALSE
45	1	1	2010-12-10	31497.65	FALSE
46	1	1	2010-12-17	44912.86	FALSE
47	1	1	2010-12-24	55931.23	FALSE
48	1	1	2010-12-31	19124.58	TRUE
49	1	1	2011-01-07	15984.24	FALSE
50	1	1	2011-01-14	17359.7	FALSE

Checking Missing Values for IsHoliday Variable

The FREQ Procedure

IsHoliday	Frequency	Percent
Nonmissing	421570	100.00

Checking Missing Values for Numeric Variables in WALMART_TRAIN dataset

The MEANS Procedure

Variable	N	N Miss	Mean	Median	Minimum	Maximum
Store	421570	0	22.2005456	22.0000000	1.0000000	45.0000000
Dept	421570	0	44.2603174	37.0000000	1.0000000	99.0000000
Date	421570	0	18796.35	18795.00	18298.00	19292.00
Weekly_Sales	421570	0	15981.26	7612.03	-4988.94	693099.36

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Variable: Store

Moments			
N	421570	Sum Weights	421570
Mean	22.2005456	Sum Observations	9359084
Std Deviation	12.7852974	Variance	163.463829
Skewness	0.0777625	Kurtosis	-1.1465028
Uncorrected SS	276688054	Corrected SS	68911283.1
Coeff Variation	57.5900144	Std Error Mean	0.01969137

Basic Statistical Measures			
Location		Variability	
Mean	22.20055	Std Deviation	12.78530
Median	22.00000	Variance	163.46383
Mode	13.00000	Range	44.00000
		Interquartile Range	22.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	1127.425	Pr > |t|	<.0001
Sign	M	210785	Pr >= |M|	<.0001
Signed Rank	S	4.443E10	Pr >= |S|	<.0001

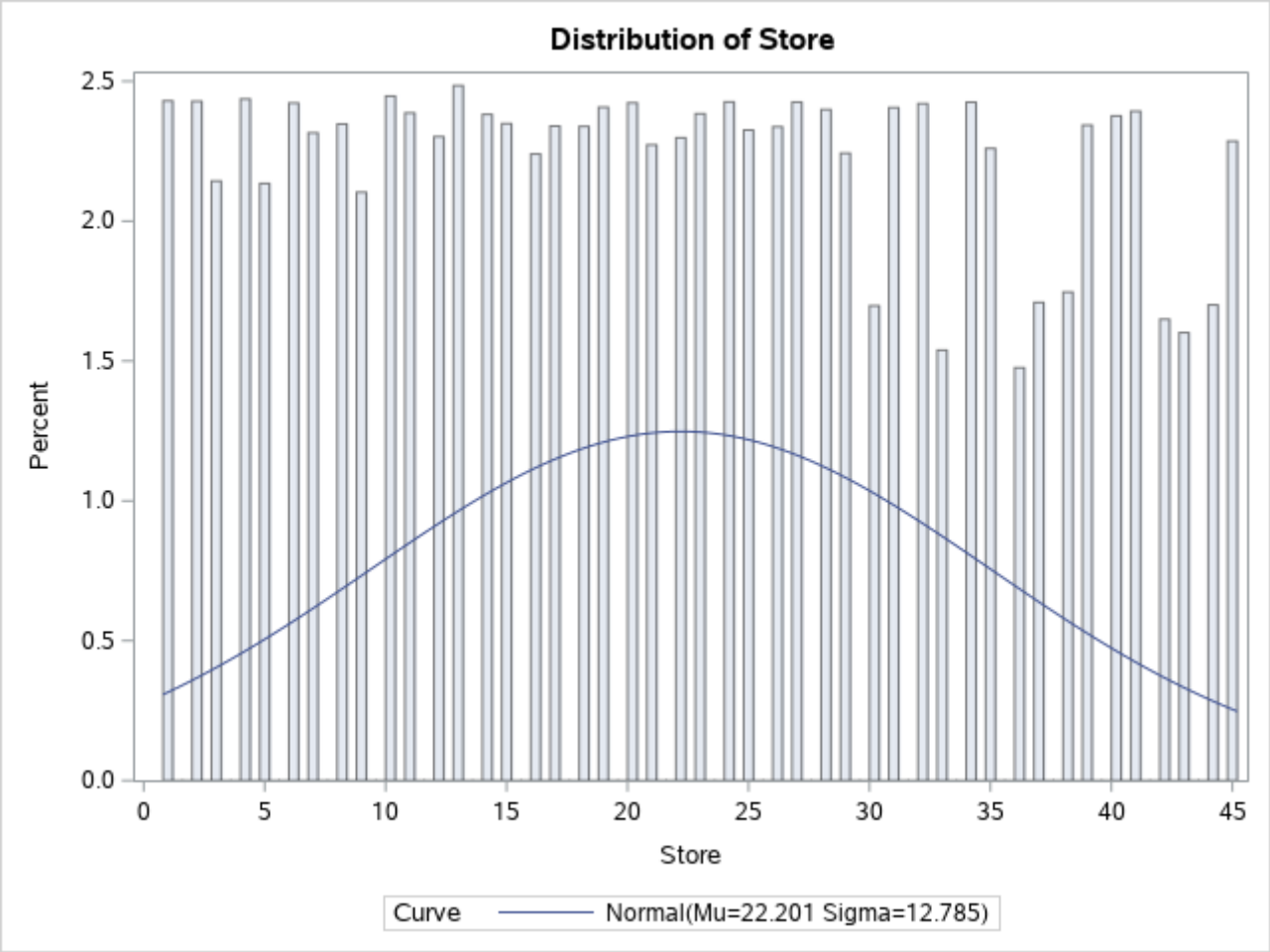
Quantiles (Definition 5)	
Level	Quantile
100% Max	45
99%	45
95%	43
90%	40
75% Q3	33
50% Median	22
25% Q1	11
10%	5
5%	3
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	10244	45	421566

