The assignment 2 report

Name: Haiyue Wang

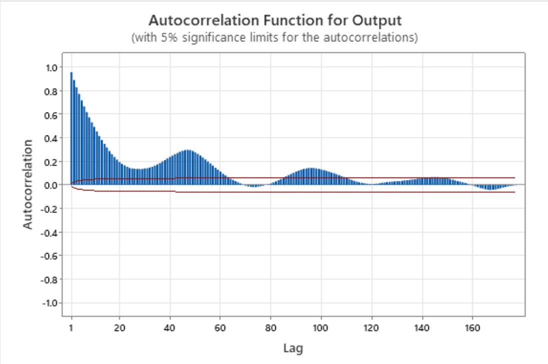
Content

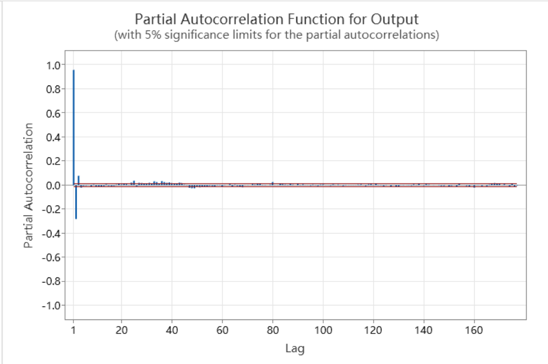
Dataset of Clements Gap Wind Farm Output 3
1.1 Find the best *ARMA(p,q)* model for the 2011 data. 3
1.2 Take the noise Zt from that model and check its *SACF*. 4
1.3 Calculate squared noise and show that it has the *ARCH* effect. 4
1.4 Find the best *ARCH* or *GARCH* model for it. 5
1.5 Apply all models to the 2012 output data. 6
1.6 Evaluate the performance 7
1.7 Compare the results 8
Dataset of Melbourne Airport Rain....................9
2.1 Test normality for December, January, February, July, August 9
2.2 Test for Gamma fit. 9
2.2.1 Get the parameters for gamma 9
2.2.2 Visualize the distribution 10
2.3 Test the correlation 10
2.4 Synthetic and CDFs 11
2.5 Synthetic and CDFs for July, August....................12
Datasets of Mt Gambier By Months Temperature and Mt Gambier Rainfall....................12
3.1 Model the seasonality and then get residuals. 13
3.2 Use exponential smoothing to see the overall trend 14
3.3 Find the trends for the smoothed data of whole series and various sections 16
3.3.1 Find the trend of whole dataset. 16
3.3.2 Split whole dataset into multiple sections. 16
3.3.3 find the trends for the two sections 17
3.4 Take the data for the month of December and the Annual mean temperature from MtGambierByMonthsTemperature.xlsx and find the trend over time. 18
3.5 How much has the mean temperature changed over time in each case? 18

Dataset of Clements Gap Wind Farm Output

1.1 Find the best *ARMA(p, q)* model for the 2011 data.

The first step is to check whether the dataset suit for ARMA model via ACF and PACF. The result like the pictures below.





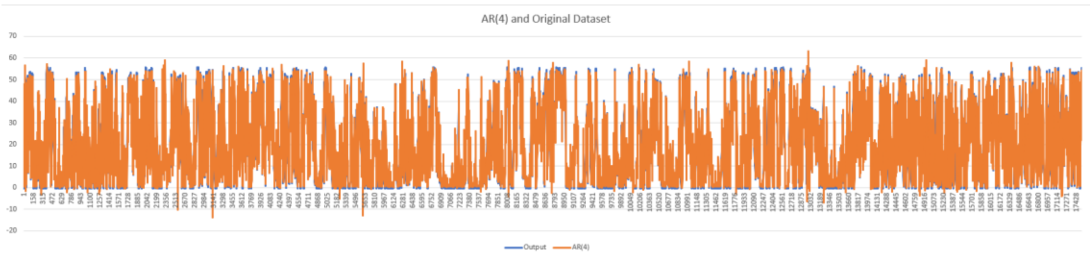
According to the graphs, it's reasonable to use ARMA model for this dataset. Next step is to find a proper ARMA(p, q) model for the dataset. There are five models that could be found for the dataset. The parameters like the picture right.

S	T	U	V	W
AR/ARMA for Output				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.4246	0.0144	-29.55	0
AR 2	0.5552	0.0225	24.64	0
AR 3	0.6599	0.021	31.37	0
MA 1	-1.6758	0.0118	-142.12	0
MA 2	-1.155	0.0186	-62.26	0
MA 3	-0.23619	0.00682	-34.63	0
Constant	4.189	0.153	27.4	0
Mean	19.999	0.73		
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.87	0.221	8.46	0
AR 2	-1.159	0.275	-4.21	0
AR 3	0.3351	0.0809	4.14	0
AR 4	-0.0663	0.0157	-4.21	0
MA 1	0.62	0.221	2.8	0.005
Constant	0.4044	0.0143	28.3	0
Mean	20.004	0.707		
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.24913	0.00753	165.83	0
AR 2	-0.3786	0.0117	-32.28	0
AR 3	0.0787	0.00753	10.45	0
Constant	1.0162	0.0376	27.02	0
Mean	19.999	0.74		
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.25076	0.00755	165.57	0
AR 2	-0.3865	0.0121	-32.01	0
AR 3	0.1046	0.0121	8.67	0
AR 4	-0.02077	0.00756	-2.75	0.006
Constant	1.0374	0.0376	27.59	0
Mean	19.999	0.725		
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.22693	0.00725	169.26	0
AR 2	-0.2821	0.00725	-38.91	0
Constant	1.1032	0.0377	29.25	0
Mean	19.997	0.684		

For selecting a proper model, we should compare the mean of squared error (MSE) and p-value of the parameters for each model, the smaller the better. The table below lists all the MSEs.

	ARMA(3,3)	ARMA(4,1)	AR(4)	AR(3)	AR(2)
mean of square error	overflow	85.89976922	24.75701538	24.76769416	24.92197907

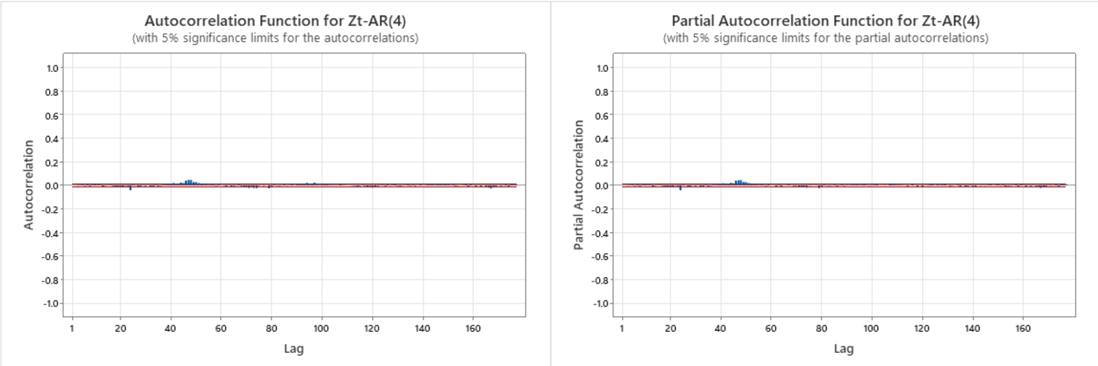
According to the mean of square error, the performance of ***AR(3)*** and ***AR(4)*** are similar, I select the ***AR(4)*** as the best model. The details of ***AR(4)*** result like the picture below.



Visualization of AR(4) with original dataset

1.2 Take the noise Zt from that model and check its *SACF*.

The residuals based on the model AR(4) could get, and name the residuals as Zt-AR(4), then, we could see the SACF according to the ACF and PACF, like the pictures below.

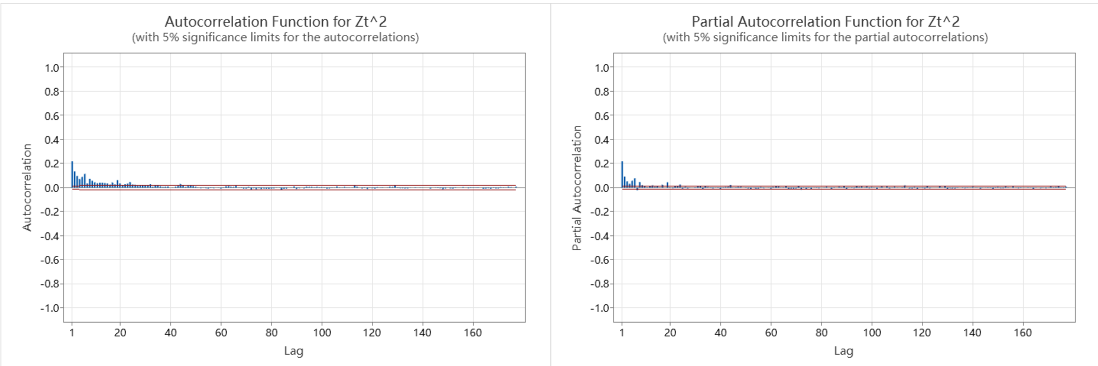


According to the result from ACF and PACF, the **Zt-AR(4)** is not suit for ARMA model.

