Time Series Analysis

With Applications in Minitab

Foreword

We live in the midst of a data deluge. According to recent estimates, 2.5 quintillion (10^{18}) bytes of data are generated on a daily basis. This is so much data that over 90 percent of the information that we store nowadays was generated in the past decade alone. Unfortunately, most of this information cannot be used by humans. Either the data is beyond the means of standard analytical methods, or it is simply too vast for our limited minds to even comprehend.

Preface

Minitab is a statistics package developed at the Pennsylvania State University by researchers Barbara F. Ryan, Thomas A. Ryan, Jr., Brian L. Joiner in 1972.

The project received funding from the Triola Statistics Company. It began as a light version of OMNITAB 80, a statistical analysis program by NIST, which was conceived by Joseph Hilsenrath in years 1962-1964 as OMNITAB program for IBM 7090.

The documentation for OMNITAB 80 was last published 1986, and there has been no significant development since then.

Minitab is distributed by Minitab, LLC, a privately owned company headquartered in State College, Pennsylvania.

In 2020, during the COVID-19 pandemic, Minitab LLC requested and received between $5 million and $10 million under the Paycheck Protection Program to avoid having to let go 250 employees.

As of 2021, Minitab LLC had subsidiaries in the UK, France, Germany, Hong Kong, and Australia.

I hope you enjoy reading this

Table of Contents :

Description of data

The data I am going to analysis for you about the average rainfall from 1976-2016, which includes 12 days in June each year.

Number of Rows :	**492**
Number of Columns:	2
Missing Values?	NO

Zt are mean in my data.

Of course, you can download my data file via the link below

https://github.com/anjell1/anjell1.git

At the first

Descriptive Statistics

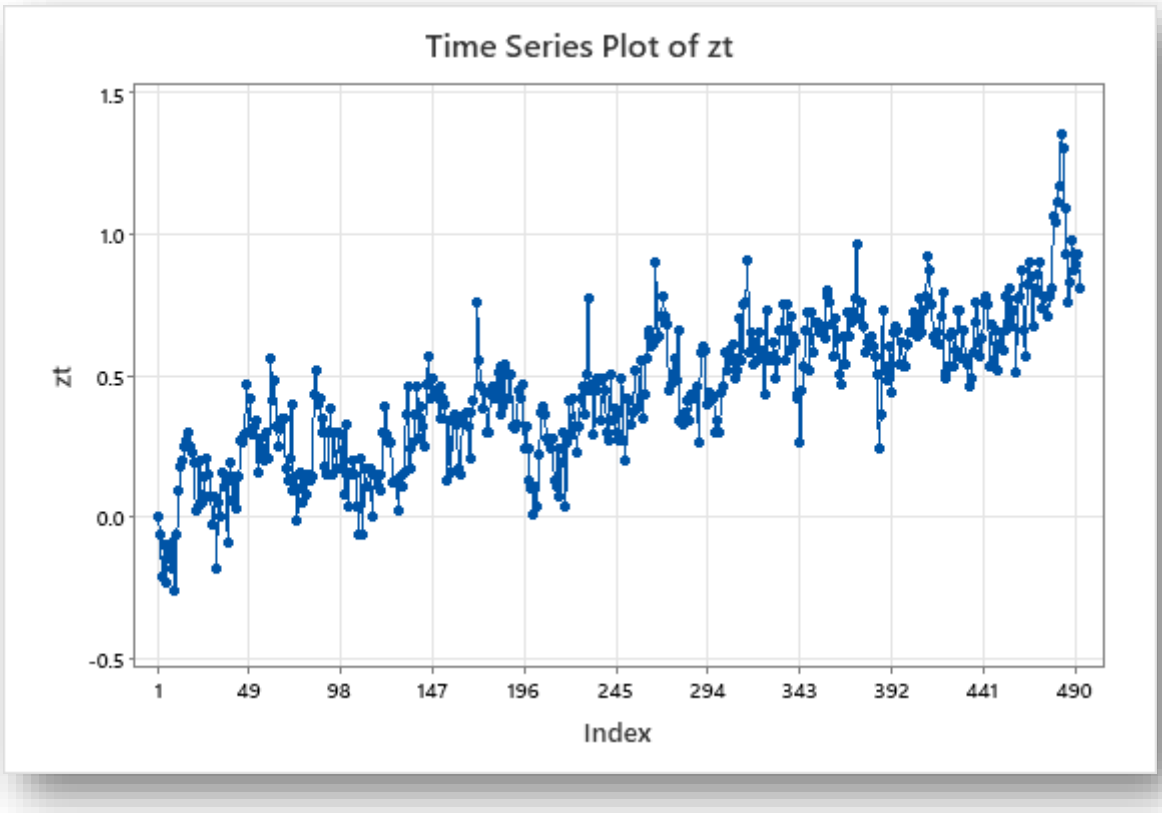
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
zt	492	0	0.4386	0.0116	0.2566	-0.2600	0.2600	0.4400	0.6300	1.3500

stDev is standard devition = 0.2566

N = Number of data

Next step :

we give time series plot of Zt , (Zt are average rainfall)



In this chart, with a large number of visible fluctuations, we can reach a decline

