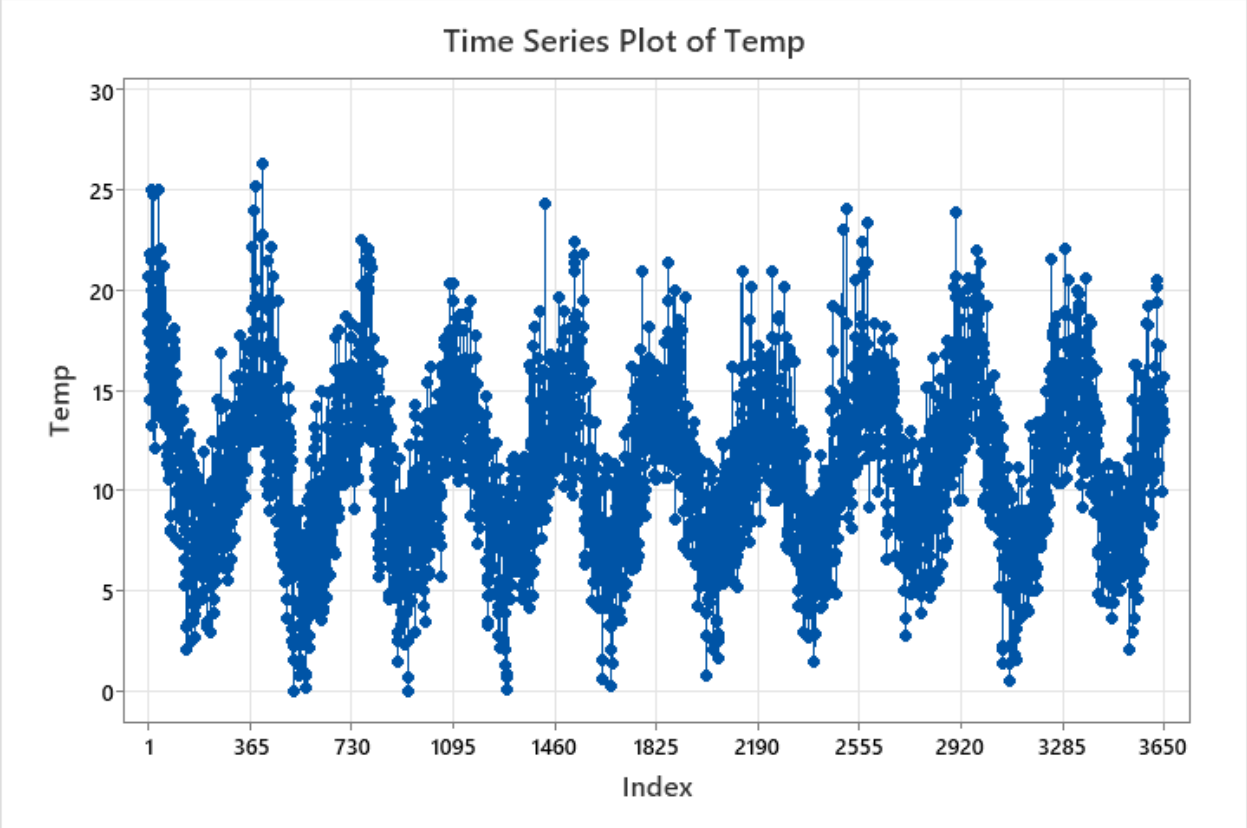
Name: Ahmed Hassan Hany

Data: Australia temperature

Link: https://gist.github.com/JacobToftgaardRasmussen/78cac8b522ce6d8cab4fc80f7de48ee9