1.3 Calculate squared noise and show that it has the *ARCH* effect.

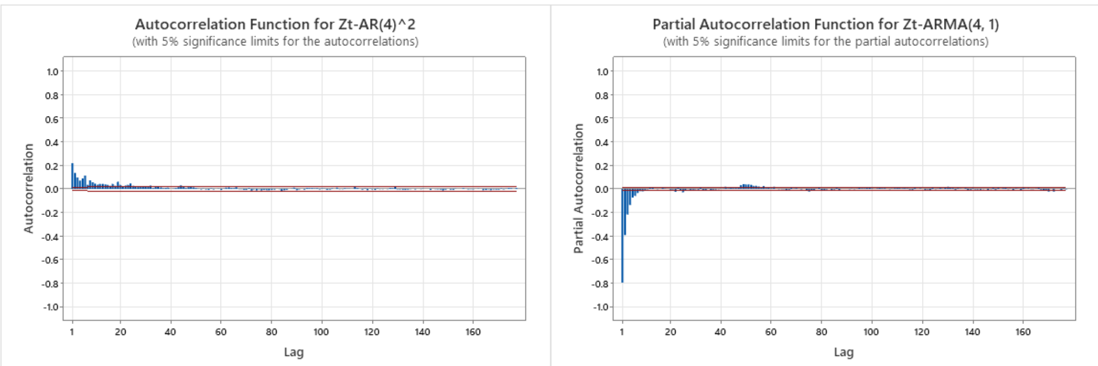
For calculating the **ARCH** effect, it should be separated into two parts, one for **ARCH** model, another for **GARCH** model.

1.3.1 Effect for ARCH model



According to the SACF above, there is ARCH effect for ARCH model.

1.3.2 Effect for GARCH model



According to the SACF above, there is also ARCH effect for GARCH model.

1.4 Find the best *ARCH* or *GARCH* model for it.

For this part, I try to find all possible ARCH and GARCH models, and then try to compare the results for finalizing the model.

1.4.1 For ARCH model

According to the squared residuals for ARCH model, five **AR** models for the dataset could be found, the parameters like the right picture. The coverage rate could be calculated like the picture below based on the parameters of ARCH models.

AG	AH	AI
ARCH(5)	Bounds Coefficient	1.96
	Coverage Rate	0.89578
ARCH(4)	Bounds Coefficient	1.96
	Coverage Rate	0.895786
ARCH(3)	Bounds Coefficient	1.96
	Coverage Rate	0.895792
ARCH(2)	Bounds Coefficient	1.96
	Coverage Rate	0.895798
ARCH(1)	Bounds Coefficient	1.96
	Coverage Rate	0.895804

According to the result, all the coverage rates for the five model are similar, and they all approach 89.6% for score 1.96, which is much less than 95%.

ARCH(5)				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.19027	0.00754	25.22	0
AR 2	0.07629	0.00768	9.94	0
AR 3	0.04172	0.00769	5.42	0
AR 4	0.0215	0.00768	2.8	0.005
AR 5	0.05725	0.00754	7.59	0
Constant	15.178	0.589	25.79	0
Mean	24.761	0.96		
ARCH(4)				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.19213	0.00755	25.44	0
AR 2	0.07894	0.00768	10.27	0
AR 3	0.04624	0.00768	6.02	0
AR 4	0.0325	0.00755	4.3	0
Constant	16.1	0.59	27.31	0
Mean	24.762	0.907		
ARCH(3)				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.19384	0.00755	25.69	0
AR 2	0.08159	0.00766	10.65	0
AR 3	0.05253	0.00755	6.96	0
Constant	16.641	0.59	28.22	0
Mean	24.762	0.878		
ARCH(2)				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.19867	0.00752	26.4	0
AR 2	0.09202	0.00752	12.23	0
Constant	17.564	0.591	29.74	0
Mean	24.763	0.833		
ARCH(1)				
Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.21881	0.00737	29.68	0
Constant	19.345	0.593	32.62	0
Mean	24.763	0.759		

1.4.2 For GARCH model

According to the squared residuals for GARCH model, 10 **ARMA** model for the dataset could be found, the parameters like the picture below.

GARCH(1,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.792	0.0147	53.72	0
MA 1	0.6343	0.0186	34.02	0
Constant	5.15	0.215	23.93	0
Mean	24.76	1.03		
θ	0.1577			

GARCH(4,3)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.9252	0.0328	28.2	0
AR 2	0.6912	0.0357	19.36	0
AR 3	-0.704	0.0175	-40.19	0
AR 4	0.068	0.0105	6.49	0
MA 1	0.7421	0.0319	23.23	0
MA 2	0.7581	0.0287	26.43	0
MA 3	-0.5453	0.0151	-36.08	0
Constant	0.4857	0.0265	18.34	0
Mean	24.78	1.35		

GARCH(3,5)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1.697	0.0264	-64.34	0
AR 2	-1.1881	0.0299	-39.74	0
AR 3	-0.3447	0.0341	-10.1	0
MA 1	-1.8959	0.0255	-74.25	0
MA 2	-1.6361	0.0322	-50.75	0
MA 3	-0.8184	0.0309	-26.48	0
MA 4	-0.2703	0.0155	-17.48	0
MA 5	-0.053	0.0104	-5.11	0
Constant	104.74	3.35	31.28	0
Mean	24.763	0.792		

GARCH(2,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.039	0.0195	53.29	0
AR 2	-0.1077	0.0104	-10.38	0
MA 1	0.8536	0.0175	48.65	0
Constant	1.7032	0.086	19.81	0
Mean	24.77	1.25		

GARCH(3,4)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.4777	0.0616	7.75	0
AR 2	-0.3075	0.0739	-4.16	0
AR 3	0.6795	0.0489	13.91	0
MA 1	0.2938	0.0618	4.75	0
MA 2	-0.3275	0.0666	-4.92	0
MA 3	0.6068	0.0395	15.36	0
MA 4	0.09922	0.00983	10.09	0
Constant	3.724	0.192	19.38	0
Mean	24.77	1.28		

GARCH(3,3)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.217	0.204	5.98	0
AR 2	-0.62	0.255	-2.43	0.015
AR 3	0.3133	0.0922	3.4	0.001
MA 1	1.037	0.203	5.11	0
MA 2	-0.513	0.219	-2.35	0.019
MA 3	0.289	0.0689	4.19	0
Constant	2.221	0.11	20.23	0
Mean	24.77	1.22		

GARCH(2,5)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1.3777	0.0279	-49.46	0
AR 2	-0.5514	0.0789	-6.99	0
MA 1	-1.5753	0.0276	-56.98	0
MA 2	-0.9339	0.076	-12.29	0
MA 3	-0.3236	0.0199	-16.25	0
MA 4	-0.1523	0.0151	-10.06	0
MA 5	-0.02124	0.00919	-2.31	0.021
Constant	72.53	2.36	30.68	0
Mean	24.763	0.807		

GARCH(3,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.0712	0.023	46.52	0
AR 2	-0.0926	0.0116	-7.98	0
AR 3	-0.02906	0.00918	-3.17	0.002
MA 1	0.888	0.0216	41.04	0
Constant	1.2491	0.0658	18.99	0
Mean	24.78	1.3		

GARCH(2,4)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1.252	0.116	-10.82	0
AR 2	-0.4643	0.0613	-7.57	0
MA 1	-1.45	0.115	-12.57	0
MA 2	-0.8231	0.068	-12.1	0
MA 3	-0.2871	0.024	-11.98	0
MA 4	-0.118	0.0171	-6.89	0
Constant	67.25	2.17	30.98	0
Mean	24.763	0.799		

GARCH(2,2)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.2923	0.022	58.73	0
AR 2	-0.3301	0.0108	-30.69	0
MA 1	1.1084	0.0199	55.81	0
MA 2	-0.19276	0.00834	-23.11	0
Constant	0.9367	0.0495	18.91	0
Mean	24.78	1.31		

For some models will occurs negative values, which will lead to the specified model unavailable. The coverage rate for each model like the picture below.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Original and Forecast component				GARCH(1,1)					GARCH(2,2)				
2	Output	AR(4)	Zt-AR(4)	Zt-AR(4)^2	Variance	StDev	LB	UB	Coverage	Variance	StDev	LB	UB	Coverage
3	1.453													
4	-0.444													
5	-0.396													
6	0.768													
7	2.149	2.07	0.0741	0.005491										
8	0.684	3.4	-2.7118	7.353945	24.7609	4.98	-6	13	1					
9	-0.384	1.15	-1.5355	2.357825	41.6255	6.45	##	14	1	2.28834	1.513	-2	4	1
10	-0.336	0.5	-0.8374	0.701194	51.5349	7.18	##	15	1	2.89671	1.702	-3	4	1
11	-0.12	0.79	-0.9124	0.832531	57.5592	7.59	##	16	1	3.51143	1.874	-3	4	1
12	-0.408	0.96	-1.371	1.879542	61.4011	7.84	##	16	1	4.32721	2.08	-3	5	1
13	4.85	0.55	4.3043	18.52702	64.0031	8	##	16	1	5.28742	2.299	-4	5	1
14	8.452	7.26	1.19588	1.430125	68.2789	8.26	-9	23	1	9.11215	3.019	1	13	1
15	17.58	9.69	7.88245	62.13309	68.2948	8.26	-7	26	1	7.7359	2.781	4	15	0
16	17.86	20.3	-2.4058	5.788077	77.8778	8.82	3	38	1	18.9846	4.357	12	29	1
17	20.47	17.4	3.0981	9.598243	75.0707	8.66	0	34	1	13.0191	3.608	10	24	1
18	9.881	21.4	-11.518	132.6608	73.891	8.6	5	38	1	12.6778	3.561	14	28	0
19	39.05	6.99	32.0671	1028.301	92.5496	9.62	##	26	0	35.5573	5.963	-5	19	0
20	29.33	47.8	-18.507	342.4992	245.627	15.7	17	79	1	208.789	14.45	20	76	1
21	17.31	23.2	-5.9237	35.09041	234.573	15.3	-7	53	1	147.264	12.14	-1	47	1
22	9.677	15.2	-5.5582	30.89308	179.084	13.4	##	41	1	83.3318	9.129	-3	33	1
23	2.005	8.71	-6.7012	44.90584	143.225	12	##	32	1	65.7771	8.11	-7	25	1