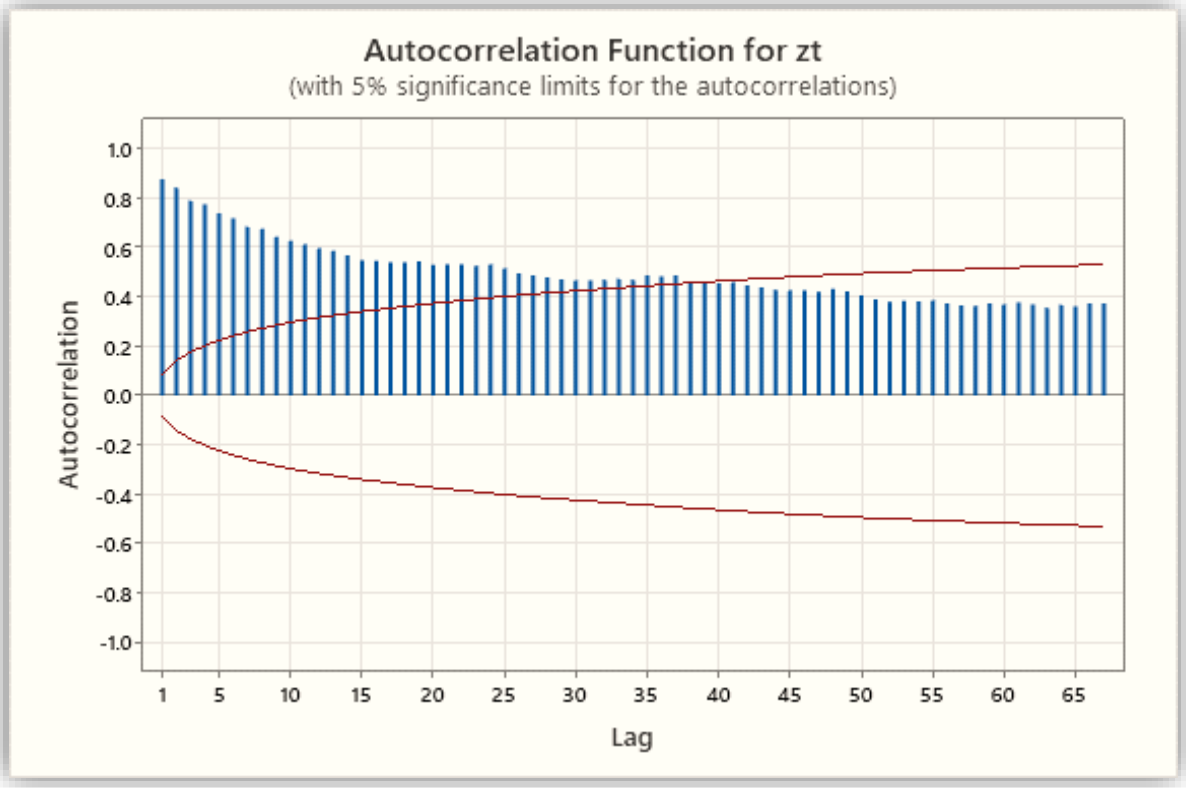
After that

I give a ACF(Autocorrelation Function: zt) with Lag 1

Autocorrelations

Lag	ACF	T	LBQ
1	0.876459	19.44	380.25
2	0.842551	11.73	732.37
3	0.790153	8.81	1042.69
4	0.775359	7.54	1342.11
5	0.737874	6.47	1613.83
6	0.717982	5.82	1871.63
7	0.682947	5.19	2105.37
8	0.675135	4.87	2334.26
9	0.642936	4.43	2542.26
10	0.626073	4.15	2739.91
11	0.612230	3.92	2929.31
12	0.597115	3.71	3109.85
13	0.586935	3.55	3284.65
14	0.567588	3.35	3448.45
15	0.548595	3.17	3601.80
16	0.545567	3.09	3753.78
17	0.539628	2.99	3902.78
18	0.540503	2.95	4052.58
19	0.543903	2.91	4204.59
20	0.530102	2.79	4349.29
21	0.531085	2.75	4494.84
22	0.530850	2.71	4640.56
23	0.524650	2.64	4783.21
24	0.530868	2.63	4929.57
25	0.515522	2.52	5067.88
26	0.495535	2.39	5195.95
27	0.487552	2.33	5320.20
28	0.479038	2.26	5440.40
29	0.470007	2.20	5556.37
30	0.467244	2.16	5671.22
31	0.465641	2.14	5785.53
32	0.468638	2.13	5901.57
33	0.471839	2.13	6019.46
34	0.469471	2.10	6136.42
35	0.487019	2.15	6262.56
36	0.483624	2.12	6387.23
37	0.486096	2.11	6513.45
38	0.462797	1.99	6628.11
39	0.454376	1.94	6738.88
40	0.453648	1.92	6849.54
41	0.457849	1.93	6962.51
42	0.445958	1.86	7069.92
43	0.438508	1.82	7174.01
44	0.428905	1.77	7273.81
45	0.425383	1.74	7372.20
46	0.425486	1.73	7470.86
47	0.419588	1.70	7567.02
48	0.432953	1.74	7669.63
49	0.421633	1.68	7767.16
50	0.406608	1.62	7858.07
51	0.389703	1.54	7941.77
52	0.379588	1.49	8021.36
53	0.384116	1.50	8103.05
54	0.381193	1.49	8183.68
55	0.385347	1.50	8266.27
56	0.373085	1.44	8343.86
57	0.364614	1.40	8418.14
58	0.363505	1.39	8492.14
59	0.373980	1.43	8570.64
60	0.369946	1.41	8647.64
61	0.377057	1.43	8727.82
62	0.368311	1.39	8804.49
63	0.354877	1.33	8875.84
64	0.368130	1.38	8952.80