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	10243	45	421567
1	10242	45	421568
1	10241	45	421569
1	10240	45	421570

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure



Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Fitted Normal Distribution for Store

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	22.20055
Std Dev	Sigma	12.7853

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.06794	Pr > D	<0.010
Cramer-von Mises	W-Sq	559.96899	Pr > W-Sq	<0.005

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Anderson-Darling	A-Sq	4199.64129	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	1.00000	-7.54250
5.0	3.00000	1.17060
10.0	5.00000	5.81553
25.0	11.00000	13.57699
50.0	22.00000	22.20055
75.0	33.00000	30.82410
90.0	40.00000	38.58556
95.0	43.00000	43.23049
99.0	45.00000	51.94359

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Variable: Dept

Moments			
N	421570	Sum Weights	421570
Mean	44.2603174	Sum Observations	18658822
Std Deviation	30.492054	Variance	929.765358
Skewness	0.35822319	Kurtosis	-1.2155706
Uncorrected SS	1217805636	Corrected SS	391960252
Coeff Variation	68.8925336	Std Error Mean	0.04696257

Basic Statistical Measures			
Location		Variability	
Mean	44.26032	Std Deviation	30.49205
Median	37.00000	Variance	929.76536
Mode	1.00000	Range	98.00000
		Interquartile Range	56.00000

Note: The mode displayed is the smallest of 22 modes with a count of 6435.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	942.4595	Pr > |t|	<.0001
Sign	M	210785	Pr >= |M|	<.0001
Signed Rank	S	4.443E10	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	99
99%	98
95%	95
90%	92