	AE	AF	AG	AH	AI	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
1	GARCH(3,3)					GARCH(3, 1)					GARCH(3,4)				
2	Variance	StDev	LB	UB	Coverage	Variance	StDev	LB	UB	Coverage	Variance	StDev	LB	UB	Coverage
3															
4															
5															
6															
7															
8															
9															
10	1.85867	1.363	-2	3	1	0.999919	1	-1	2	1					
11	4.201069	2.05	-3	5	1	1.833446	1.35	-2	3	1	4.43474	2.106	-3	5	1
12	5.756134	2.399	-4	6	1	2.896271	1.7	-2	4	1	5.36547	2.316	-4	6	1
13	6.838394	2.615	-5	6	1	4.067852	2.02	-3	4	0	4.26127	2.064	-3	5	0
14	10.72761	3.275	1	14	1	8.057264	2.84	2	13	1	9.41501	3.068	1.2	13	1
15	9.821662	3.134	4	16	0	6.895727	2.63	5	15	0	9.56052	3.092	3.6	16	0
16	20.36023	4.512	11	29	1	18.08446	4.25	12	29	1	19.3694	4.401	12	29	1
17	15.8247	3.978	10	25	1	12.5734	3.55	10	24	1	14.8305	3.851	9.8	25	1
18	13.64306	3.694	14	29	0	11.83111	3.44	15	28	0	14.8712	3.856	14	29	0
19	37.12747	6.093	-5	19	0	35.00158	5.92	-5	19	0	40.9471	6.399	-6	20	0
20	209.4293	14.47	19	76	1	208.1518	14.4	20	76	1	214.26	14.64	19	77	1
21	159.141	12.62	-1	48	1	149.758	12.2	-1	47	1	156.955	12.53	-1	48	1
22	65.19936	8.075	-1	31	1	79.06497	8.89	-2	33	1	94.0499	9.698	-4	34	1
23	58.84729	7.671	-6	24	1	63.916	7.99	-7	24	1	145.312	12.05	-15	32	1

BE	BF	BG
GARCH(1,1)	Bounds Coefficient	1.96
	Coverage Rate	0.991436
GARCH(4,3)	Bounds Coefficient	1.96
	Coverage Rate	#NUM!
GARCH(3,5)	Bounds Coefficient	1.96
	Coverage Rate	#NUM!
GARCH(2,1)	Bounds Coefficient	1.96
	Coverage Rate	#NUM!
GARCH(3,4)	Bounds Coefficient	1.96
	Coverage Rate	0.947924
GARCH(3,3)	Bounds Coefficient	1.96
	Coverage Rate	0.947924
GARCH(2,5)	Bounds Coefficient	1.96
	Coverage Rate	#NUM!
GARCH(3,1)	Bounds Coefficient	1.96
	Coverage Rate	0.947296
GARCH(2,4)	Bounds Coefficient	1.96
	Coverage Rate	#NUM!
GARCH(2,2)	Bounds Coefficient	1.96
	Coverage Rate	0.947528

According to the result the best model for the residuals is ***GARCH(1,1)***, the coverage could be 99.15%.

1.5 Apply all models to the 2012 output data.

According to previous steps, we got two models for the dataset. One is AR(4), another one is GARCH(1,1). All the models will apply to 2012 dataset. The result likes the picture below.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Output	Model	Zt	Zt^2	Model	StDev	LB	UB			Width
2	1/01/2012	55.35891										18.87457
3	1/01/2012	54.77062										
4	1/01/2012	51.58908										
5	1/01/2012	44.87781										
6	1/01/2012	47.13491	41.8088	5.326112	28.36747							
7	1/01/2012	49.53607	46.90522	2.630858	6.921415	9.623549	3.102185	40.82493	52.9855	TRUE	1	12.16056
8	1/01/2012	48.08337	48.40021	-0.31684	0.100388	12.34572	3.513648	41.51346	55.28696	TRUE	1	13.7735
9	1/01/2012	48.17942	46.03066	2.148752	4.617136	12.99672	3.605097	38.96467	53.09665	TRUE	1	14.13198
10	1/01/2012	48.65965	46.91655	1.743105	3.038414	14.12194	3.757918	39.55103	54.28206	TRUE	1	14.73104
11	1/01/2012	48.22744	47.27826	0.949186	0.900954	14.58671	3.819255	39.79252	54.764	TRUE	1	14.97148
12	1/01/2012	42.1645	46.59228	-4.42778	19.60524	14.54443	3.813716	39.11739	54.06716	TRUE	1	14.94977
13	1/01/2012	34.3487	39.22427	-4.87557	23.7712	17.46728	4.179387	31.03267	47.41587	TRUE	1	16.3832
14	1/01/2012	26.22075	31.73673	-5.51598	30.42607	19.97821	4.469699	22.97612	40.49734	TRUE	1	17.52122
15	1/01/2012	28.53787	23.96621	4.57166	20.90008	22.62037	4.756088	14.64428	33.28814	TRUE	1	18.64386
16	1/01/2012	37.39818	29.31423	8.083953	65.3503	22.79404	4.774311	19.95658	38.67188	TRUE	1	18.7153
17	1/01/2012	34.63684	38.81293	-4.17609	17.43974	29.914	5.46937	28.09296	49.53289	TRUE	1	21.43993
18	1/01/2012	29.49834	32.34583	-2.84749	8.108217	26.8747	5.184081	22.18503	42.50663	TRUE	1	20.3216
19	1/01/2012	24.37184	27.86472	-3.49288	12.20019	23.47529	4.84513	18.36826	37.36118	TRUE	1	18.99291
20	1/01/2012	21.43042	22.96587	-1.53546	2.35763	21.96434	4.686613	13.78011	32.15163	TRUE	1	18.37152
21	1/01/2012	8.716238	20.78811	-12.0719	145.73	19.45378	4.410644	12.14324	29.43297	FALSE	0	17.28973
22	1/01/2012	10.61316	5.593081	5.020079	25.20119	40.47116	6.361695	-6.87584	18.062	TRUE	1	24.93784

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.25076	0.00755	165.57	0
AR 2	-0.3865	0.0121	-32.01	0
AR 3	0.1046	0.0121	8.67	0
AR 4	-0.02077	0.00756	-2.75	0.006
Constant	1.0374	0.0376	27.59	0
Mean	19.999	0.725		

GARCH(1,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.792	0.0147	53.72	0
MA 1	0.6343	0.0186	34.02	0
Constant	5.15	0.215	23.93	0
Mean	24.76	1.03		
θ	0.1577			

Bounds Coefficient	1.96
Coverage Rate	0.947474

According to the result of coverage also approaches 95% that means quite well.

1.6 Evaluate the performance.

The score of 90% is about 1.65 and the score of 95% is about 1.96. The 95% result like the picture below.

The 95% coverage result like the picture below.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Output	Model	Zt	Zt^2	Model	StDev	LB	UB			Width
2	1/01/2012	55.35891										18.87457
3	1/01/2012	54.77062										
4	1/01/2012	51.58908										
5	1/01/2012	44.87781										
6	1/01/2012	47.13491	41.8088	5.326112	28.36747							
7	1/01/2012	49.53607	46.90522	2.630858	6.921415	9.623549	3.102185	40.82493	52.9855	TRUE	1	12.16056
8	1/01/2012	48.08337	48.40021	-0.31684	0.100388	12.34572	3.513648	41.51346	55.28696	TRUE	1	13.7735
9	1/01/2012	48.17942	46.03066	2.148752	4.617136	12.99672	3.605097	38.96467	53.09665	TRUE	1	14.13198
10	1/01/2012	48.65965	46.91655	1.743105	3.038414	14.12194	3.757918	39.55103	54.28206	TRUE	1	14.73104
11	1/01/2012	48.22744	47.27826	0.949186	0.900954	14.58671	3.819255	39.79252	54.764	TRUE	1	14.97148
12	1/01/2012	42.1645	46.59228	-4.42778	19.60524	14.54443	3.813716	39.11739	54.06716	TRUE	1	14.94977
13	1/01/2012	34.3487	39.22427	-4.87557	23.7712	17.46728	4.179387	31.03267	47.41587	TRUE	1	16.3832
14	1/01/2012	26.22075	31.73673	-5.51598	30.42607	19.97821	4.469699	22.97612	40.49734	TRUE	1	17.52122
15	1/01/2012	28.53787	23.96621	4.57166	20.90008	22.62037	4.756088	14.64428	33.28814	TRUE	1	18.64386
16	1/01/2012	37.39818	29.31423	8.083953	65.3503	22.79404	4.774311	19.95658	38.67188	TRUE	1	18.7153
17	1/01/2012	34.63684	38.81293	-4.17609	17.43974	29.914	5.46937	28.09296	49.53289	TRUE	1	21.43993
18	1/01/2012	29.49834	32.34583	-2.84749	8.108217	26.8747	5.184081	22.18503	42.50663	TRUE	1	20.3216
19	1/01/2012	24.37184	27.86472	-3.49288	12.20019	23.47529	4.84513	18.36826	37.36118	TRUE	1	18.99291
20	1/01/2012	21.43042	22.96587	-1.53546	2.35763	21.96434	4.686613	13.78011	32.15163	TRUE	1	18.37152
21	1/01/2012	8.716238	20.78811	-12.0719	145.73	19.45378	4.410644	12.14324	29.43297	FALSE	0	17.28973
22	1/01/2012	10.61316	5.593081	5.020079	25.20119	40.47116	6.361695	-6.87584	18.062	TRUE	1	24.93784

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.25076	0.00755	165.57	0
AR 2	-0.3865	0.0121	-32.01	0
AR 3	0.1046	0.0121	8.67	0
AR 4	-0.02077	0.00756	-2.75	0.006
Constant	1.0374	0.0376	27.59	0
Mean	19.999	0.725		

GARCH(1,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.792	0.0147	53.72	0
MA 1	0.6343	0.0186	34.02	0
Constant	5.15	0.215	23.93	0
Mean	24.76	1.03		
θ	0.1577			

Bounds Coefficient	1.96
Coverage Rate	0.947474

The 90% coverage result like the picture below.