First Step: Identification

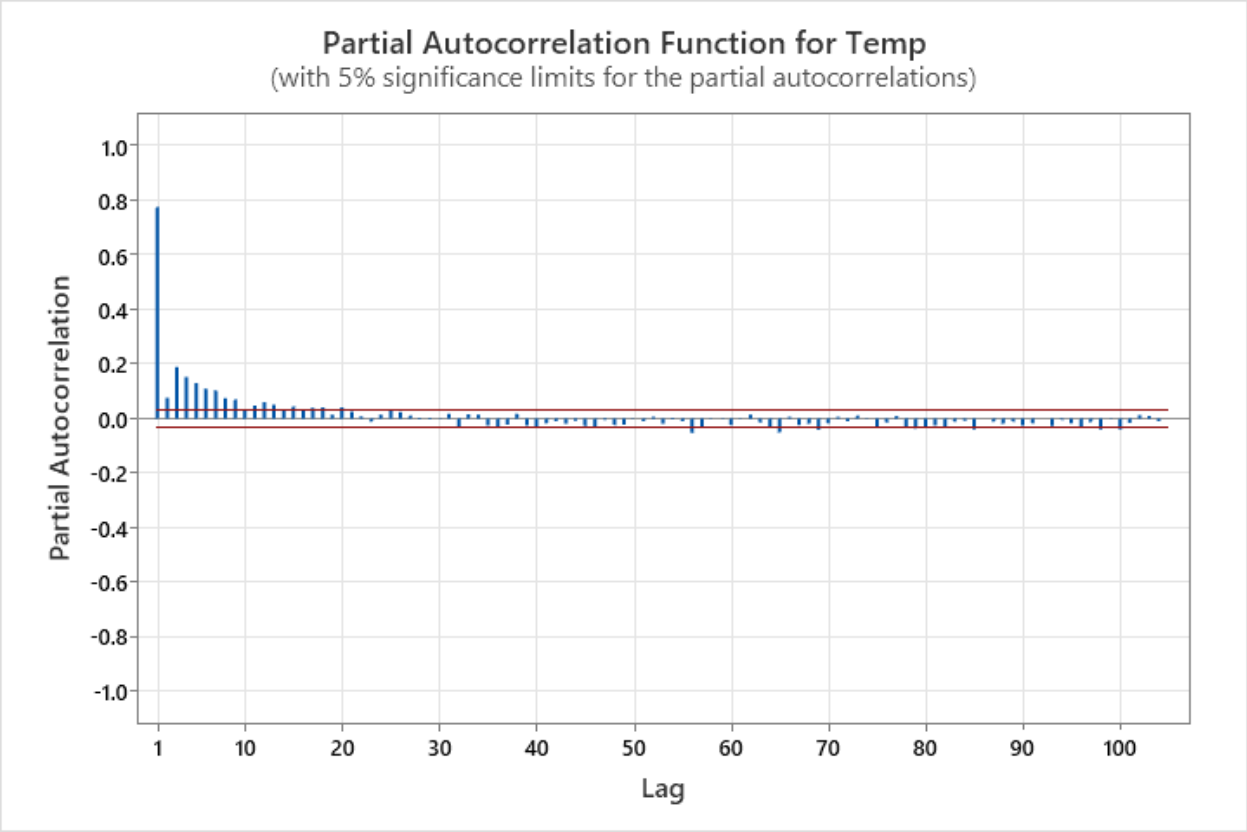
Time Series Plot of Temp

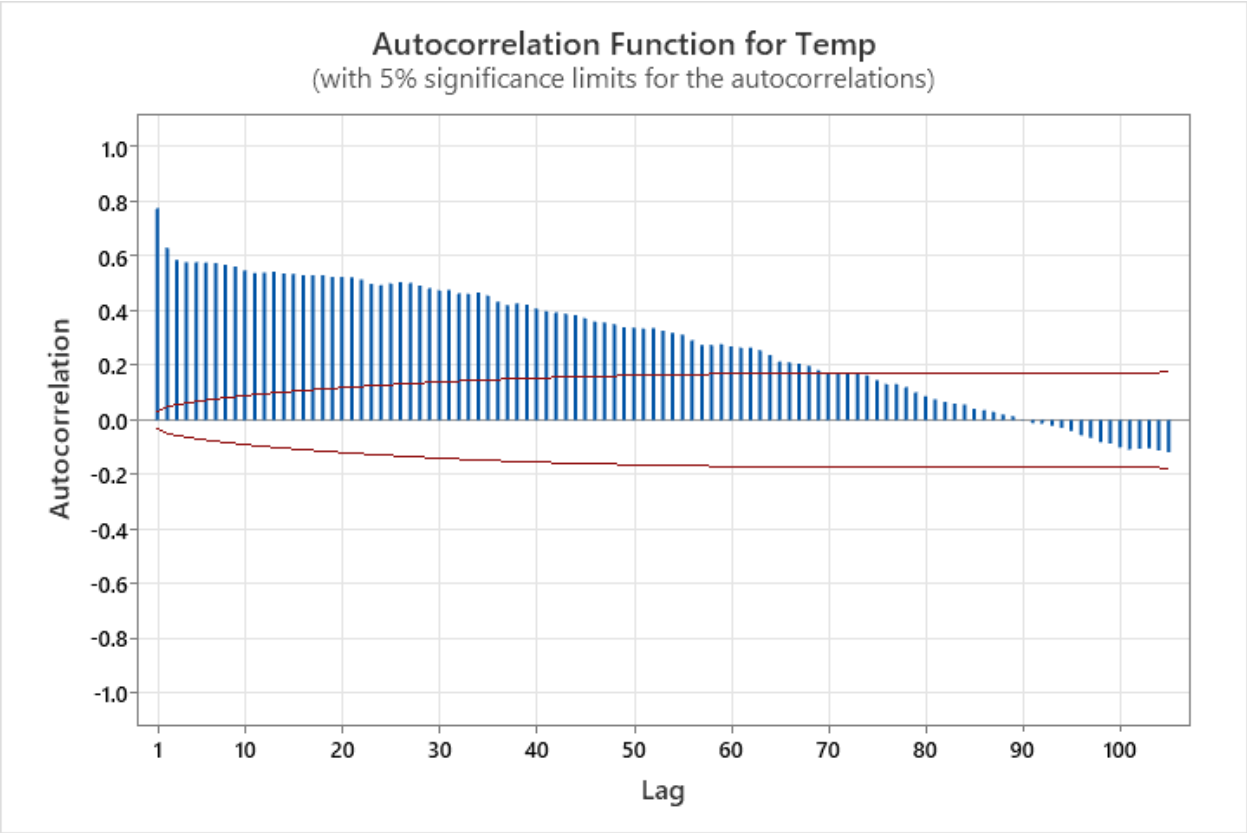


From the time series plot:

- Data seems stationary in mean and variance.
- Data seems to have outliers.
- No general trend.

Conclusion: We don't need to transform the data.


Partial Autocorrelation Function for Temp
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
10
20
30
40
50
60
70
80
90
100
Lag


Autocorrelation Function for Temp
(with 5% significance limits for the autocorrelations)
Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
10
20
30
40
50
60
70
80
90
100
Lag

Observations tail-off to zero quickly suggesting an ARMA model.

2) Suggesting a model

Method

Criterion for best model	Minimum AICc
Rows used	3650
Rows unused	0

Model Selection

Model (d = 0)	LogLikelihood	AICc	AIC	BIC
p = 2, q = 2*	-8384.68	16781.4	16781.4	16818.6
p = 1, q = 2	-8403.31	16816.6	16816.6	16847.6
p = 2, q = 1	-8432.04	16874.1	16874.1	16905.1
p = 1, q = 1	-8605.43	17218.9	17218.9	17243.7
p = 2, q = 0	-8618.98	17246.0	17246.0	17270.8
p = 1, q = 0	-8629.86	17265.7	17265.7	17284.3
p = 0, q = 2	-8969.89	17947.8	17947.8	17972.6
p = 0, q = 1	-9274.08	18554.2	18554.2	18572.8

** Best model with minimum AICc. Output for the best model follows.*

ARMA (2,0,2) has minimum AIC.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.2398	0.0398	31.17	0.000
AR 2	-0.2449	0.0390	-6.27	0.000
MA 1	0.6478	0.0391	16.57	0.000
MA 2	0.2303	0.0311	7.40	0.000
Constant	0.05702	0.00486	11.73	0.000
Mean	11.300	0.963		