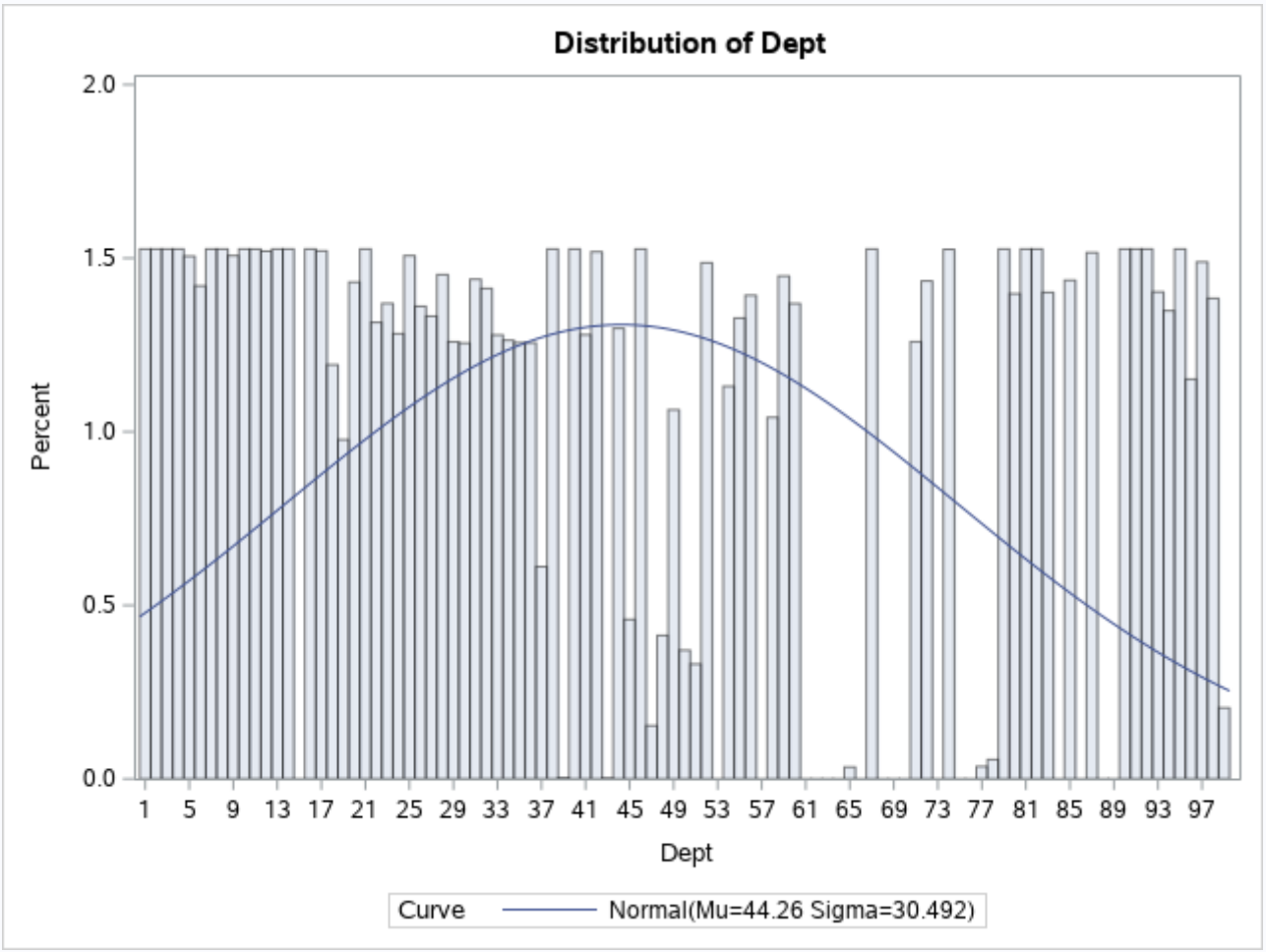
Quantiles (Definition 5)	
Level	Quantile
75% Q3	74
50% Median	37
25% Q1	18
10%	7
5%	4
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	412076	99	404762
1	412075	99	404763
1	412074	99	404764
1	412073	99	411932
1	412072	99	411933

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure

Decide whether we drop it now or later.



Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure

Fitted Normal Distribution for Dept

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	44.26032
Std Dev	Sigma	30.49205

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.1069	**Pr > D**	<0.010
Cramer-von Mises	**W-Sq**	1568.9910	**Pr > W-Sq**	<0.005
Anderson-Darling	**A-Sq**	10473.5350	**Pr > A-Sq**	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	**Observed**	**Estimated**
1.0	1.00000	-26.67481
5.0	4.00000	-5.89465
10.0	7.00000	5.18318
25.0	18.00000	23.69374
50.0	37.00000	44.26032
75.0	74.00000	64.82690
90.0	92.00000	83.33746
95.0	95.00000	94.41528
99.0	98.00000	115.19544

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Variable: Date

Moments			
N	421570	**Sum Weights**	421570
Mean	18796.3545	**Sum Observations**	7923979182
Std Deviation	288.967647	**Variance**	83502.301
Skewness	-0.0082245	**Kurtosis**	-1.1992392
Uncorrected SS	1.48977E14	**Corrected SS**	3.5202E10
Coeff Variation	1.53736006	**Std Error Mean**	0.44505571

Basic Statistical Measures			
Location		**Variability**	
Mean	18796.35	**Std Deviation**	288.96765
Median	18795.00	**Variance**	83502
Mode	18984.00	**Range**	994.00000
		Interquartile Range	504.00000

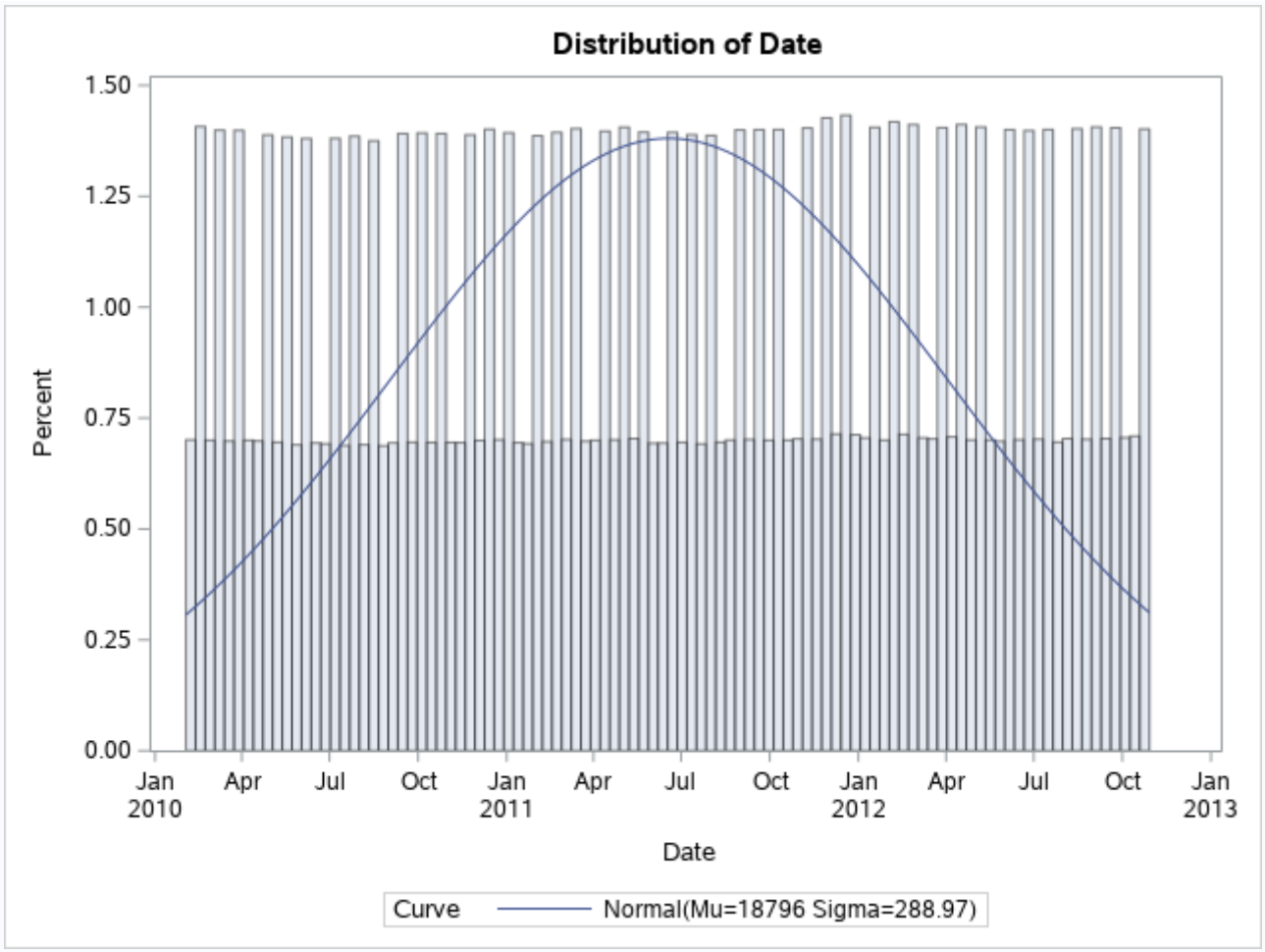
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	42233.71	**Pr > |t|**	<.0001
Sign	**M**	210785	**Pr >= |M|**	<.0001
Signed Rank	**S**	4.443E10	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	19292
99%	19285
95%	19243
90%	19194
75% Q3	19047
50% Median	18795
25% Q1	18543
10%	18396
5%	18347
1%	18305
0% Min	18298