B	C	D	E	F	G	H	I	J	K	L
Output	Model	Zt	Zt^2	Model	StDev	LB	UB			Width
55.35891										15.88931
54.77062										
51.58908										
44.87781										
47.13491	41.8088	5.326112	28.36747							
49.53607	46.90522	2.630858	6.921415	9.623549	3.102185	41.78661	52.02382	TRUE	1	10.23721
48.08337	48.40021	-0.31684	0.100388	12.34572	3.513648	42.60269	54.19773	TRUE	1	11.59504
48.17942	46.03066	2.148752	4.617136	12.99672	3.605097	40.08225	51.97907	TRUE	1	11.89682
48.65965	46.91655	1.743105	3.038414	14.12194	3.757918	40.71598	53.11711	TRUE	1	12.40113
48.22744	47.27826	0.949186	0.900954	14.58671	3.819255	40.97649	53.58003	TRUE	1	12.60354
42.1645	46.59228	-4.42778	19.60524	14.54443	3.813716	40.29964	52.88491	TRUE	1	12.58526
34.3487	39.22427	-4.87557	23.7712	17.46728	4.179387	32.32828	46.12026	TRUE	1	13.79198
26.22075	31.73673	-5.51598	30.42607	19.97821	4.469699	24.36173	39.11173	TRUE	1	14.75001
28.53787	23.96621	4.57166	20.90008	22.62037	4.756088	16.11867	31.81376	TRUE	1	15.69509
37.39818	29.31423	8.083953	65.3503	22.79404	4.774311	21.43661	37.19184	FALSE	0	15.75523
34.63684	38.81293	-4.17609	17.43974	29.914	5.46937	29.78847	47.83739	TRUE	1	18.04892
29.49834	32.34583	-2.84749	8.108217	26.8747	5.184081	23.7921	40.89956	TRUE	1	17.10747
24.37184	27.86472	-3.49288	12.20019	23.47529	4.84513	19.87026	35.85919	TRUE	1	15.98893
21.43042	22.96587	-1.53546	2.35763	21.96434	4.686613	15.23296	30.69878	TRUE	1	15.46582
8.716238	20.78811	-12.0719	145.73	19.45378	4.410644	13.51054	28.06567	FALSE	0	14.55513
10.61316	5.593081	5.020079	25.20119	40.47116	6.361695	-4.90371	16.08988	TRUE	1	20.99359

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.25076	0.00755	165.57	0
AR 2	-0.3865	0.0121	-32.01	0
AR 3	0.1046	0.0121	8.67	0
AR 4	-0.02077	0.00756	-2.75	0.006
Constant	1.0374	0.0376	27.59	0
Mean	19.999	0.725		

GARCH(1,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.792	0.0147	53.72	0
MA 1	0.6343	0.0186	34.02	0
Constant	5.15	0.215	23.93	0
Mean	24.76	1.03		
θ	0.1577			

Bounds Coefficient	1.65
Coverage Rate	0.915672

The statistical summary for the residuals of 2012 data likes the right picture. According to the statistical summary, the standard deviation is about 5.05. The mean prediction interval width of 95% coverage is about 18.87, for 90% coverage is about 15.89. The real coverage with 1.96 score is 94.75%, and 91.57% for 1.65 score. So we could get the formulas like below:

T	U
Zt	
Mean	-0.01922
Standard	0.038172
Median	-0.54386
Mode	-1.0374
Standard	5.051877
Sample V	25.52146
Kurtosis	8.370171
Skewness	0.248611
Range	99.58921
Minimum	-49.3423
Maximum	50.24694
Sum	-336.695
Count	17515

94.75% ≈ 95%

91.57% > 90%

5.05*1.96*2=19.8>18.87

5.05*1.65*2=16.665>15.89

The results suggest that the model is quiet well for the dataset.

1.7 Compare the results.

The picture below shows two types of error bounds, the ***GARCH(1,1)*** and ***Quantile.***

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Original Data		AR(4)			GARCH(1,1)						Quantile		
2	Date	Output	Model	Zt	Zt^2	Model	StDev	LB	UB	Coverage	Width	LB	UP	Coverage
3	1/01/2012	55.35891									18.87457			
4	1/01/2012	54.77062												
5	1/01/2012	51.58908												
6	1/01/2012	44.87781												
7	1/01/2012	47.13491	41.8088	5.326112	28.36747									
8	1/01/2012	49.53607	46.90522	2.630858	6.921415	9.623549	3.102185	40.82493	52.9855	1	12.16056	37.05942	58.12175	1
9	1/01/2012	48.08337	48.40021	-0.31684	0.100388	12.34572	3.513648	41.51346	55.28696	1	13.7735	38.55442	59.61674	1
10	1/01/2012	48.17942	46.03066	2.148752	4.617136	12.99672	3.605097	38.96467	53.09665	1	14.13198	36.18487	57.2472	1
11	1/01/2012	48.65965	46.91655	1.743105	3.038414	14.12194	3.757918	39.55103	54.28206	1	14.73104	37.07075	58.13308	1
12	1/01/2012	48.22744	47.27826	0.949186	0.900954	14.58671	3.819255	39.79252	54.764	1	14.97148	37.43246	58.49479	1
13	1/01/2012	42.1645	46.59228	-4.42778	19.60524	14.54443	3.813716	39.11739	54.06716	1	14.94977	36.74648	57.80881	1
14	1/01/2012	34.3487	39.22427	-4.87557	23.7712	17.46728	4.179387	31.03267	47.41587	1	16.3832	29.37847	50.4408	1
15	1/01/2012	26.22075	31.73673	-5.51598	30.42607	19.97821	4.469699	22.97612	40.49734	1	17.52122	21.89093	42.95326	1
16	1/01/2012	28.53787	23.96621	4.57166	20.90008	22.62037	4.756088	14.64428	33.28814	1	18.64386	14.12041	35.18274	1
17	1/01/2012	37.39818	29.31423	8.083953	65.3503	22.79404	4.774311	19.95658	38.67188	1	18.7153	19.46843	40.53076	1
18	1/01/2012	34.63684	38.81293	-4.17609	17.43974	29.914	5.46937	28.09296	49.53289	1	21.43993	28.96713	50.02946	1
19	1/01/2012	29.49834	32.34583	-2.84749	8.108217	26.8747	5.184081	22.18503	42.50663	1	20.3216	22.50003	43.56236	1
20	1/01/2012	24.37184	27.86472	-3.49288	12.20019	23.47529	4.84513	18.36826	37.36118	1	18.99291	18.01892	39.08125	1
21	1/01/2012	21.43042	22.96587	-1.53546	2.35763	21.96434	4.686613	13.78011	32.15163	1	18.37152	13.12008	34.1824	1
22	1/01/2012	8.716238	20.78811	-12.0719	145.73	19.45378	4.410644	12.14324	29.43297	0	17.28973	10.94231	32.00464	0
23	1/01/2012	10.61316	5.593081	5.020079	25.20119	40.47116	6.361695	-6.87584	18.062	1	24.93784	-4.25272	16.80961	1
24	1/01/2012	24.16775	12.67851	11.48924	132.0026	34.79508	5.898736	1.116986	24.24003	1	23.12304	2.832711	23.89504	0
25	1/01/2012	18.56102	27.63007	-9.06905	82.24766	48.03733	6.930896	14.04551	41.21463	1	27.16911	17.78427	38.8466	1
26	1/01/2012	21.26233	15.84105	5.421283	29.39031	48.59053	6.970691	2.178496	29.5036	1	27.32511	5.995253	27.05758	1
27	1/01/2012	25.69249	22.76515	2.927338	8.569308	40.60583	6.37227	10.2755	35.2548	1	24.9793	12.91935	33.98168	1
28	1/01/2012	20.74608	26.39417	-5.64809	31.90087	32.25766	5.679582	15.26218	37.52615	1	22.26396	16.54837	37.6107	1
29	1/01/2012	8.163968	18.89415	-10.7302	115.1367	30.6418	5.535503	8.044561	29.74373	1	21.69917	9.04835	30.11068	0
30	1/01/2012	1.800878	5.476021	-3.67514	13.50667	42.74316	6.537825	-7.33812	18.29016	1	25.62827	-4.36978	16.69255	1

GARCH PART

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.25076	0.00755	165.57	0
AR 2	-0.3865	0.0121	-32.01	0
AR 3	0.1046	0.0121	8.67	0
AR 4	-0.02077	0.00756	-2.75	0.006
Constant	1.0374	0.0376	27.59	0
Mean	19.999	0.725		

GARCH(1,1)

Final Estimates of Parameters				
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.792	0.0147	53.72	0
MA 1	0.6343	0.0186	34.02	0
Constant	5.15	0.215	23.93	0
Mean	24.76	1.03		
θ	0.1577			

Bounds Coefficient	1.96	1.65
Coverage Rate	0.947474	0.915672
Width	18.87457	15.88931

Quantile

	Boundary	Quantile	Boundary	Quantile
Lower	-9.8458	0.025	-7.28988	0.05
Upper	11.21653	0.975	8.046939	0.95
Width	21.06232911		15.33681395	
Coverage Rate	0.949985727		0.899971453	

Zt	
Mean	-0.01922
Standard	0.038172
Median	-0.54386
Mode	-1.0374
Standard	5.051877
Sample V	25.52146
Kurtosis	8.370171
Skewnes	0.248611
Range	99.58921
Minimum	-49.3423
Maximum	50.24694
Sum	-336.695
Count	17515

The right picture focusses on the results of the two approaches. According to the results the performance of two approaches is similar.