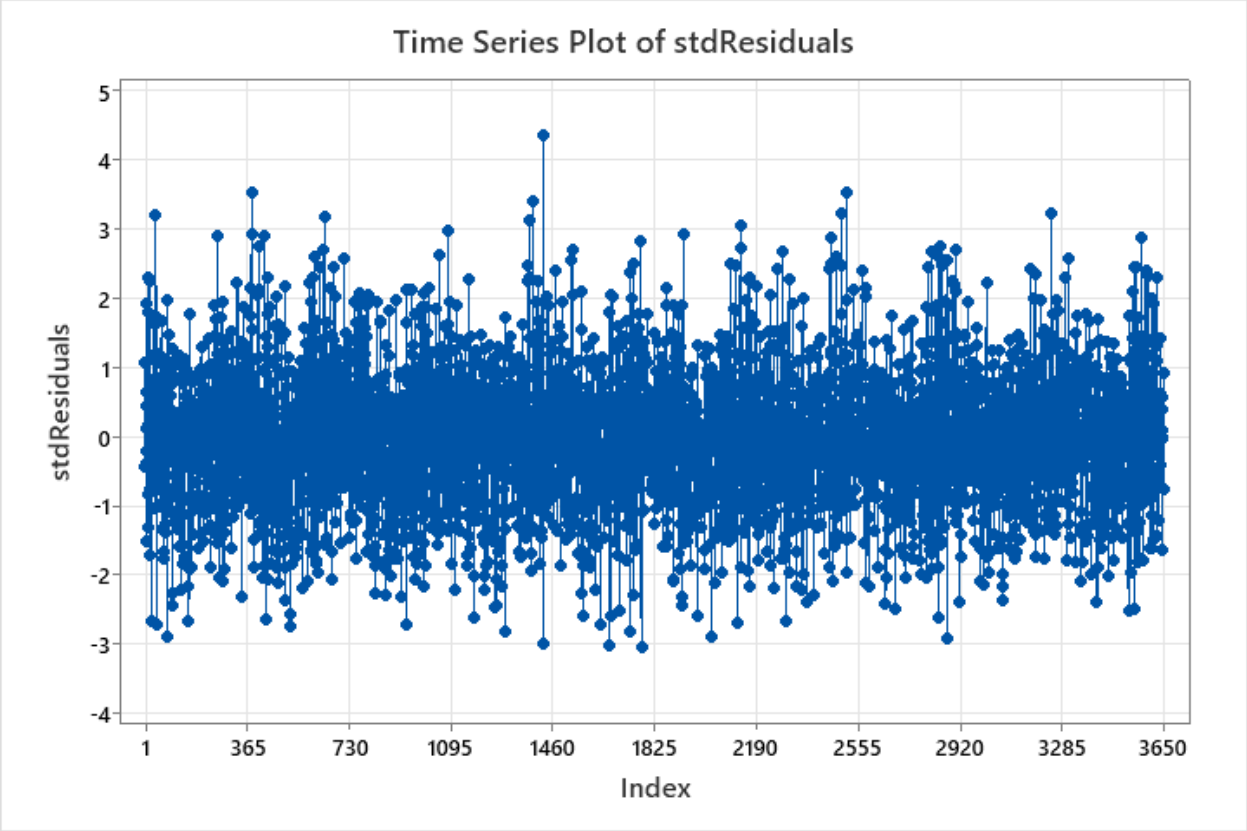
The model is statistically significant.

Third step: Verification of Diagnostics

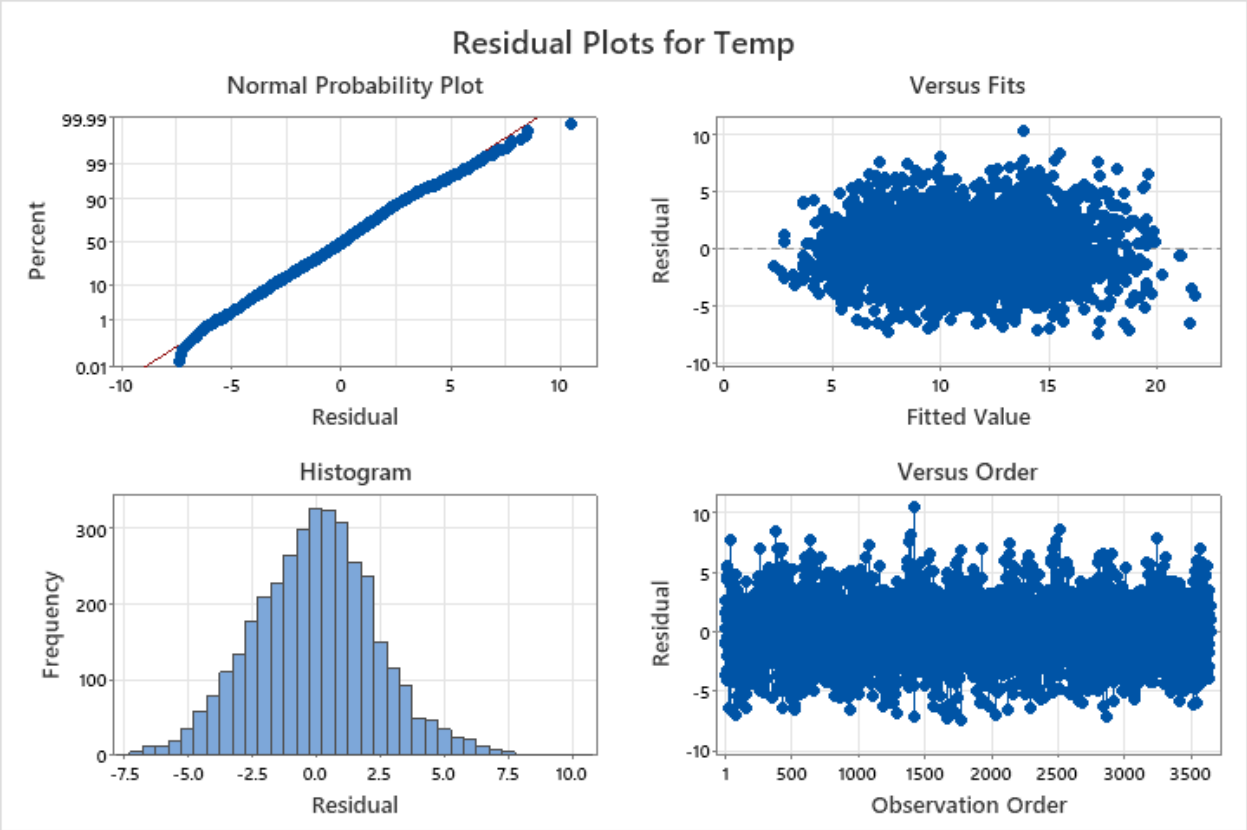
Stationarity and invertibility analysis:

The model is invertible but not stationary.

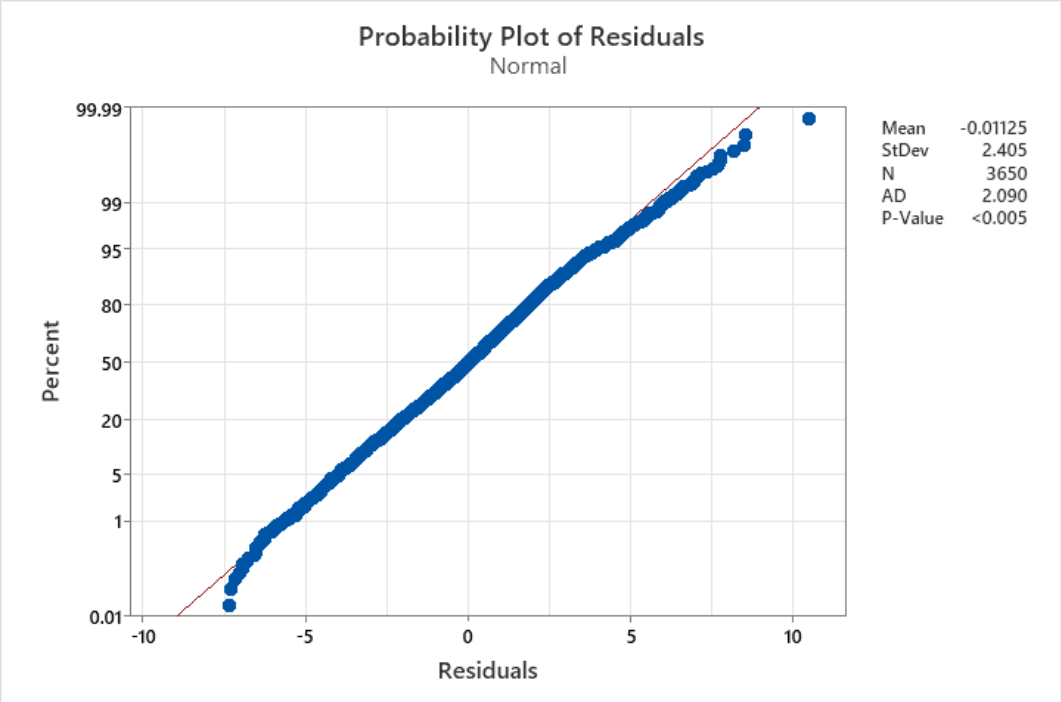
Residual analysis:



Plotting standardized residuals shows that we have outliers, and the data is stationary.



Normality test:



Null Hypothesis: residuals are normal.

Alternative Hypothesis: residuals are not normal.

p-value < 0.05, we reject null hypothesis. The model doesn't come from normal distribution.

Randomness test by Runs:

Test

Null hypothesis	H_0: The order of the data is random
Alternative hypothesis	H_1: The order of the data is not random

Number of Runs

Observed	Expected	P-Value
1848	1825.76	0.461

p-value > 0.05; the model is random.

