QUALITY DATA ANALYSIS

08/07/2022

General recommendations:

a) write the solutions in CLEAR and READABLE way on paper and show (qualitatively) all the relevant plots;
b) avoid (if not required) theoretical introductions or explanations covered during the course;
c) always state the assumptions and report all relevant steps/discussion/formulas/expression to present and motivate your solution;
d) when using hypothesis tests provide the numerical value of the test statistic and the test conclusion in terms of p-value.
e) For exams in presence: to access the software on the provided laptops, go on browser → Favourites → Managed favourites → Virtual Desktop and enter your Polimi credentials.
f) Exam duration: 2h 10min
g) Multichance students should skip: point d) of exercise 1, point b) of exercise 2.

Exercise 1 (15 points)

In a thermal process, the cooling profile has a direct effect on the final quality and performance of the material. For Ti6Al4V components, it is known that the temperature of the material during the cooling process follows an exponential decay in the form $temp_t = \beta_1 \cdot e^{-0.1t} + \varepsilon_t$, where time t is expressed in seconds. Table 1 shows the material temperature during the cooling phase for one thermally treated part.

Table 1

Time (s)	Temperature (°C)	Time (s)	Temperature (°C)	Time (s)	Temperature (°C)	Time (s)	Temperature (°C)
1	361,9	11	123,3	21	50,5	31	16
2	322,7	12	114,7	22	46,7	32	18,8
3	298,1	13	104,8	23	38,5	33	10,5
4	265,6	14	104,3	24	29	34	3,4
5	240,7	15	89,8	25	35	35	8,8
6	213,7	16	80	26	39	36	10,9
7	194,5	17	66,4	27	32	37	10,2
8	164,8	18	60,9	28	22,9	38	11,1
9	150,8	19	51,9	29	26,1	39	13,5
10	135,5	20	50,3	30	19,6	40	13,9

a) Is the exponential decay model appropriate for designing a control chart procedure? If not, what is the appropriate model for fitting data in Table 1?
b) Based on the model selected in point a), design a suitable control chart with $ARL_0 = 250$.
c) Using the control chart designed at point b), determine if the cooling process of a different component of the same material (Table 2) is in-control or not.

Table 2

Time (s)	Temperature (°C)	Time (s)	Temperature (°C)	Time (s)	Temperature (°C)	Time (s)	Temperature (°C)
1	373	11	189,1	21	95,8	31	46,6
2	345,7	12	177,8	22	83,2	32	51,6
3	318,4	13	167,8	23	73	33	46,4
4	302	14	150,5	24	73,9	34	34,2
5	280,3	15	136,4	25	66,1	35	32,3
6	262,2	16	123,7	26	64,4	36	29,1
7	241,6	17	120,3	27	65,8	37	30
8	220,5	18	123,9	28	60,6	38	29,6
9	209,8	19	113,9	29	58,4	39	28,5
10	196,7	20	99,6	30	45,5	40	24,1

d) Design and implement a statistical test of hypothesis to check whether the first and second components (referring to Table 1 and Table 2, respectively) have a cooling history that is statistically different or not.

Exercise 2 (15 points)

In a finishing process for the production of an oil & gas component, two critical hole diameters are measured and monitored by randomly sampling $n = 3$ parts every hour. Under in-control conditions, it is known that the two diameters are independent and normally distributed with mean and standard deviation:

$$\mu_1 = 20.5\ mm,\ \mu_2 = 25.5\ mm,\ \sigma_1 = 0.2\ mm,\ \sigma_2 = 0.28\ mm$$

a) Design two univariate control charts for the mean with a familywise Type I error $\alpha = 0.01$ and determine if the data shown in Table 3 are in-control or not.

Table 3

Sample	Diameter hole 1 (mm)			Diameter hole 2 (mm)		
1	20,78	19,89	20,22	24,88	25,19	25,16
2	20,49	20,63	20,35	25,34	25,59	25,42
3	20,59	20,44	20,73	25,56	25,17	25,25
4	20,43	20,46	20,62	25,76	25,23	25,42
5	20,3	20,58	20,24	25,53	25,32	25,37

b) Determine the ARL_0 value if no familywise correction on the Type I error is applied and compare it with the ARL_0 of the chart designed at point a). Discuss the result.

c) In case both the diameters exhibit a shift of the mean $\Delta\mu_1 = \Delta\mu_2 = 0.3\ mm$, determine the probability of detecting it at the first sample after the shift using the control chart designed at point a).