Bounds Coefficient	1.96	1.65
Coverage Rate	0.947474	0.915672
Width	18.87457	15.88931

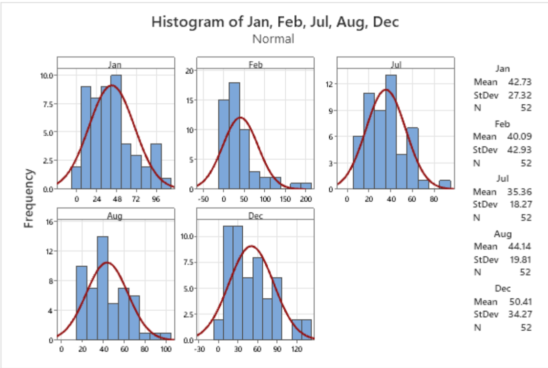
Quantile

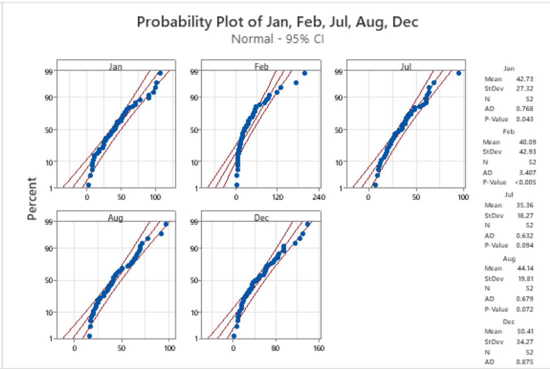
	Boundary	Quantile	Boundary	Quantile
Lower	-9.8458	0.025	-7.28988	0.05
Upper	11.21653	0.975	8.046939	0.95
Width	21.06232911		15.33681395	
Coverage Rate	0.949985727		0.899971453	

Dataset of Melbourne Airport Rain

2.1 Test normality for December, January, February, July, August

For normality test, the pp-plot and histogram could be used for testing.





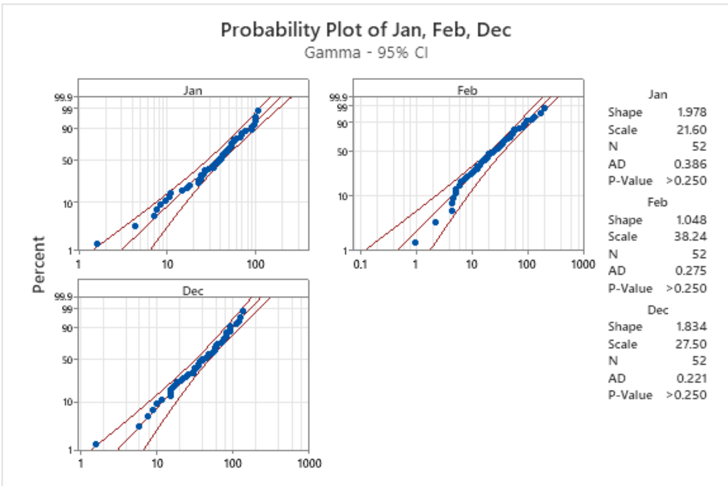
According to the histogram all the distribution of the months are right skewed. According to the pp-plot result, we reject January, February and December datasets follow normal distribution because of p-value is less than 0.05.

2.2 Test for Gamma fit.

There are two steps for this question. The first step is to test the gamma distribution, and then to calculate the α and β parameters. Another step is to get the distribution and visualize them. According to previous step, the datasets of January, February and December will be processed.

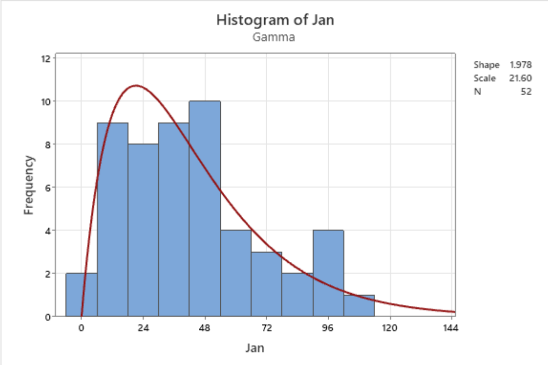
2.2.1 Gamma distribution test

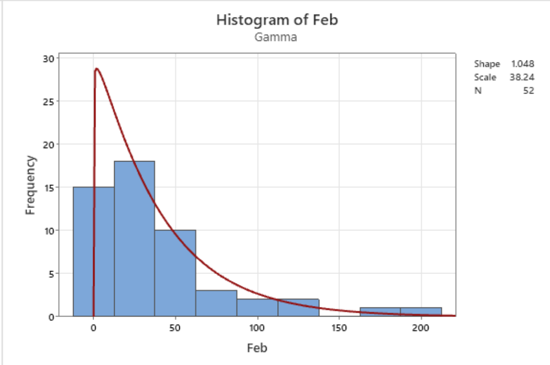
According to the right graph, the p-values are all greater than 0.05, it means we cannot reject the datasets follow gamma distribution. And the parameters of gamma distribution for each dataset are also listed on the graph.

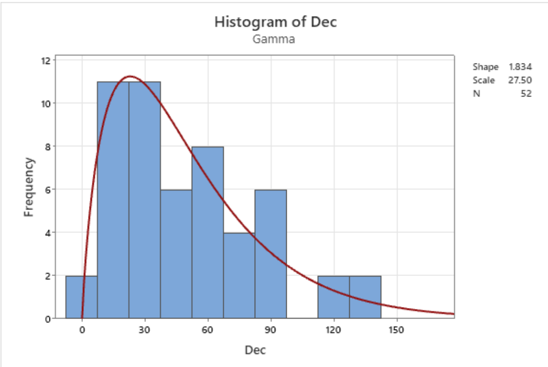


2.2.2 Visualize the distribution.

This part will show the gamma distribution for each month.

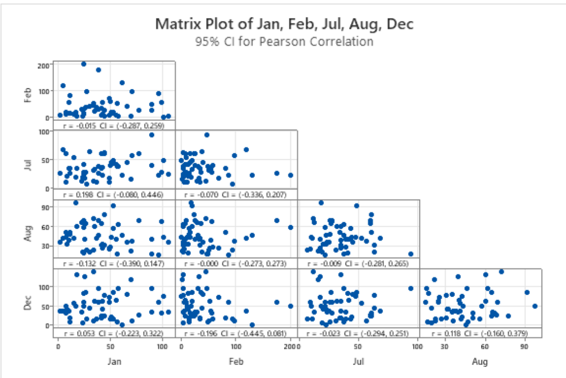






2.3 Test the correlation.

For the correlation test, the Pearson correlation is used here. The Matrix Plot and correlation matrix like the pictures below.



Correlations

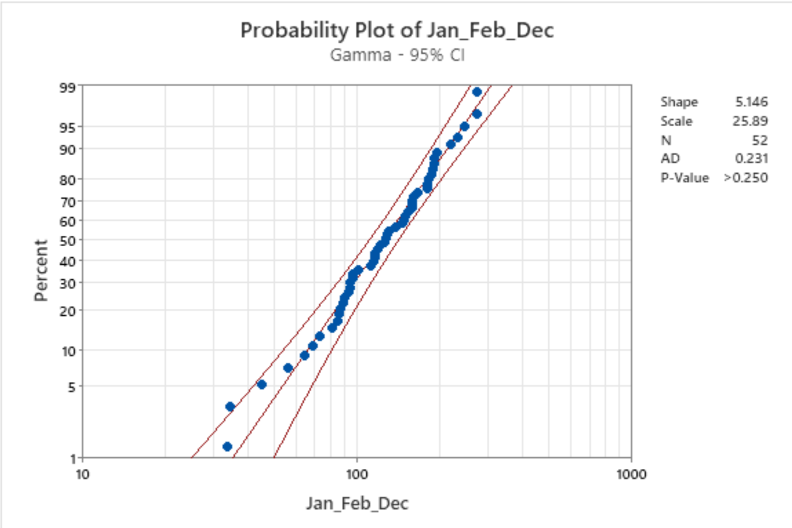
	Jan	Feb	Jul	Aug
Feb	-0.015			
Jul	0.198	-0.070		
Aug	-0.132	-0.000	-0.009	
Dec	0.053	-0.196	-0.023	0.118

According to the correlation matrix, the data of Jan with July have the highest correlation, even the highest correlation is very week, so we regard there is no correlation among the datasets.