Outlier test:

Method

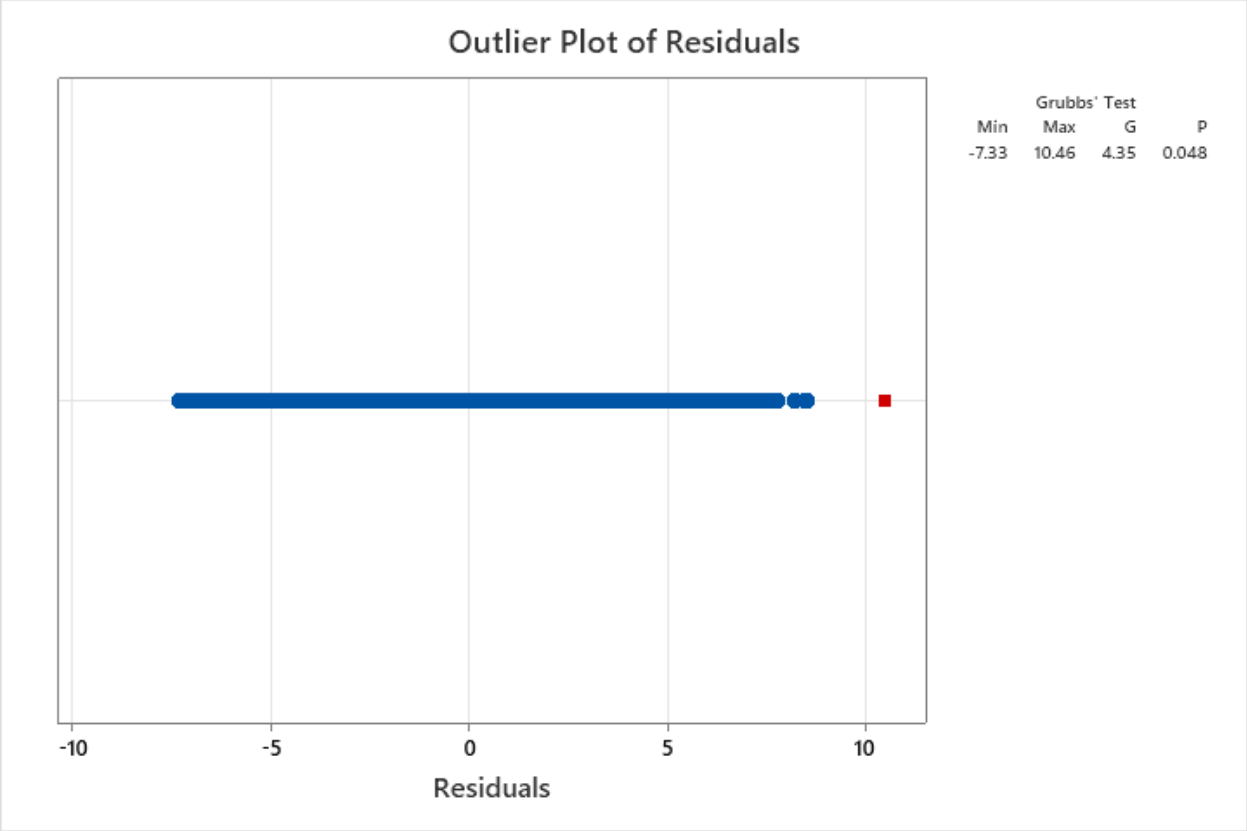
Null hypothesis	All data values come from the same normal population
Alternative hypothesis	Smallest or largest data value is an outlier
Significance level	$\alpha = 0.05$

Grubbs' Test

Variable	N	Mean	StDev	Min	Max	G	P
Residuals	3650	-0.0113	2.4051	-7.3316	10.4616	4.35	0.048

Outlier

Variable	Row	Outlier
Residuals	1426	10.4616



p-value = 0.048 which is less than alpha by 0.02.

We have only one outlier.

If we remove the outlier, the p-value will probably be bigger than alpha.

Auto-Correlation test:

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	7.32	15.96	43.86	53.94
DF	7	19	31	43
P-Value	0.396	0.660	0.063	0.122

Null hypothesis: auto-correlation = 0

Alternative hypothesis: auto-correlation != 0.

p-value > 0.05, we fail to reject the null hypothesis; there is no auto-correlation.

Forecast:

Forecasts from Time Period 3650

			95% Limits	
Time Period	Forecast	SE Forecast	Lower	Upper	Actual
3651	12.9934	2.40642	8.27585	17.7109	
3652	13.3986	2.79652	7.91633	18.8809	
3653	13.4872	2.86507	7.87059	19.1039	