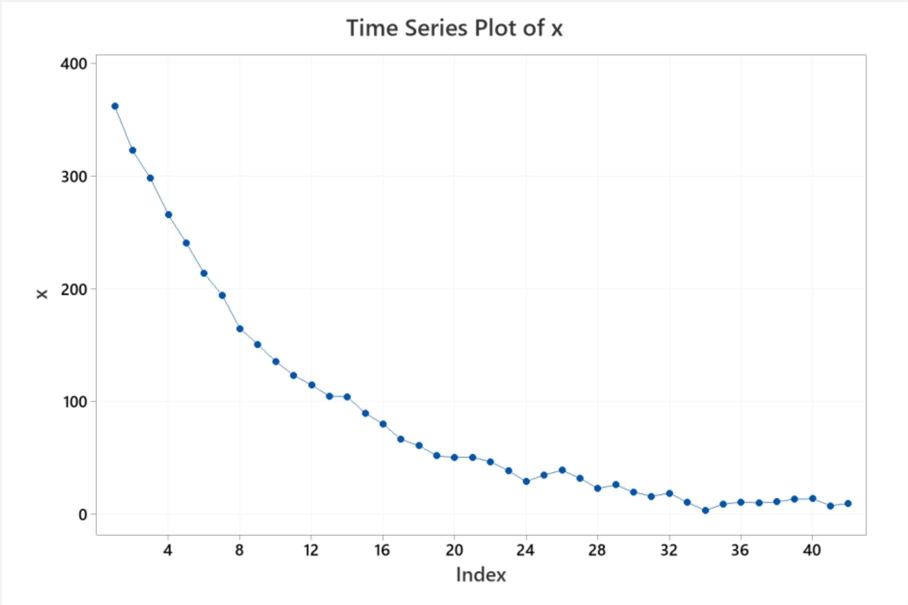
d) What is the minimum sample size to be used to detect a simultaneous shift of the means $\Delta\mu_1 = \Delta\mu_2 = 0.3\ mm$ with a probability $P > 90\%$?

Exercise 3 (3 points)

A quality characteristic X_t follows a stationary AR(1) model $X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with positive autocorrelation coefficient and known σ_ε^2. Let $E(X_t) = \mu$ and $V(X_t) = \sigma^2$. Compute the expressions of ξ and ϕ_1 as functions of μ, σ^2 and σ_ε^2.

Exercise 1 (solution)

a) The cooling process for data in Table 1 is:



By fitting a model in the form $temp_t = \beta_1 \cdot e^{-0.1t} + \varepsilon_t$, we get:

WORKSHEET 1

Regression Analysis: x versus exp2

Regression Equation

x = 392,37 exp2

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
exp2	392,37	2,46	159,32	0,000	1,00