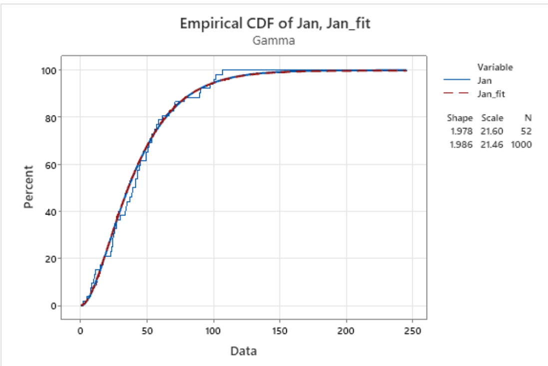
2.4 Synthetic and empirical CDFs for January, February, December and total.

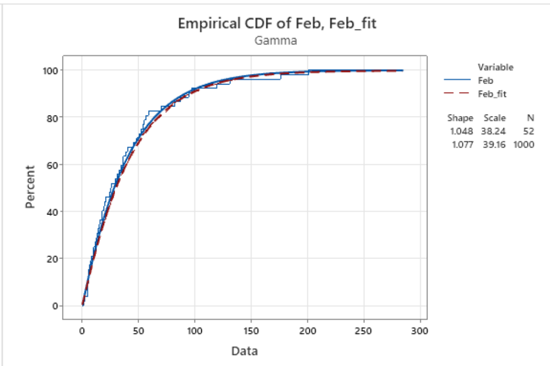
The first step is to get the total data then to test gamma distribution for it. The second step is to generate synthetic data for the datasets that follow gamma distribution, then to do empirical CDFs for them.

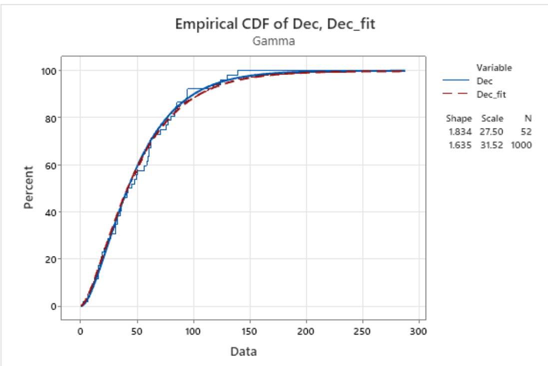
Make sum of January, February, and December to form the total data, then to do gamma test like the picture below.

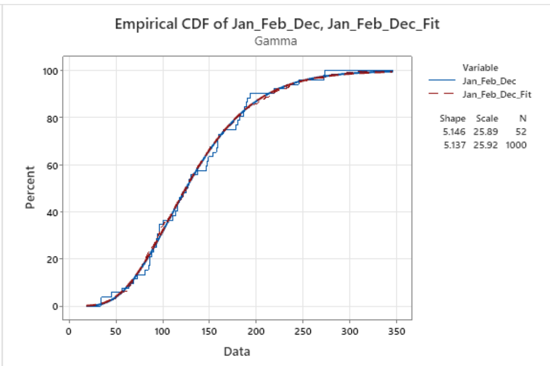


According to this graph above, it can not to reject the total data follow gamma distribution. So, next step is to do empirical CDFs for the four datasets, details like the pictures below.





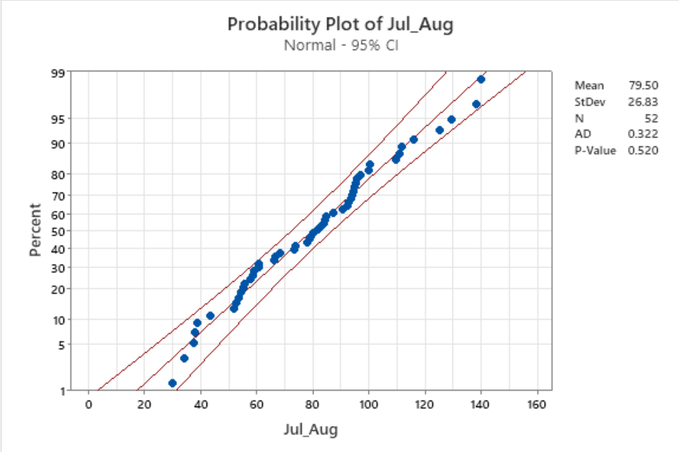


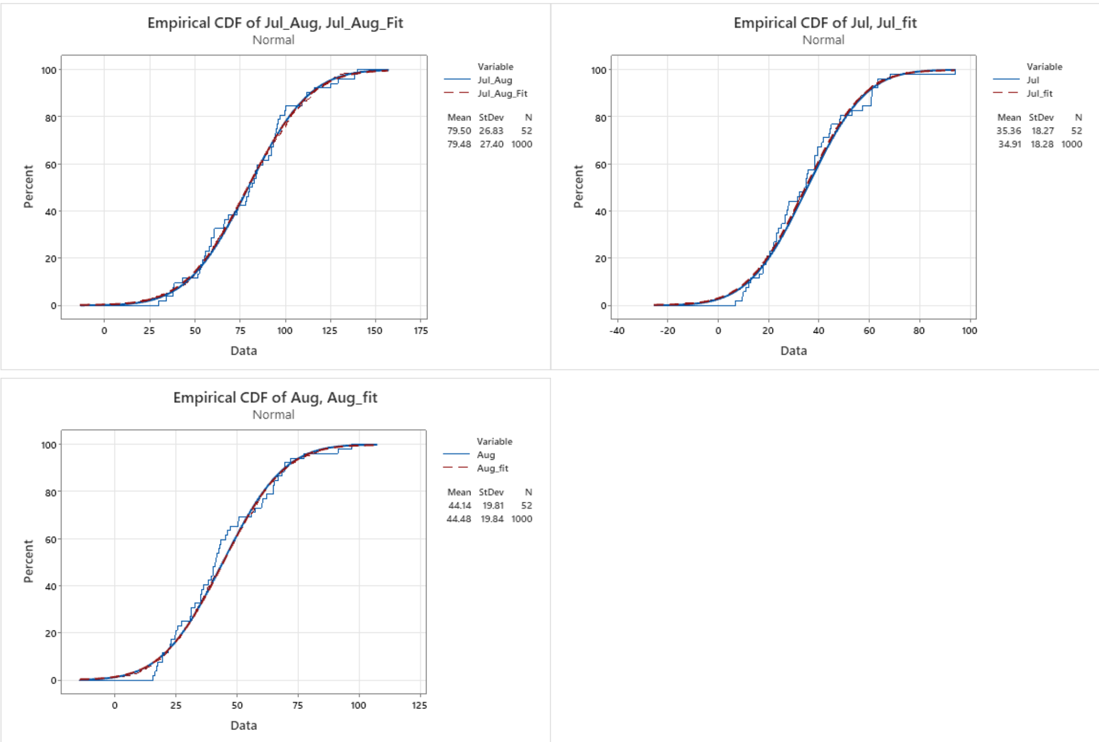


2.5 Synthetic and CDFs for July, August.

The process for this question is the same with chapter 2.4. The difference is the datasets of July and August follow normal distribution. When generate Synthetic data using normal distribution.

Test the total value of July and August, then to test the normality for it, details like the picture below. According to the result, we cannot reject the total data of July and August follows normal distribution because of the p-value is greater than 0.05. The next step is to do empirical CDFs for the three datasets, details like the pictures below.





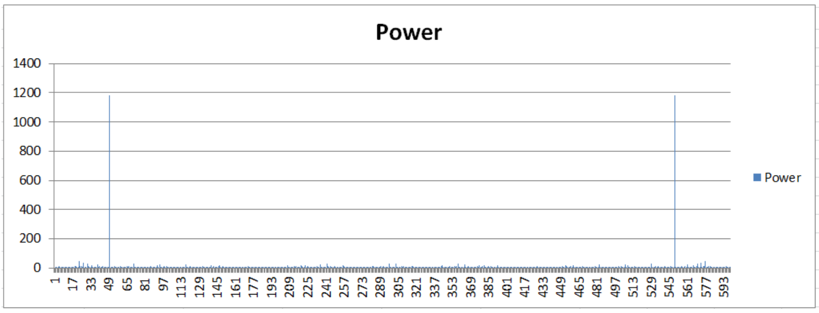
Empirical CDF for July, August, and Sum of them

Datasets of Mt Gambier By Months Temperature and Mt Gambier Rainfall

3.1 Model the seasonality and then get residuals.

There are three steps for this question. The first step is to find the best frequencies, the second step is to find the proper parameters for seasonality, and the last step is to visualize the seasonality result.

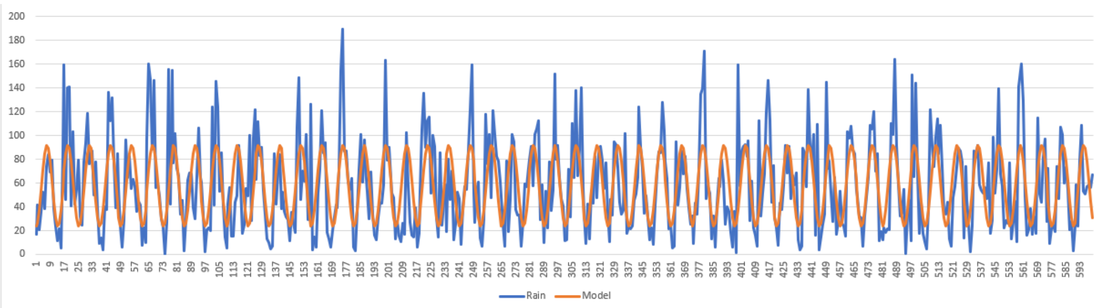
The picture right is the DFT result for the dataset. The 50 and 550 is the best for the dataset.