65	0.361512	1.35	9027.19
66	0.374268	1.39	9107.11
67	0.372568	1.38	9186.49



The red lines are confidence interval

Descending is a good sign because it is in the confidence interval

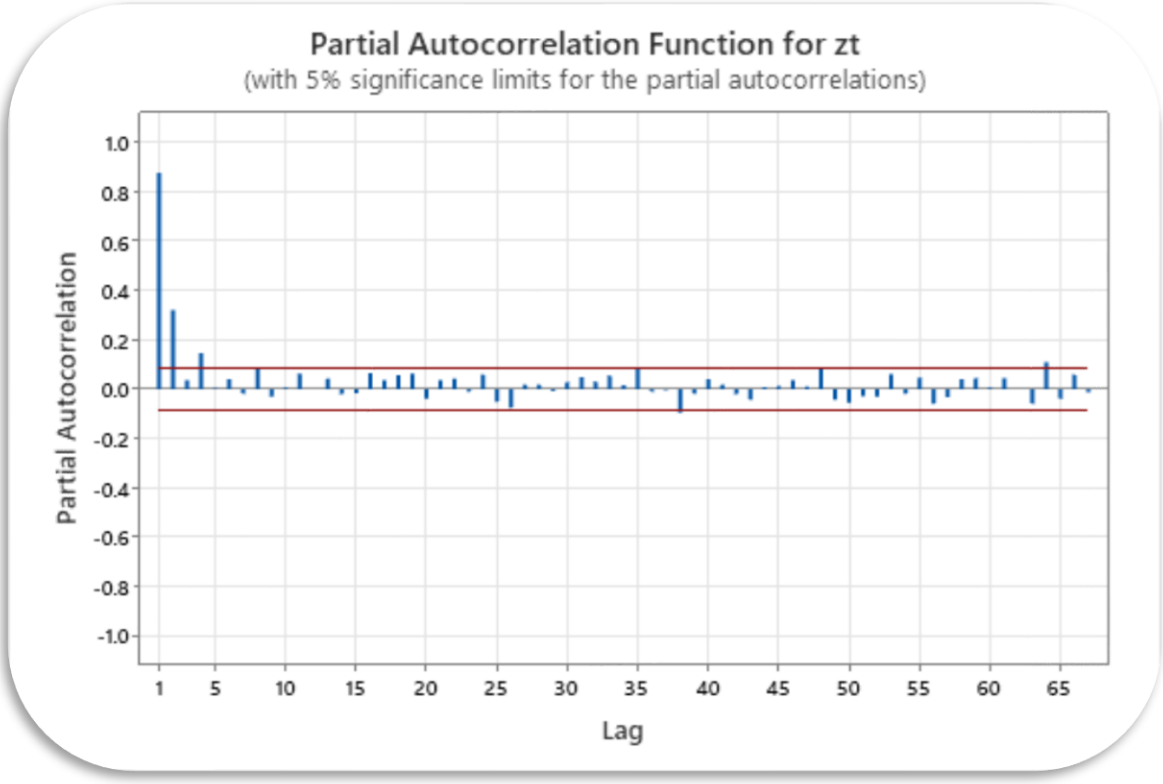
We can give PACF (Partial Autocorrelation Function): zt

Lag	PACF	T
1	0.876459	19.44
2	0.320816	7.12
3	0.035680	0.79
4	0.146189	3.24
5	0.004205	0.09
6	0.040817	0.91
7	-0.017327	-0.38
8	0.081055	1.80
9	-0.030395	-0.67
10	0.007071	0.16
11	0.062447	1.39
12	-0.000568	-0.01
13	0.041827	0.93
14	-0.020759	-0.46
15	-0.017018	-0.38
16	0.065373	1.45
17	0.036659	0.81
18	0.057424	1.27
19	0.062090	1.38
20	-0.038231	-0.85
21	0.036417	0.81
22	0.043026	0.95
23	-0.010272	-0.23
24	0.059623	1.32
25	-0.050454	-1.12
26	-0.076075	-1.69
27	0.017673	0.39
28	0.017746	0.39
29	-0.007942	-0.18
30	0.028913	0.64
31	0.047724	1.06
32	0.030277	0.67
33	0.054179	1.20
34	0.016785	0.37
35	0.087228	1.93
36	-0.009794	-0.22
37	-0.003450	-0.08
38	-0.096345	-2.14
39	-0.018711	-0.42
40	0.041176	0.91
41	0.018902	0.42
42	-0.020912	-0.46
43	-0.042017	-0.93
44	0.007799	0.17
45	0.013101	0.29
46	0.037079	0.82
47	0.010773	0.24
48	0.088992	1.97

49	-0.043151	-0.96
50	-0.054695	-1.21
51	-0.027991	-0.62
52	-0.031392	-0.70
53	0.061640	1.37
54	-0.017481	-0.39
55	0.047315	1.05
56	-0.058970	-1.31
57	-0.033433	-0.74
58	0.040802	0.91
59	0.044155	0.98
60	0.005935	0.13
61	0.044243	0.98
62	0.000638	0.01
63	-0.059847	-1.33
64	0.109570	2.43
65	-0.038948	-0.86
66	0.058156	1.29
67	-0.011346	-0.25