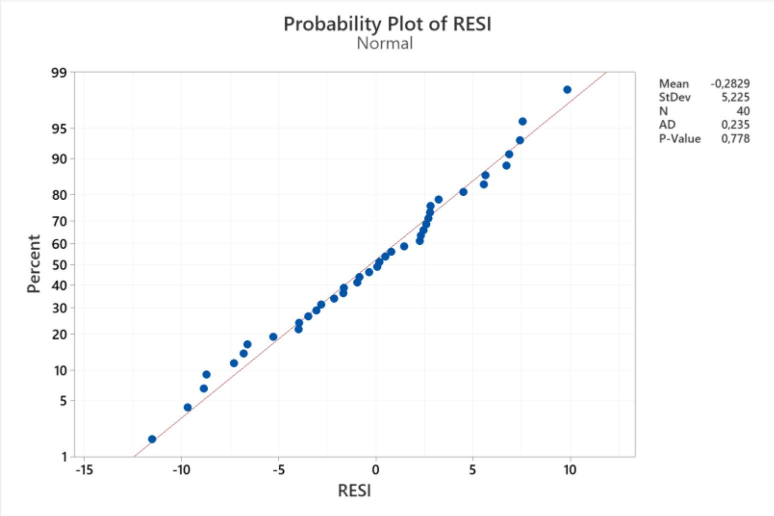
Model Summary

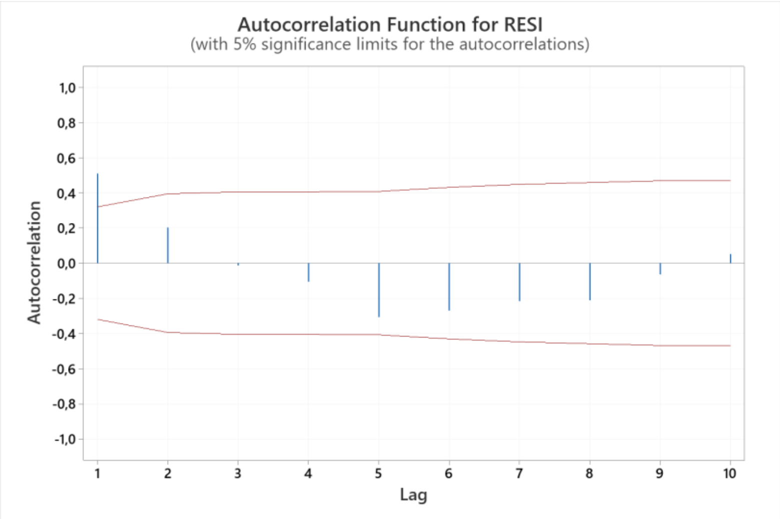
S	R-sq	R-sq(adj)	R-sq(pred)
5,23300	99,85%	99,84%	99,84%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	695111	695111	25383,52	0,000
exp2	1	695111	695111	25383,52	0,000
Error	39	1068	27		
Total	40	696179			

The residuals are normal but not independent:





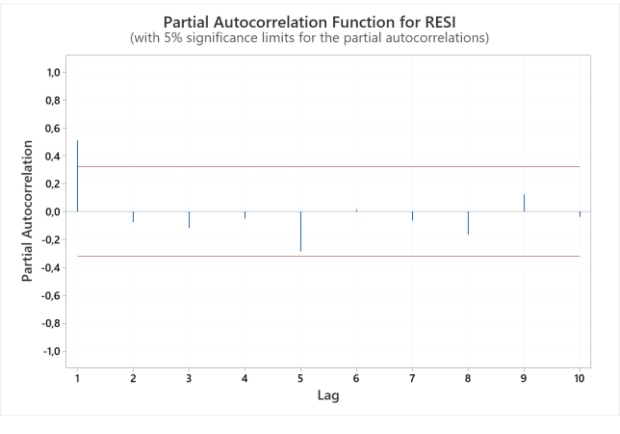
Bartlett's test at lag 1 (95% confidence):

$$|r_k| = 0.509$$

$$\frac{z_{\alpha/2}}{\sqrt{n}} = 0.354$$

The autocorrelation at lag 1 is significant.

PACF:



A more appropriate model should include an AR(1) term, i.e.: $temp_t = \beta_1 \cdot e^{-0.1t} + \beta_2 temp_{t-1} + \varepsilon_t$.

By fitting this model, we get:

WORKSHEET 1

Regression Analysis: x versus exp2; AR1

Method

Rows unused 1

Regression Equation

x = 157,8 exp2 + 0,537 AR1

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
exp2	157,8	59,7	2,65	0,012	680,40
AR1	0,537	0,137	3,91	0,000	680,40