The right picture is the seasonality parameters using the frequencies got from step1. And also we got the final model and the residuals.

	A	B	C	D	E	F	G	H	I
1					1/year	11/year		n=	600
2				Frequencies	0.523598776	5.759587		SME	520408.3
3	Mean	57.868		alpha	-13.568258	-13.5683			
4	Variance	38.18741		beta	-10.495358	10.49536			
5									
6									
7	Year	Month	Time	Rain	Frequencies		Model	Diff^2	Residuals
8					50	550			
9	1950	1	1	17.1	-16.9981352	-16.9981	23.87173	45.85632	-6.77173
10	1950	2	2	41.9	-15.8733757	-15.8734	26.12125	248.969	15.77875
11	1950	3	3	20.5	-10.495358	-10.4954	36.87728	268.2154	-16.3773
12	1950	4	4	35.7	-2.30511767	-2.30512	53.25776	308.2751	-17.5578

The picture below is the visualization for the final model of seasonality.

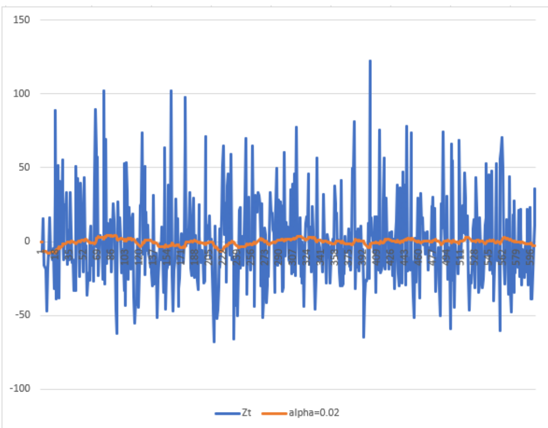


3.2 Use exponential smoothing to see the overall trend

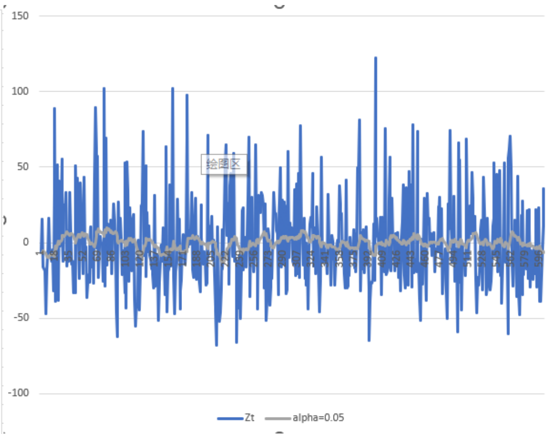
For this question, I will the dataset calculated using exponential smoothing with various values, then to visualize them. The datasets get from exponential smoothing with different values like the pictures below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1					1/year	11/year		n=	600						Exponential Smoothing	
2				Frequencies	0.523598776	5.759587		SME	520408.3						alpha	0.05
3	Mean	57.868		alpha	-13.568258	-13.5683										
4	Variance	38.18741		beta	-10.495358	10.49536										
5																
6	Year	Month	Time	Rain	Frequencies		Seasonality	Diff^2	Residuals	Exponential Smoothing						
7					50	550	Model			0.02	0.05	0.1	0.15	0.2		
8	1950	1	1	17.1	-16.9981352	-16.9981	23.87172968	45.85632	-6.77173							
9	1950	2	2	41.9	-15.8733757	-15.8734	26.1212486	248.969	15.77875	-6.77	-6.77	-6.77	-6.77	-6.77		
10	1950	3	3	20.5	-10.495358	-10.4954	36.87728392	268.2154	-16.3773	-6.32	-5.64	-4.52	-3.39	-2.26		
11	1950	4	4	35.7	-2.30511767	-2.30512	53.25776466	308.2751	-17.5578	-6.52	-6.18	-5.7	-5.34	-5.08		
12	1950	5	5	52.3	6.502777119	6.502777	70.87355424	344.9769	-18.5736	-6.74	-6.75	-6.89	-7.17	-7.58		
13	1950	6	6	38.2	13.56825803	13.56826	85.00451606	2190.663	-46.8045	-6.98	-7.34	-8.06	-8.88	-9.78		
14	1950	7	7	74.1	16.99813516	16.99814	91.86427032	315.5693	-17.7643	-7.78	-9.31	-11.9	-14.6	-17.2		
15	1950	8	8	83.6	15.8733757	15.87338	89.6147514	36.17723	-6.01475	-7.98	-9.74	-12.5	-15	-17.3		
16	1950	9	9	69.6	10.49535804	10.49536	78.85871608	85.72382	-9.25872	-7.94	-9.55	-11.9	-13.7	-15		
17	1950	10	10	78.9	2.305117668	2.305118	62.47823534	269.6744	16.42176	-7.96	-9.54	-11.6	-13	-13.9		
18	1950	11	11	38.4	-6.50277712	-6.50278	44.86244576	41.76321	-6.46245	-7.48	-8.24	-8.8	-8.61	-7.82		
19	1950	12	12	24.1	-13.568258	-13.5683	30.73148394	43.97658	-6.63148	-7.45	-8.15	-8.57	-8.29	-7.55		
20	1951	13	13	11.7	-16.9981352	-16.9981	23.87172968	148.151	-12.1717	-7.44	-8.07	-8.37	-8.04	-7.37		
21	1951	14	14	21.4	-15.8733757	-15.8734	26.1212486	22.29019	-4.72125	-7.53	-8.28	-8.75	-8.66	-8.33		

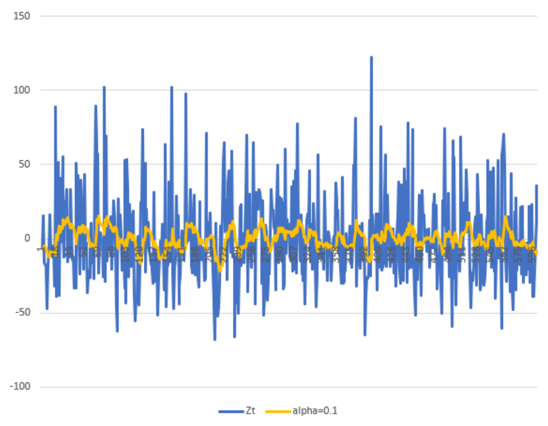
According to the picture above, five different values of alpha are used to calculate. Next step will show the overall trend based on the five parameters.



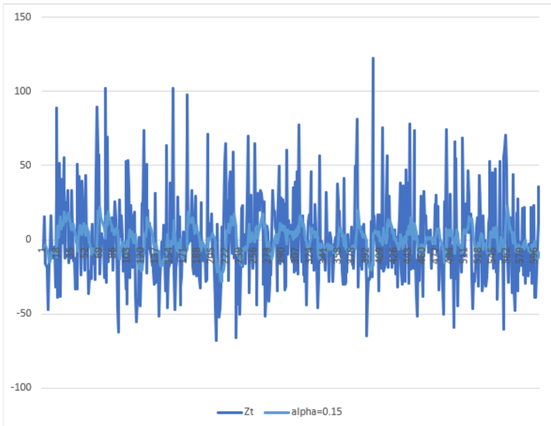
α=0.02



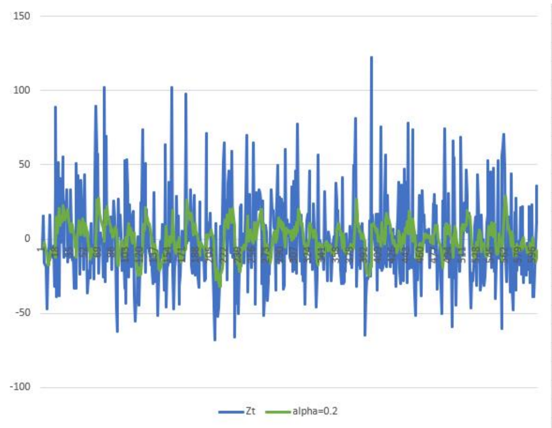
α=0.05



α=0.1



α=0.15



α=0.2

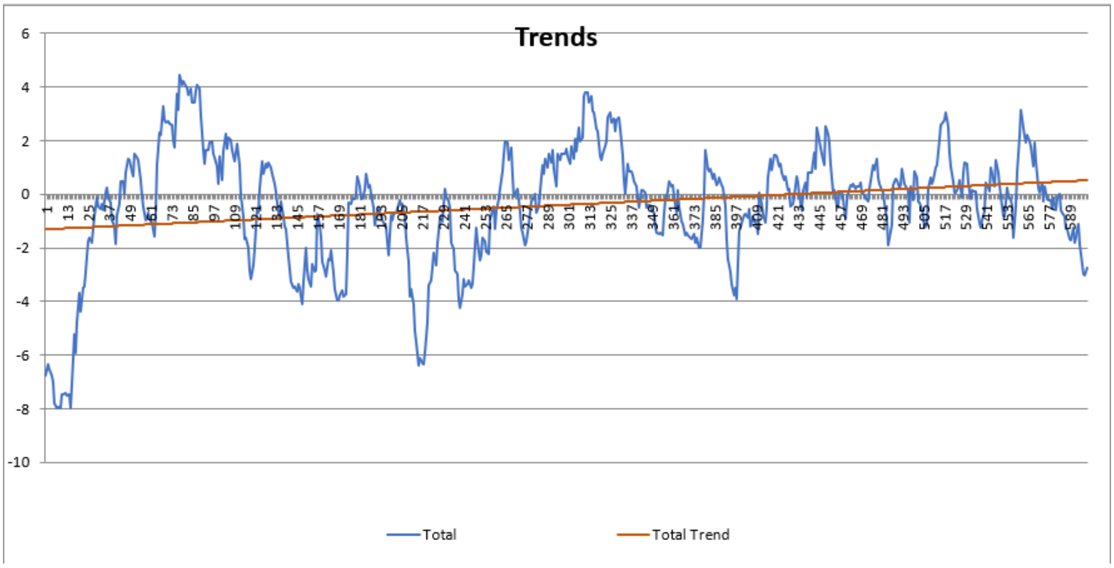
According to the graphs, it is much easier to extract the trend when the α=0.02. For the first several data, the trend is rising, the following data has no obvious changing.