PACF with Lag 1

And see PAF chart I think is good

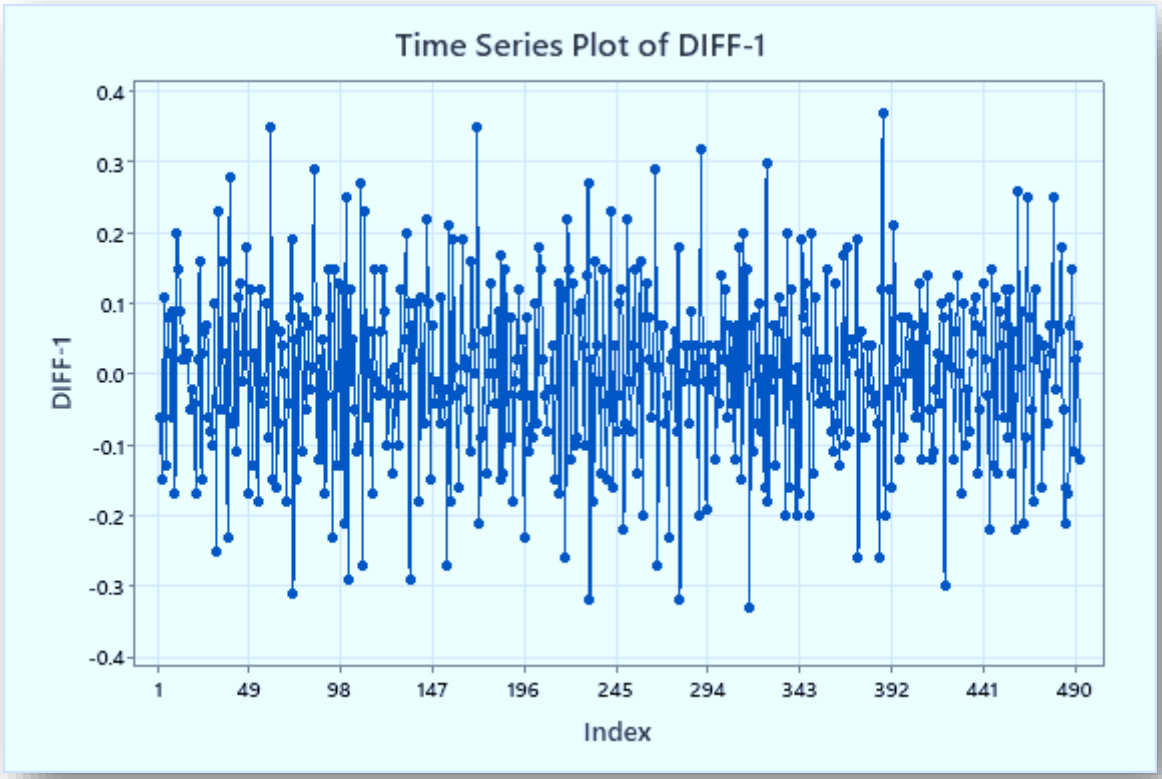


We can see outlines of interval confidence that is not bad but are low

But We can accept this .

In this part I'm create a difference column with LAG 1 .

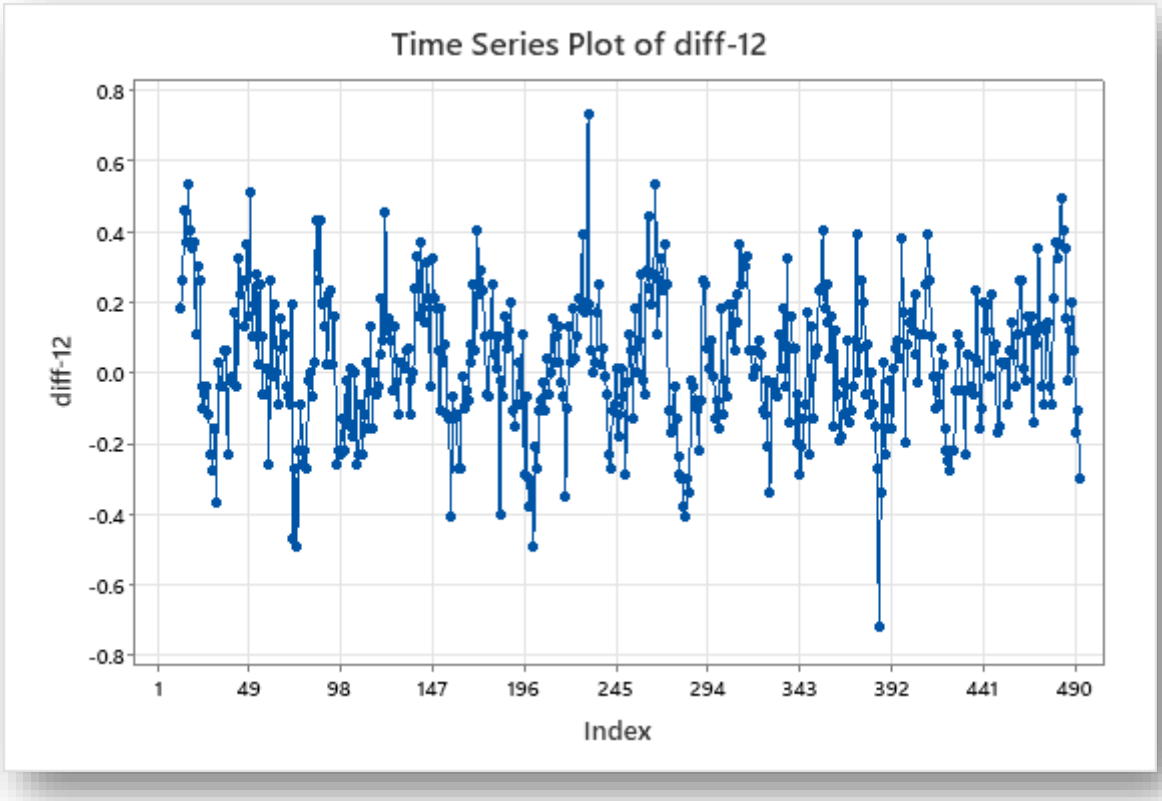
And then I give a time series plot.



We can see the fluctuations around the number zero

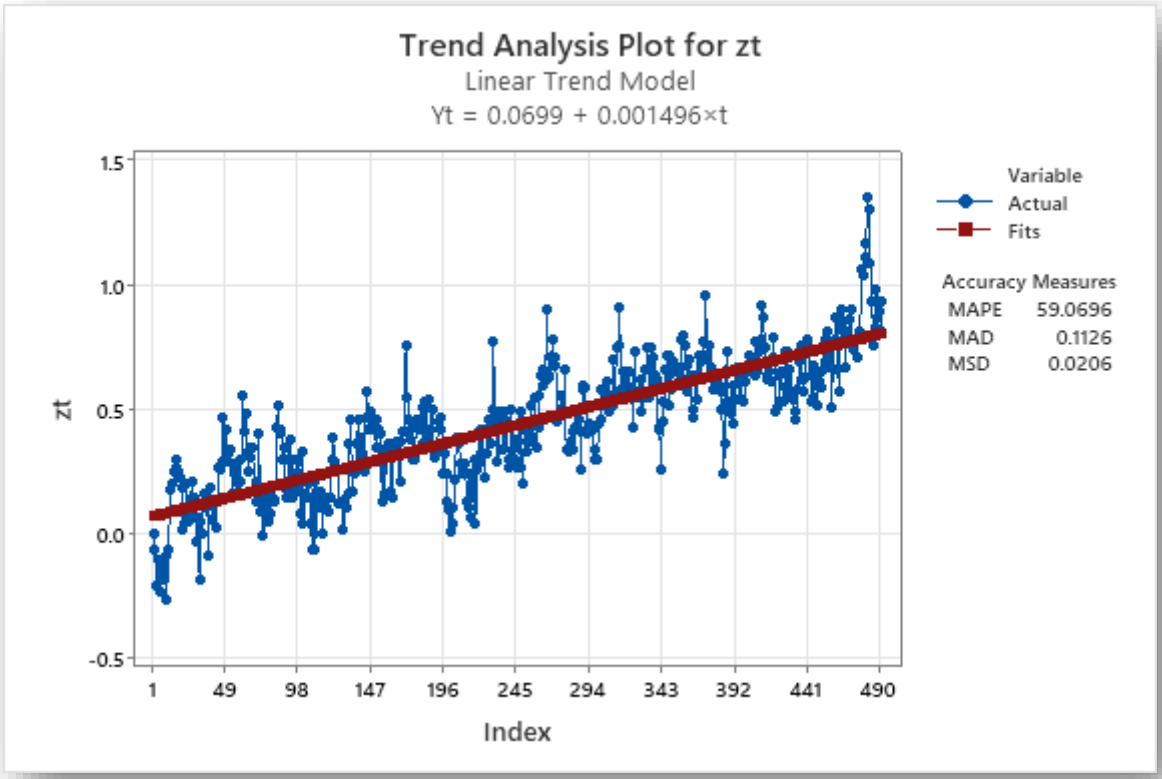
Attention oscillation and scattering are high