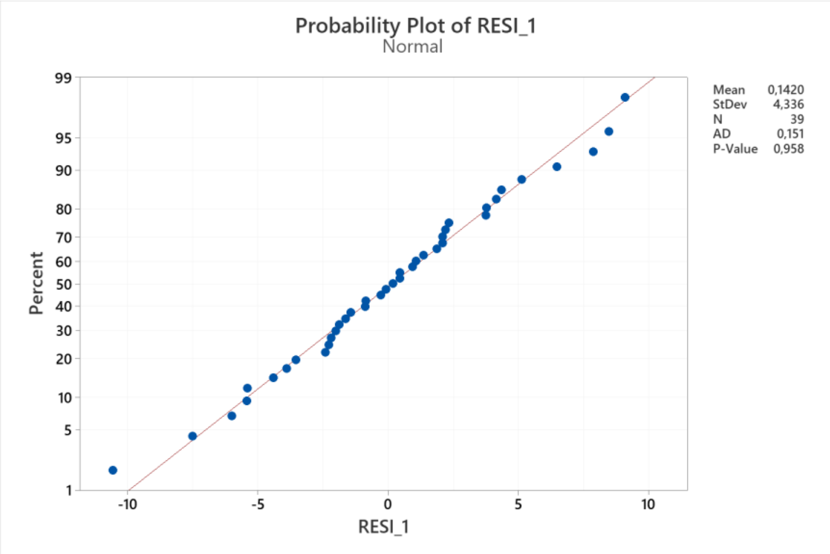
Model Summary

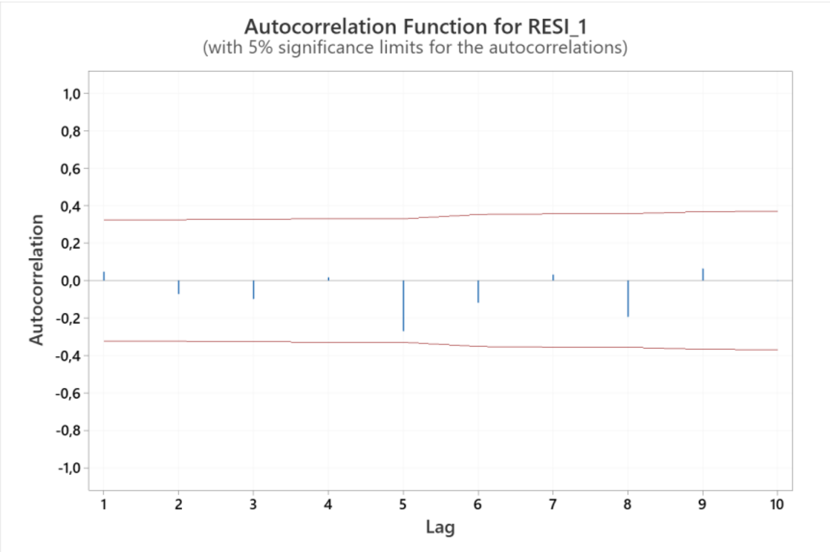
S	R-sq	R-sq(adj)	R-sq(pred)
4,39697	99,87%	99,87%	99,86%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	564492	282246	14598,89	0,000
exp2	1	135	135	7,00	0,012
AR1	1	295	295	15,26	0,000
Error	37	715	19		
Total	39	565207			

The residuals of this model are normal and independent, thus the model is adequate.





Test

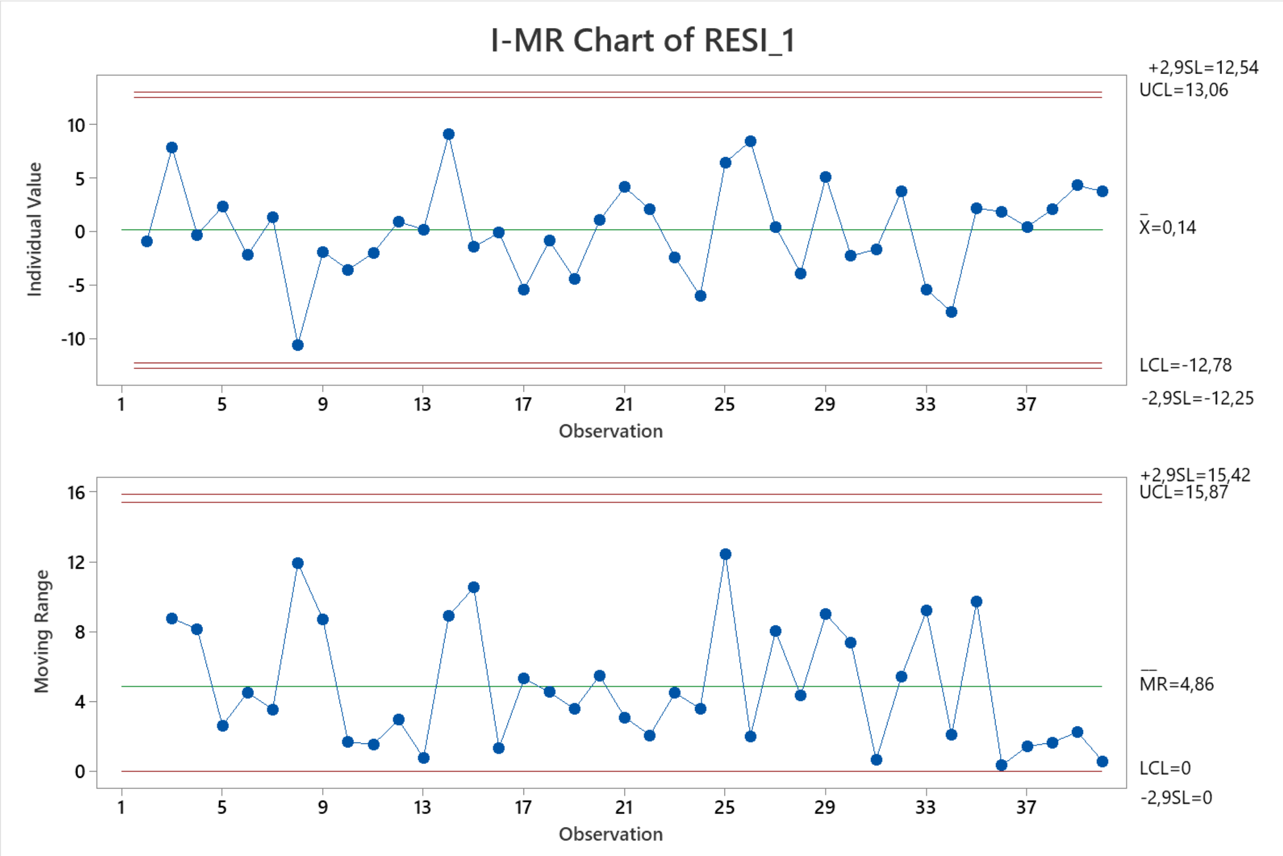
Null hypothesis H_0: The order of the data is random