3.3 Find the trends for the smoothed data of whole series and various sections.

I will set the parameter **α** equals 0.02 of smoothed data, and then to process the smoothed data. There are 3 steps to do. The first step is to find the trend of the whole dataset. The second step is to split the dataset into multiple sections, and the last step is to find the trends for each section.

3.3.1 Find the trend of whole dataset.

We could find the trend from the smoothed data and linear regression. The whole dataset is rising based the smoothed dataset got from the previous step. The next step is using linear regression approach to get, I use the univariate linear regression to model the trend. The result like the picture below.



According to the result above, there is a rising trend for the dataset.

3.3.2 Split whole dataset into multiple sections.

According the visualization of the dataset, the dataset could be split into two sections, the first section (from the beginning to 60) rise rapidly, and the second section oscillate around a variable. So, the dataset could be split into two sections like the picture below.

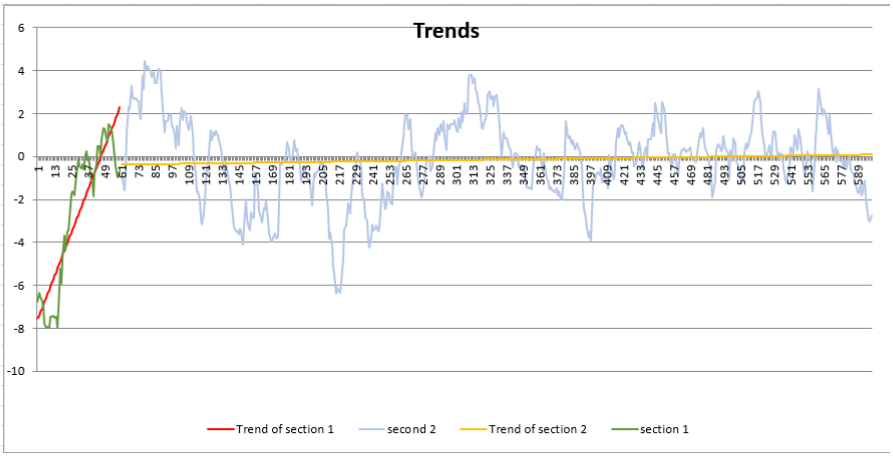
	A	B	C	D	E	F	G	H	I	J	K
1	Time	Exponential Smoothing			Trends			squared error			
2		Total	second1	second 2	second 1	second 2	Total	second1	second2	Total	
60	58	-0.959194581	-0.95919458		1.957334		-1.12994	8.506138		0.029154	
61	59	-0.939259604	-0.9392596		2.1246		-1.12689	9.387234		0.035204	
62	60	-0.709104091	-0.70910409		2.291866		-1.12383	9.005819		0.172001	
63	61	-1.020356603		-1.02036		-0.37164	-1.12078		0.420833	0.010085	
64	62	-1.076374443		-1.07637		-0.37078	-1.11773		0.497863	0.00171	
65	63	-1.586392632		-1.58639		-0.36992	-1.11467		1.479803	0.222518	
66	64	-0.707820073		-0.70782		-0.36906	-1.11162		0.114757	0.163055	
67	65	1.094865244		1.094865		-0.3682	-1.10857		2.140566	4.855117	
68	66	2.316877618		2.316878		-0.36734	-1.10551		7.205038	11.71277	

3.3.3 Find the trends for the two sections

According to the smoothed data itself, we could find the first section has a rapid rising, and the second section data oscillate up and down around 0. Next step we could use univariate linear regression to model the trend, details like the picture below.

L	M	N	O
Trend Parameters			
	second 1	second 2	Total
a	0.167266	0.00086	0.003053
b	-7.74409	-0.38883	-1.30703
SSE	128.5533	1924.723	2737.405

The trends parameters for the two sections



The trends for the two sections

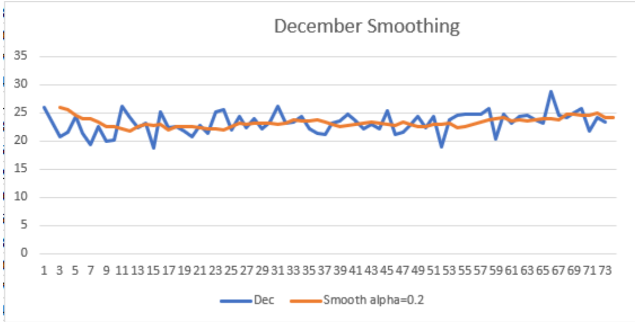
According to the results above, the two sections both have a rising trend, the first section has a rapid rising trend, the second section only has a slight rising trend that could almost be ignored.

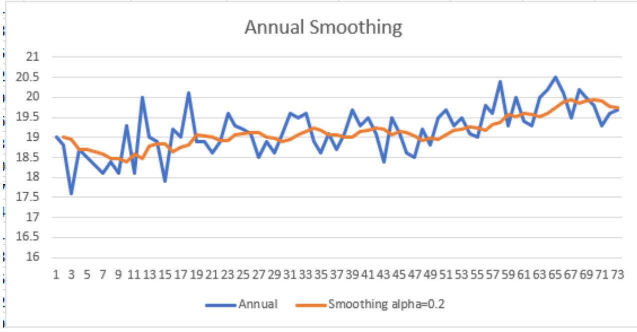
3.4 Take the data for the month of December and the Annual mean temperature from MtGambierByMonthsTemperature.xlsx and find the trend over time.

For finding trending there are two ways to do, the first way is using smoothing method, another approach is using linear regression. I will try to use the two approaches.

3.4.1 Smoothing method

I set the alpha = 0.02 to smoothing the dataset like the picture right. According to the graph, the temperature of December does not have too much change, on the other hand, the temperature of Annual temperature has an obvious rise.

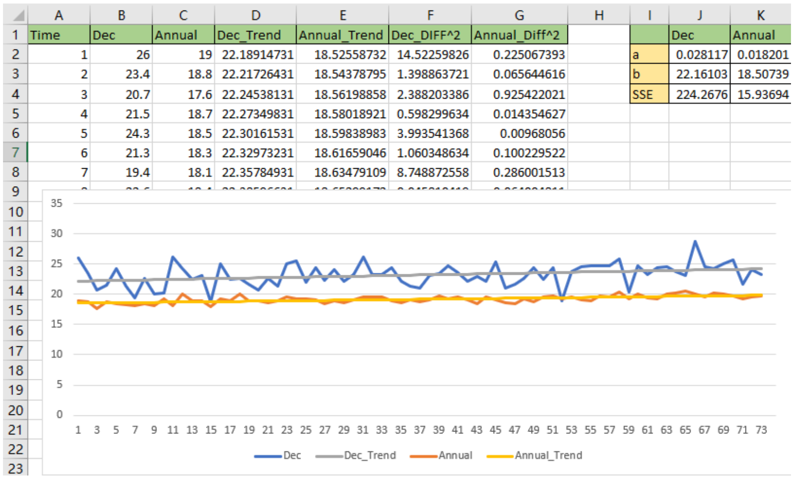




3.4.2 Linear regression

The univariate linear regression to model trends for the two datasets. The results like the picture right.

According to the picture, the temperature of Annual and December are both rise, but the upward trend of Annual is more obvious.



	A	B	C	D	E	F	G	H	I	J	K
1	Time	Dec	Annual	Dec_Trend	Annual_Trend	Dec_DIFF^2	Annual_Diff^2			Dec	Annual
2	1	26	19	22.18914731	18.52558732	14.52259826	0.225067393		a	0.028117	0.018201
3	2	23.4	18.8	22.21726431	18.54378795	1.398863721	0.065644616		b	22.16103	18.50739
4	3	20.7	17.6	22.24538131	18.56198858	2.388203386	0.925422021		SSE	224.2676	15.93694
5	4	21.5	18.7	22.27349831	18.58018921	0.598299634	0.014354627				
6	5	24.3	18.5	22.30161531	18.59838983	3.993541368	0.00968056				
7	6	21.3	18.3	22.32973231	18.61659046	1.060348634	0.100229522				
8	7	19.4	18.1	22.35784931	18.63479109	8.748872558	0.286001513				

3.5 How much has the mean temperature changed over time in each case?

According to the smoothing approach, the mean temperature changed over time is about rise from 18.5 to 20 that approaches 1.5.

According to the linear regression result from the last step, the mean temperature changed over time should be calculated via the linear regression parameters with the whole 73 years on this dataset. The mean temperature changed over time for should be, a*year = 0.018201 * 73 ≈ 1.32.