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
18298	421436	19292	421013
18298	421293	19292	421147
18298	421148	19292	421290
18298	420871	19292	421435
18298	420728	19292	421570

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure



Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Fitted Normal Distribution for Date

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	18796.35
Std Dev	Sigma	288.9676

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.06079	**Pr > D**	<0.010
Cramer-von Mises	**W-Sq**	643.71170	**Pr > W-Sq**	<0.005
Anderson-Darling	**A-Sq**	4688.61073	**Pr > A-Sq**	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	**Observed**	**Estimated**
1.0	18305.0	18124.1
5.0	18347.0	18321.0
10.0	18396.0	18426.0
25.0	18543.0	18601.4
50.0	18795.0	18796.4
75.0	19047.0	18991.3
90.0	19194.0	19166.7

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
95.0	19243.0	19271.7
99.0	19285.0	19468.6

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Variable: Weekly_Sales

Derive Variable: TargetVariable = WeeklySales per Sq meter= Weekly Sales / Size

Moments			
N	421570	Sum Weights	421570
Mean	15981.2581	Sum Observations	6737218987
Std Deviation	22711.1835	Variance	515797857
Skewness	3.26200819	Kurtosis	21.4912899
Uncorrected SS	3.25114E14	Corrected SS	2.17444E14
Coeff Variation	142.111362	Std Error Mean	34.9788009

Basic Statistical Measures			
Location		Variability	
Mean	15981.26	Std Deviation	22711
Median	7612.03	Variance	515797857
Mode	10.00	Range	698088
		Interquartile Range	18126

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	456.8841	Pr > |t|	<.0001
Sign	M	209463.5	Pr >= |M|	<.0001
Signed Rank	S	4.44E10	Pr >= |S|	<.0001