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

Number of Runs		
Observed	Expected	P-Value
18	20,49	0,419

b) Given $ARL_0 = 250$, the Type I error is $\alpha = 0.004$, thus $K = z_{\alpha/2} = 2.878$.

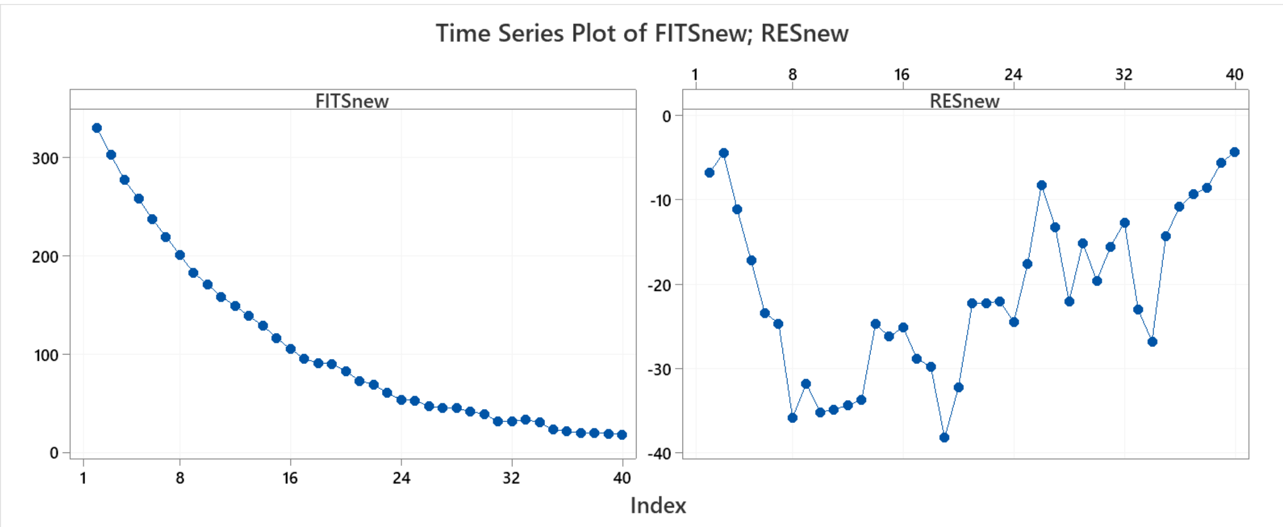
The special cause control chart for the process is the following:



The process is in-control.