Time Series Plot of diff-12



We can not accept this because we're going to werse model

Trend Analysis for zt



It is relatively good, but we have to do other tests

Regression Analysis: zt versus t, t^2

Regression Equation

zt = 0.0840 + 0.001324 t + 0.000000 t^2

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0840	0.0195	4.31	0.000	
t	0.001324	0.000183	7.24	0.000	16.06
t^2	0.000000	0.000000	0.97	0.332	16.06

Beta 0 (t) is 0 and under 0.05 so we can't accept 0 and this model is wonderful good.

About beta 0(t^2) and higher 0.05 so we can accept 0 and this model is a bad model.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.143710	68.76%	68.63%	68.31%

R-sq = is so good because every more value is a good R-sq.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	22.2269	11.1135	538.11	0.000
t	1	1.0832	1.0832	52.45	0.000
t^2	1	0.0195	0.0195	0.94	0.332
Error	489	10.0991	0.0207		
Total	491	32.3261			

Fits and Diagnostics for Unusual Observations

Obs	zt	Fit	Resid	Std Resid	
3	-0.2100	0.0880	-0.2980	-2.09	R
5	-0.2300	0.0907	-0.3207	-2.25	R
10	-0.2600	0.0973	-0.3573	-2.51	R
32	-0.1800	0.1268	-0.3068	-2.15	R
48	0.4700	0.1484	0.3216	2.25	R
61	0.5600	0.1661	0.3939	2.75	R
63	0.4800	0.1688	0.3112	2.17	R
85	0.5200	0.1991	0.3209	2.24	R
108	-0.0600	0.2311	-0.2911	-2.03	R
110	-0.0600	0.2339	-0.2939	-2.05	R
171	0.7600	0.3206	0.4394	3.06	R
201	0.0100	0.3642	-0.3542	-2.47	R
203	0.0400	0.3672	-0.3272	-2.28	R
215	0.0700	0.3848	-0.3148	-2.20	R
218	0.0400	0.3892	-0.3492	-2.44	R
230	0.7700	0.4070	0.3630	2.53	R
266	0.9000	0.4609	0.4391	3.06	R
270	0.7800	0.4669	0.3131	2.18	R
315	0.9100	0.5357	0.3743	2.61	R
343	0.2600	0.5792	-0.3192	-2.23	R
373	0.9600	0.6264	0.3336	2.33	R
385	0.2400	0.6455	-0.4055	-2.83	R
386	0.3600	0.6470	-0.2870	-2.00	R
480	1.1100	0.7999	0.3101	2.17	R
481	1.1700	0.8015	0.3685	2.58	R
482	1.3500	0.8032	0.5468	3.83	R
483	1.3000	0.8049	0.4951	3.47	R

R Large residual

This data, we've are seasonality

We've 12 season after fix these season in data after that :

Regression Analysis: zt versus s1, s2, s3, s4, s5, s6, s7, s8, s9, s10, s11, s12

Regression Equation

zt = 0.4605 s1 + 0.4698 s2 + 0.4934 s3 + 0.4505 s4 + 0.4276 s5 + 0.4044 s6 + 0.4095 s7 + 0.4132 s8 + 0.4271 s9 + 0.4312 s10 + 0.4412 s11 + 0.4351 s12

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
s1	0.4605	0.0403	11.42	0.000	1.00
s2	0.4698	0.0403	11.65	0.000	1.00
s3	0.4934	0.0403	12.23	0.000	1.00
s4	0.4505	0.0403	11.17	0.000	1.00
s5	0.4276	0.0403	10.60	0.000	1.00
s6	0.4044	0.0403	10.03	0.000	1.00
s7	0.4095	0.0403	10.15	0.000	1.00
s8	0.4132	0.0403	10.24	0.000	1.00
s9	0.4271	0.0403	10.59	0.000	1.00
s10	0.4312	0.0403	10.69	0.000	1.00
s11	0.4412	0.0403	10.94	0.000	1.00
s12	0.4351	0.0403	10.79	0.000	1.00

We've good VIF and P-value

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.258259	74.79%	74.16%	73.51%