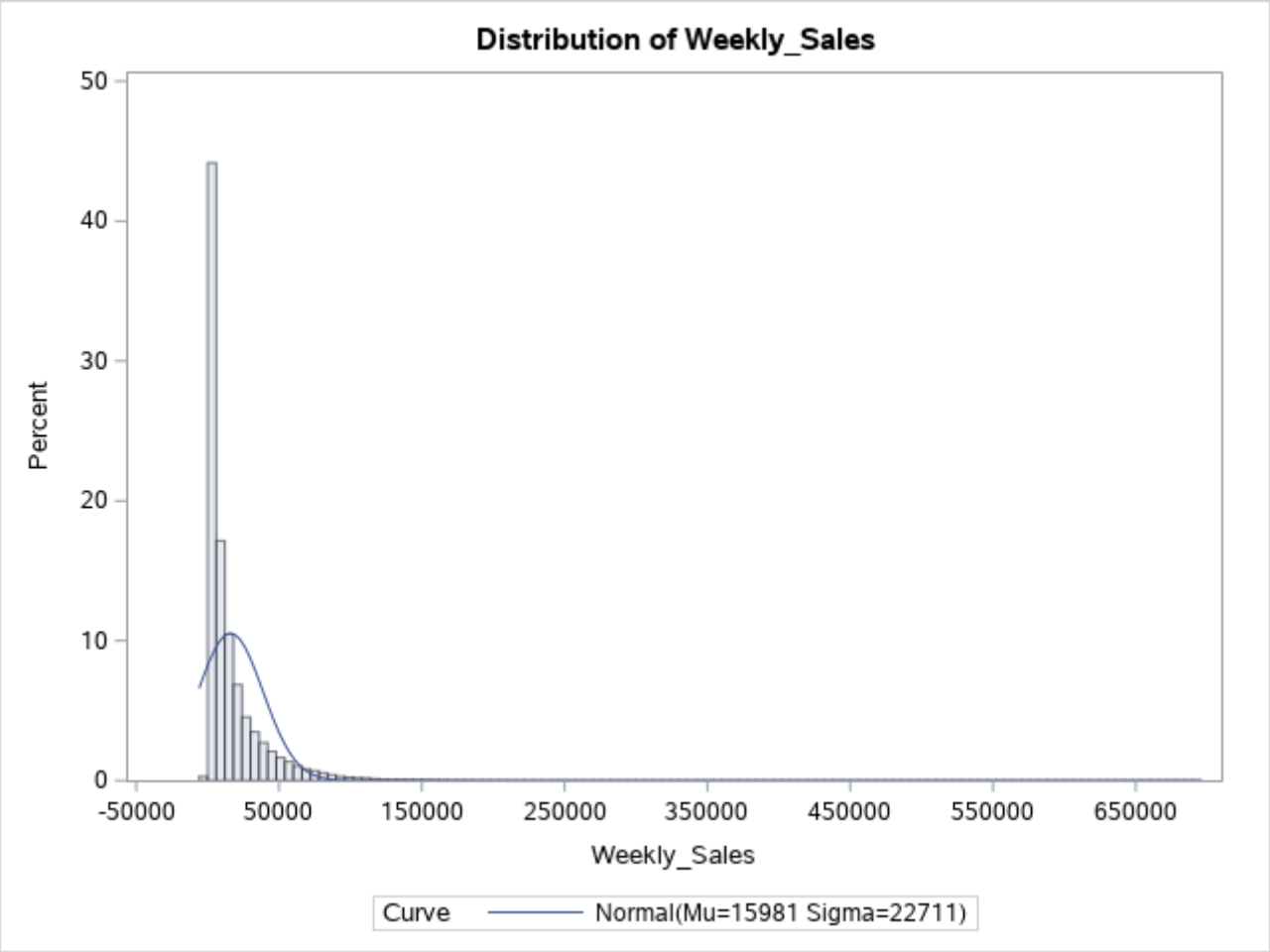
Quantiles (Definition 5)	
Level	Quantile
100% Max	693099.360
99%	106485.520
95%	61202.050
90%	42846.245
75% Q3	20205.860
50% Median	7612.030
25% Q1	2079.640
10%	291.085
5%	59.970
1%	5.000
0% Min	-4988.940

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-4988.94	267731	474330	135666

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-3924.00	336496	627963	337962
-1750.00	417802	630999	95426
-1699.00	153917	649770	338014
-1321.48	271301	693099	95374

Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure



Using PROC UNIVARIATE to Examine Store Dept Date Weekly_Sales IsHoliday

The UNIVARIATE Procedure
Fitted Normal Distribution for Weekly_Sales

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	15981.26
Std Dev	Sigma	22711.18

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.2395	Pr > D	<0.010
Cramer-von Mises	W-Sq	7085.3597	Pr > W-Sq	<0.005

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Anderson-Darling	A-Sq	38132.4564	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	5.0000	-36852.855
5.0	59.9700	-21375.314
10.0	291.0850	-13124.295
25.0	2079.6400	662.798
50.0	7612.0300	15981.258
75.0	20205.8600	31299.719
90.0	42846.2450	45086.811
95.0	61202.0500	53337.831
99.0	106485.5200	68815.372