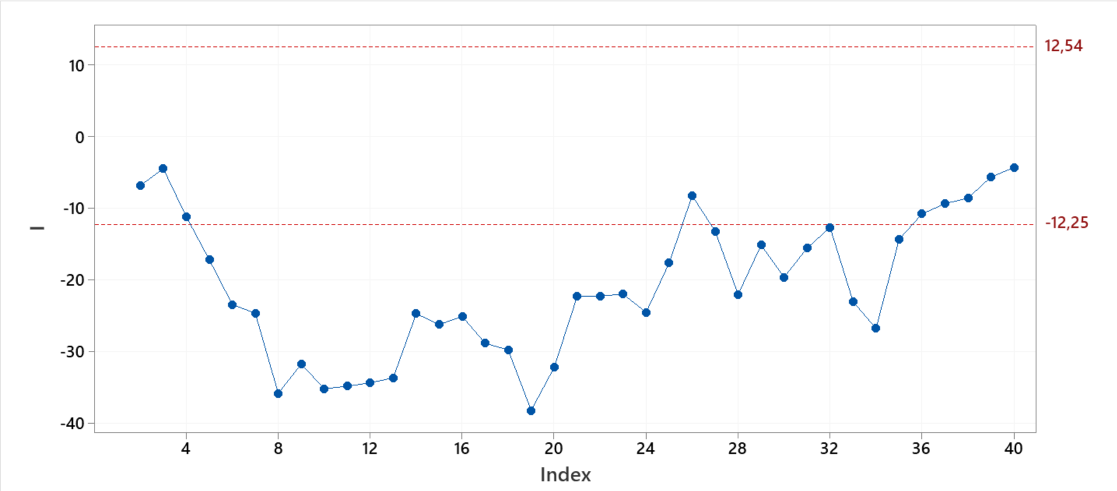
c) The same model shall be applied to the new data.

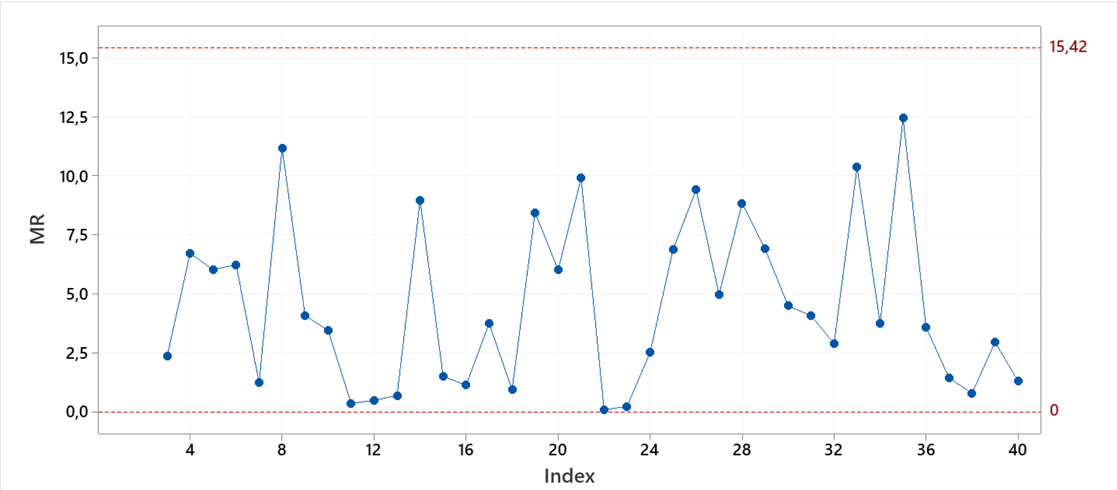
The resulting fits and residuals are the following:



FITSnew	RESnew
329,4967	-6,79671
302,542	-4,44202
276,7573	-11,1573
257,8845	-17,1845
237,1236	-23,4236
219,1626	-24,6626
200,6433	-35,8433
182,5652	-31,7652
170,714	-35,214
158,155	-34,855
149,0751	-34,3751
138,4841	-33,6841
129,0216	-24,7216
116,0284	-26,2284
105,1061	-25,1061
95,25436	-28,8544
90,68526	-29,7853
90,13623	-38,2362
82,52021	-32,2202
72,80882	-22,3088
68,92934	-22,2293
60,49925	-21,9992
53,51629	-24,5163
52,63731	-17,6373
47,21607	-8,21607
45,18783	-13,1878
44,93043	-22,0304
41,22486	-15,1249
39,2172	-19,6172
31,54226	-15,5423
31,45648	-12,6565
33,52936	-23,0294
30,1831	-26,7831
23,13055	-14,3305
21,65678	-10,7568
19,52807	-9,32807
19,64011	-8,54011
19,08937	-5,58937
18,19471	-4,29471

Using the previously designed control chart, the new cooling pattern results to be out-of-control.





Indeed, as shown in the following figure, the new cooling process is characterized by a slower decay than the previous (in-control) one.

d) A suitable way to check whether the first and second components (referring to Table 1 and Table 2, respectively) have a cooling time series that is statistically different is to fit the same model to the two time series and check if process parameters are statistically different.