R-sq is better than past

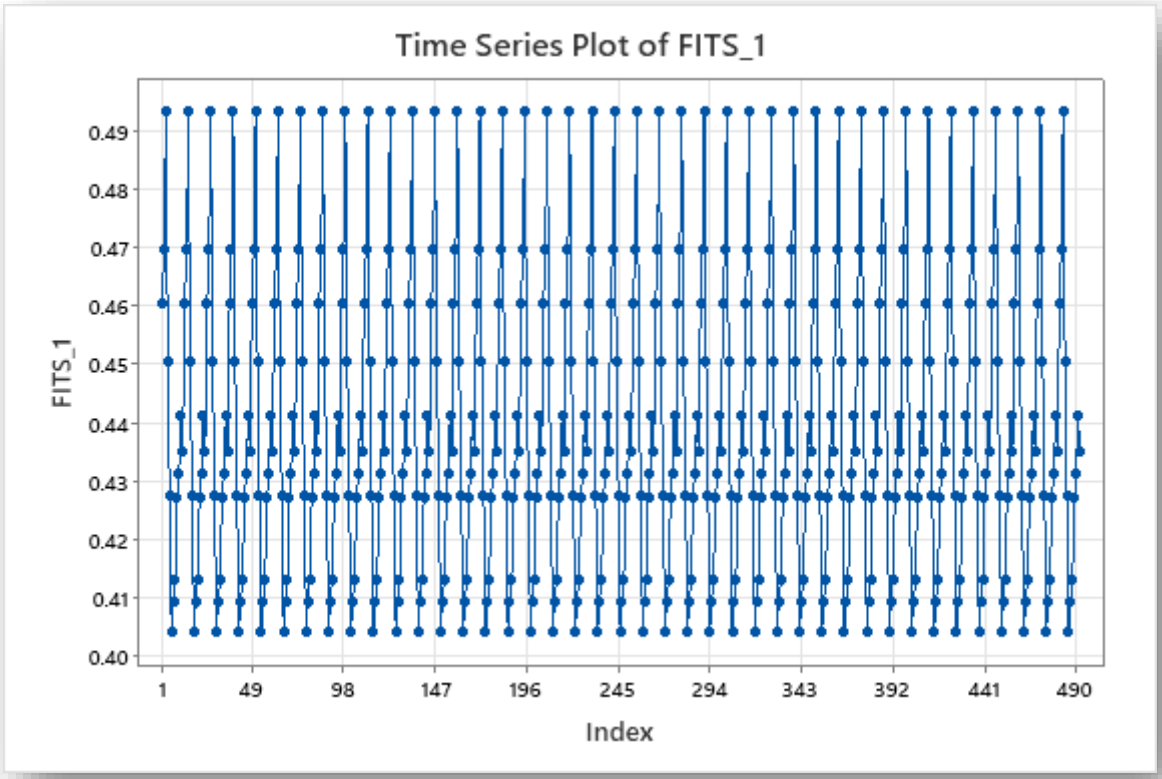
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	12	94.965	7.91373	118.65	0.000
s1	1	8.694	8.69401	130.35	0.000
s2	1	9.048	9.04750	135.65	0.000
s3	1	9.982	9.98178	149.66	0.000
s4	1	8.321	8.32051	124.75	0.000
s5	1	7.495	7.49514	112.37	0.000
s6	1	6.705	6.70479	100.52	0.000
s7	1	6.876	6.87571	103.09	0.000
s8	1	6.999	6.99911	104.94	0.000
s9	1	7.478	7.47805	112.12	0.000
s10	1	7.624	7.62396	114.31	0.000
s11	1	7.982	7.98166	119.67	0.000
s12	1	7.763	7.76258	116.38	0.000
Error	480	32.015	0.06670		
Total	492	126.980			

Fits and Diagnostics for Unusual Observations

Obs	zt	Fit	Resid	Std Resid	
2	-0.0600	0.4698	-0.5298	-2.08	R
3	-0.2100	0.4934	-0.7034	-2.76	R
4	-0.1000	0.4505	-0.5505	-2.16	R
5	-0.2300	0.4276	-0.6576	-2.58	R
6	-0.1500	0.4044	-0.5544	-2.17	R
7	-0.1200	0.4095	-0.5295	-2.08	R
8	-0.1800	0.4132	-0.5932	-2.33	R
9	-0.0900	0.4271	-0.5171	-2.03	R
10	-0.2600	0.4312	-0.6912	-2.71	R
32	-0.1800	0.4132	-0.5932	-2.33	R
38	-0.0900	0.4698	-0.5598	-2.19	R
110	-0.0600	0.4698	-0.5298	-2.08	R
478	1.0600	0.4312	0.6288	2.46	R
479	1.0400	0.4412	0.5988	2.35	R
480	1.1100	0.4351	0.6749	2.65	R
481	1.1700	0.4605	0.7095	2.78	R
482	1.3500	0.4698	0.8802	3.45	R
483	1.3000	0.4934	0.8066	3.16	R
484	1.0900	0.4505	0.6395	2.51	R
488	0.9800	0.4132	0.5668	2.22	R