Time series 1

Regression Equation

x = 157,8 exp2 + 0,537 AR1

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
exp2	157,8	59,7	2,65	0,012	680,40
AR1	0,537	0,137	3,91	0,000	680,40

Time series 2

Regression Equation

xnew = 1,2 exp2 + 0,9294 ARnew

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
exp2	1,2	13,8	0,09	0,931	31,22
ARnew	0,9294	0,0258	36,07	0,000	31,22

Both models have normal and independent residuals. The results of two 2-sample t tests on the model coefficients (with different variances) with a familywise confidence of 95% are:

Test on β_1:

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	40	157,8	59,7	9,4
Sample 2	40	1,2	13,8	2,2

Estimation for Difference

Difference	97,5% CI for Difference
156,60	(134,10; 179,10)

Test

Null hypothesis H₀: μ₁ - μ₂ = 0
Alternative hypothesis H₁: μ₁ - μ₂ ≠ 0

T-Value	DF	P-Value
16,16	43	0,000

Test on β_2:

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	40	0,537	0,137	0,022
Sample 2	40	0,9294	0,0258	0,0041

Estimation for Difference

Difference	97,5% CI for Difference
-0,3924	(-0,4437; -0,3411)

Test

Null hypothesis H₀: μ₁ - μ₂ = 0
Alternative hypothesis H₁: μ₁ - μ₂ ≠ 0

T-Value	DF	P-Value
-17,80	41	0,000

The two cooling histories are statistically significant.

Exercise 2 (solution)

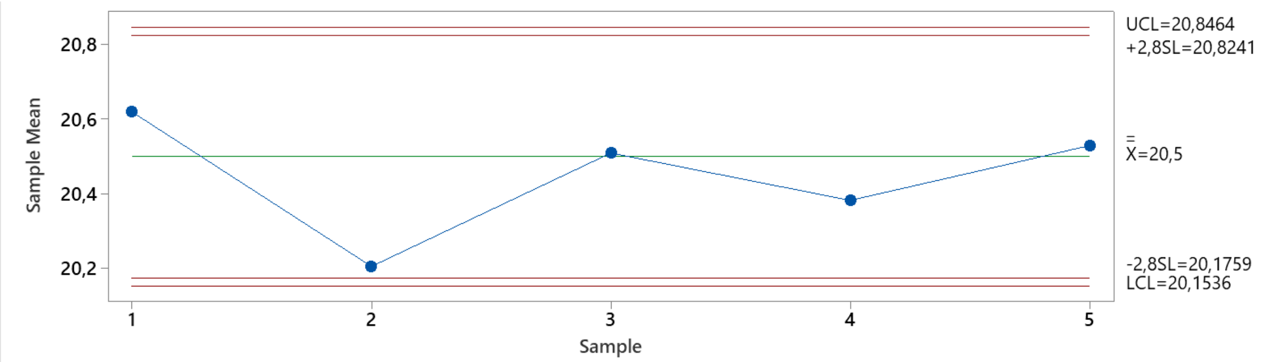
a) The two quality characteristics are independent. Therefore, the appropriate familywise correction is $\alpha^* = 1-(1-\alpha)^{1/2} = 0{,}005013$.

The control charts for the mean with $K = z_{\alpha^*/2} = 2.807$ have the following control limits:

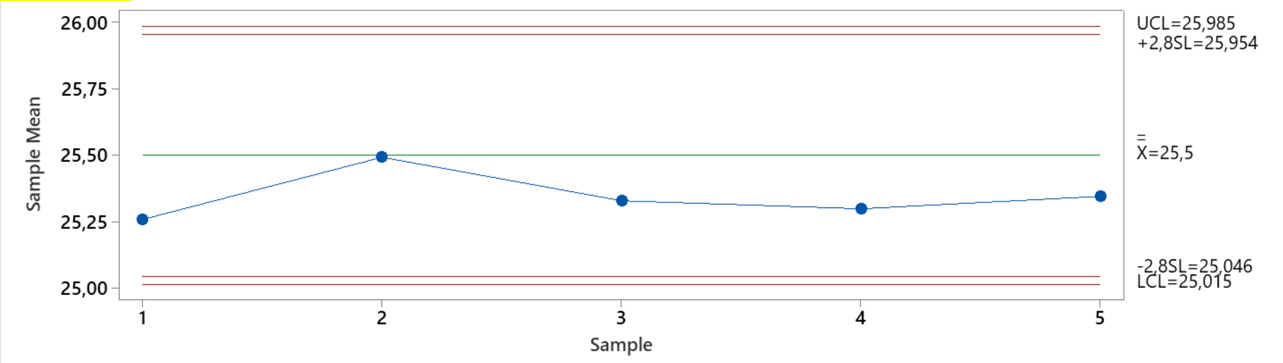
Diameter 1	Diameter 2
Xbar	Xbar
LCL =20.1536, UCL =20.8241	LCL =25.015, UCL =25.954

By applying these control charts to the provided data, no alarm is signalled, but all sample mean values of diameter 2 are below the center line, which may possibly indicate a small shift in the mean has occurred.