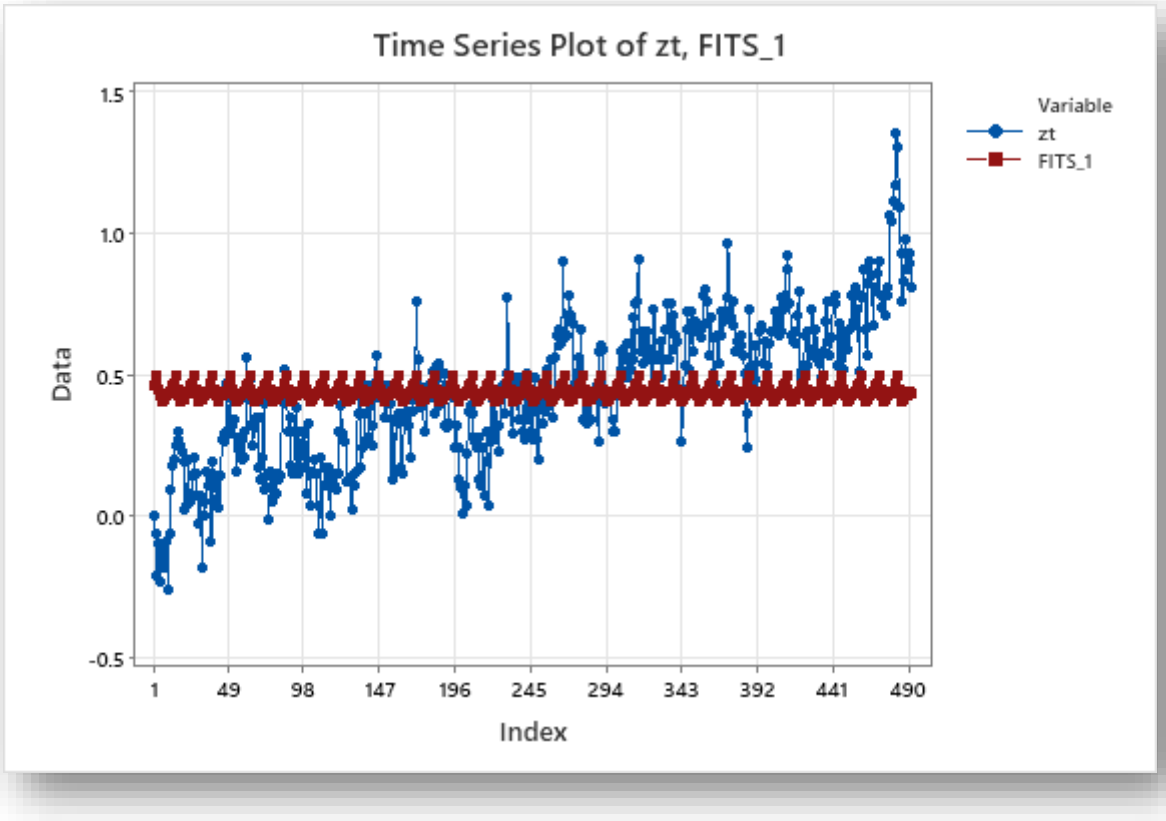
R Large residual



*After give regression analysis I selesct FIT and then I give a timr series plot of column FIT .

Dispersion and fit are good. These changes can also be seen in the important numbers on the previous page.

Time Series Plot of zt, FITS_1



we must fix it (red line data).

Regression Analysis: zt versus t, s1, s2, s3, s4, s5, s6, s7, s8, s9, s10, s11, s12

Regression Equation

zt = 0.001499 t + 0.0992 s1 + 0.1070 s2 + 0.1291 s3 + 0.0847 s4 + 0.0603 s5 + 0.0356 s6
+ 0.0392 s7 + 0.0414 s8 + 0.0538 s9 + 0.0564 s10 + 0.0650 s11 + 0.0574 s12

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
t	0.001499	0.000045	33.13	0.000	4.01
s1	0.0992	0.0248	4.00	0.000	1.24
s2	0.1070	0.0248	4.31	0.000	1.24
s3	0.1291	0.0248	5.20	0.000	1.24
s4	0.0847	0.0248	3.41	0.001	1.25
s5	0.0603	0.0249	2.42	0.016	1.25
s6	0.0356	0.0249	1.43	0.153	1.25
s7	0.0392	0.0249	1.58	0.116	1.25
s8	0.0414	0.0249	1.66	0.097	1.25
s9	0.0538	0.0249	2.16	0.031	1.26
s10	0.0564	0.0250	2.26	0.024	1.26
s11	0.0650	0.0250	2.60	0.010	1.26
s12	0.0574	0.0250	2.29	0.022	1.26

We've strong P-value in this table and of course we've best VIF

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.142489	92.34%	92.13%	91.91%

Oh WOW R-sq is so good because is change to higher value

Analysis of Variance

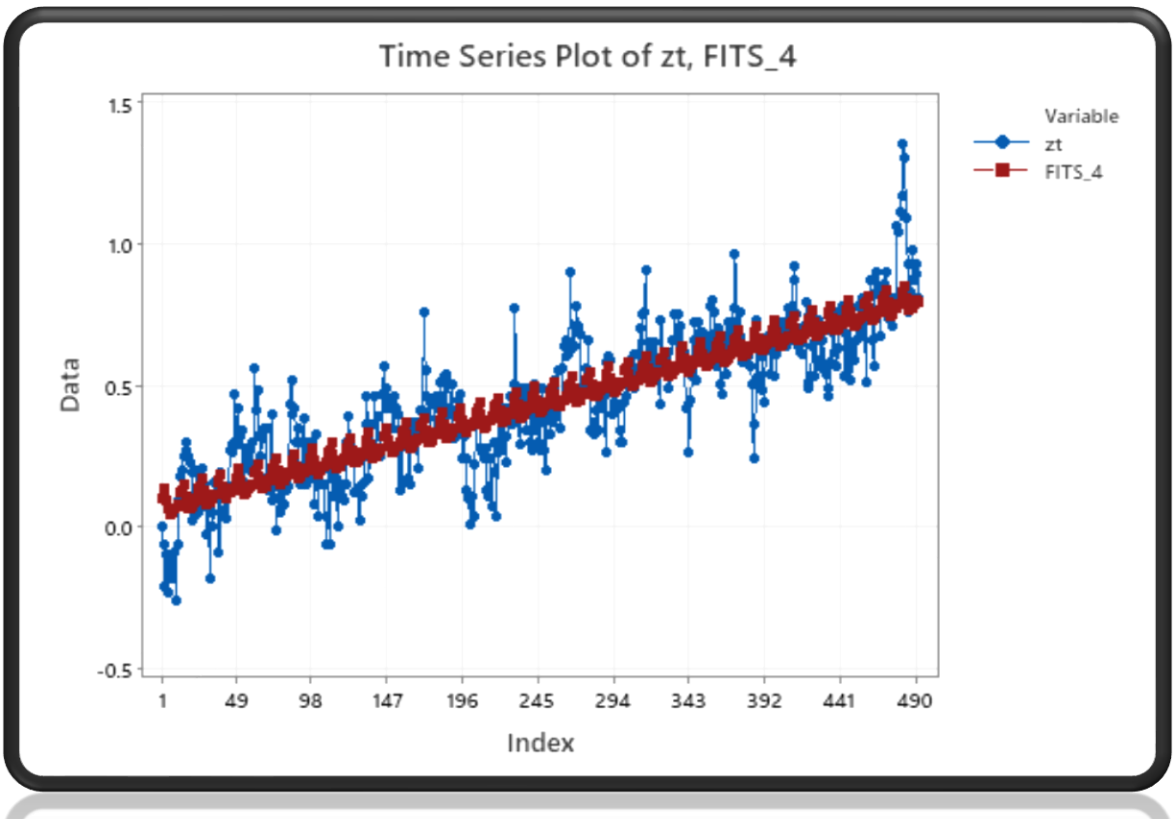
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	13	117.255	9.0196	444.24	0.000
t	1	22.290	22.2898	1097.85	0.000
s1	1	0.325	0.3254	16.03	0.000
s2	1	0.378	0.3778	18.61	0.000
s3	1	0.550	0.5496	27.07	0.000
s4	1	0.236	0.2361	11.63	0.001
s5	1	0.119	0.1194	5.88	0.016
s6	1	0.042	0.0416	2.05	0.153
s7	1	0.050	0.0504	2.48	0.116
s8	1	0.056	0.0560	2.76	0.097
s9	1	0.094	0.0945	4.65	0.031
s10	1	0.104	0.1038	5.11	0.024
s11	1	0.137	0.1372	6.76	0.010
s12	1	0.107	0.1068	5.26	0.022
Error	479	9.725	0.0203		
Total	492	126.980			