Diameter 1



Diameter 2



b) If no familywise correction was applied, $\alpha^* = 0.01$, and hence $\alpha = 1-(1-\alpha^*)^2 = 0{,}0199$. The corresponding Average Run Length is $ARL_0 = 50{,}25$. The expected one (with familywise correction) was $ARL_0 = 100$. Failing in using the proper correction would result is a much lower ARL.

c) The probability of detecting the shift is $P = 1 - \beta_{\bar{x},1} \cdot \beta_{\bar{x},2}$.

Let $\Delta\mu_1 = \Delta\mu_2 = 0.3\ mm$, then:

$$\beta_{\bar{x},1} = \phi\left(\frac{UCL_1 - \mu_1 - \Delta\mu_1}{\sigma_1/\sqrt{n}}\right) - \phi\left(\frac{LCL_1 - \mu_1 - \Delta\mu_1}{\sigma_1/\sqrt{n}}\right) =$$

$$\phi\left(\frac{20.8241-20.5-0.3}{0.2/\sqrt{3}}\right)-\phi\left(\frac{20.1536-20.5-0.3}{0.2/\sqrt{3}}\right)=0.5266$$

$$\beta_{\bar{x},2}=\phi\left(\frac{UCL_2-\mu_2-\Delta\mu_2}{\sigma_2/\sqrt{n}}\right)-\phi\left(\frac{LCL_2-\mu_2-\Delta\mu_2}{\sigma_2/\sqrt{n}}\right)=$$
$$\phi\left(\frac{25.954-25.5-0.3}{0.28/\sqrt{3}}\right)-\phi\left(\frac{25.015-25.5-0.3}{0.28/\sqrt{3}}\right)=0.8296$$

The resulting power is P =0.5166.

d) It is possible to express the power $P(n)=1-\beta_{\bar{x},1}(n)\cdot\beta_{\bar{x},2}(n)$ as a function of the sample size, keeping in mind that also control limits are functions of the sample size n.

By increasing the sample size, we get:

n	UCL1	LCL1	UCL2	LCL2	beta1	beta2	P
3	20,82412444	20,17588	25,95377	25,04623	0,582746	0,829254948	0,516755
4	20,7807	20,2193	25,89298	25,10702	0,423479	0,746700187	0,683788
5	20,75106571	20,24893	25,85149	25,14851	0,292154	0,65954168	0,807312
6	20,72919059	20,27081	25,82087	25,17913	0,192907	0,572423119	0,889576
7	20,71218926	20,28781	25,79706	25,20294	0,122694	0,488937325	0,940011
8	20,69848487	20,30152	25,77788	25,22212	0,075552	0,411589674	0,968903
9	20,68713333	20,31287	25,76199	25,23801	0,045228	0,341899076	0,984537
10	20,67753027	20,32247	25,74854	25,25146	0,026408	0,280568119	0,992591

The minimum sample size to have a power P>90% is $n=7$.

<u>Exercise 3 (solution)</u>

Given a stationary AR(1) model $X_t=\xi+\phi_1 X_{t-1}+\varepsilon_t,\ \varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, it is known that:

$$\mu=\frac{\xi}{1-\phi_1}$$
$$\sigma^2=\frac{\sigma_\varepsilon^2}{1-\phi_1{}^2}$$

Therefore:

$$1-\phi_1=\frac{\xi}{\mu}$$
$$1-\phi_1{}^2=\frac{\sigma_\varepsilon^2}{\sigma^2}$$

By solving the two equations with two unknowns:

$$\phi_1=\sqrt{1-\frac{\sigma_\varepsilon^2}{\sigma^2}}$$

$$\xi = \mu\left(1 - \sqrt{1 - \frac{\sigma_{\varepsilon}^{2}}{\sigma^{2}}}\right)$$