Fits and Diagnostics for Unusual Observations

Obs	zt	Fit	Resid	Std Resid	
3	-0.2100	0.1336	-0.3436	-2.45	R
5	-0.2300	0.0678	-0.2978	-2.12	R
10	-0.2600	0.0714	-0.3314	-2.36	R
48	0.4700	0.1293	0.3407	2.43	R
61	0.5600	0.1907	0.3693	2.63	R
85	0.5200	0.2266	0.2934	2.09	R
110	-0.0600	0.2719	-0.3319	-2.36	R
171	0.7600	0.3855	0.3745	2.66	R
201	0.0100	0.3551	-0.3451	-2.45	R
203	0.0400	0.3693	-0.3293	-2.34	R
215	0.0700	0.3873	-0.3173	-2.25	R
218	0.0400	0.4338	-0.3938	-2.80	R
230	0.7700	0.4518	0.3182	2.26	R
266	0.9000	0.5057	0.3943	2.80	R
270	0.7800	0.4404	0.3396	2.41	R
315	0.9100	0.6013	0.3087	2.19	R
343	0.2600	0.5534	-0.2934	-2.09	R
373	0.9600	0.6584	0.3016	2.15	R
385	0.2400	0.6764	-0.4364	-3.10	R
386	0.3600	0.6856	-0.3256	-2.32	R
433	0.4600	0.7483	-0.2883	-2.05	R
458	0.5100	0.7936	-0.2836	-2.02	R
478	1.0600	0.7730	0.2870	2.04	R
480	1.1100	0.7769	0.3331	2.37	R
481	1.1700	0.8203	0.3497	2.49	R
482	1.3500	0.8295	0.5205	3.71	R
483	1.3000	0.8532	0.4468	3.18	R

R Large residual

Time Series Plot of zt, FITS_4



Oh look at there we fix it this model is fix an it's a good situation .

AT THE END WE MUST GIVE THIS REGRESSION ANALYSIS

Regression Analysis: zt versus t, SIN, COS, SIN*2, Cos*2

Regression Equation

zt	=	0.0691 + 0.001499 t + 0.03071 SIN + 0.01606 COS + 0.00295 SIN*2 - 0.01715 Cos*2

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0691	0.0128	5.41	0.000	
t	0.001499	0.000045	33.33	0.000	1.00
SIN	0.03071	0.00903	3.40	0.001	1.00
COS	0.01606	0.00903	1.78	0.076	1.00
SIN*2	0.00295	0.00903	0.33	0.744	1.00
Cos*2	-0.01715	0.00903	-1.90	0.058	1.00

Oops ! we've not good P-vale and we can accept 0 but they are not good model with these

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.141630	69.84%	69.53%	69.07%

R-sq : decreased

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	22.5773	4.5155	225.11	0.000
t	1	22.2860	22.2860	1111.02	0.000
SIN	1	0.2319	0.2319	11.56	0.001
COS	1	0.0635	0.0635	3.16	0.076
SIN*2	1	0.0021	0.0021	0.11	0.744
Cos*2	1	0.0724	0.0724	3.61	0.058
Error	486	9.7487	0.0201		
Total	491	32.3261			

Fits and Diagnostics for Unusual Observations

Obs	zt	Fit	Resid	Std Resid	
3	-0.2100	0.1215	-0.3315	-2.36	R
5	-0.2300	0.0670	-0.2970	-2.11	R
10	-0.2600	0.0716	-0.3316	-2.36	R
48	0.4700	0.1400	0.3300	2.35	R
61	0.5600	0.1838	0.3762	2.67	R
85	0.5200	0.2198	0.3002	2.13	R
108	-0.0600	0.2299	-0.2899	-2.06	R
110	-0.0600	0.2798	-0.3398	-2.41	R
171	0.7600	0.3733	0.3867	2.74	R
201	0.0100	0.3569	-0.3469	-2.46	R
203	0.0400	0.3608	-0.3208	-2.28	R
215	0.0700	0.3788	-0.3088	-2.19	R
218	0.0400	0.4417	-0.4017	-2.85	R
230	0.7700	0.4596	0.3104	2.20	R
266	0.9000	0.5136	0.3864	2.74	R
270	0.7800	0.4406	0.3394	2.41	R
315	0.9100	0.5892	0.3208	2.28	R
343	0.2600	0.5480	-0.2880	-2.04	R
373	0.9600	0.6515	0.3085	2.19	R
385	0.2400	0.6694	-0.4294	-3.05	R
386	0.3600	0.6935	-0.3335	-2.37	R
433	0.4600	0.7414	-0.2814	-2.00	R
458	0.5100	0.8014	-0.2914	-2.07	R
478	1.0600	0.7731	0.2869	2.04	R
480	1.1100	0.7875	0.3225	2.29	R
481	1.1700	0.8133	0.3567	2.54	R
482	1.3500	0.8374	0.5126	3.65	R
483	1.3000	0.8410	0.4590	3.27	R

R Large residual