QUALITY DATA ANALYSIS

29/01/2021

General recommendations:

- write the solutions in a CLEAR and READABLE way on paper and show (qualitatively) all the relevant plots;
- avoid (if not required) theoretical introductions or explanations covered during the course;
- always state the assumptions and report all relevant steps/discussion/formulas/expression to present and motivate your solution;
- when using hypothesis tests provide the numerical value of the test statistic and the test conclusion in terms of p-value.
- Exam duration: 2h 10min

Exercise 1 (15 points)

A service company that produces metal parts via Additive Manufacturing processes is keeping under control the pre-heating temperature applied in a laser powder bed fusion process on aluminium parts. Every time a build is produced, the temperature in Celsius degrees is recorded and the data are included in Table 1.

Table 1

build	T (°C)	build	T (°C)	build	T (°C)	build	T (°C)
1	111,0	11	111,1	21	112,7	31	112,3
2	110,8	12	111,0	22	111,1	32	111,0
3	111,1	13	111,0	23	113,2	33	112,1
4	110,3	14	112,3	24	113,8	34	113,0
5	111,0	15	112,4	25	113,3	35	113,4
6	109,9	16	112,2	26	113,8	36	112,7
7	109,0	17	111,9	27	111,6	37	113,9
8	110,8	18	111,9	28	113,8	38	114,0
9	110,0	19	113,6	29	111,4	39	114,9
10	109,9	20	111,4	30	109,0	40	115,3

a) Fit a suitable model to these data;
b) Estimate the confidence interval for the model parameters with a confidence of 95% (show the formula for the interval too)
c) Based on the result of point a), estimate the 95% prediction interval for the expected pre-heating temperature in the next build.
d) Based on the result of point a), design an appropriate control chart for these data with ARL0 = 250. If needed, use the following information: during the production of the 30th build a partial detachment of the thermocouple used to measure the temperature was observed.

Exercise 2 (15 points)

In order to monitor a process for the production of special valves for the oil & gas sector, two quality characteristics kept under statistical control through periodic individual measurements. The two quality characteristics are known to follow a multinormal distribution with $\mu_1 = 70{,}05$ mm, $\mu_2 = 89{,}89$ mm and variance-covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} 0{,}8 & 0{,}65 \\ 0{,}65 & 0{,}9 \end{bmatrix}$$

Hint: to import a matrix in Minitab use the command Data→Copy→Columns to Matrix, selecting the columns of the worksheet that include the values of the matrix.

a) Design two univariate I control charts for the two distinct quality characteristics (with a familywise type I error $\alpha = 0{,}01$) and check if the measurements in Table 2 are in-control or not.
b) Estimate the principal components of the quality characteristics: show the weights of the linear combination and the percentage of variance explained by each of them.
c) Design two univariate I control charts for the two principal components estimated in point b) (with a familywise type I error $\alpha = 0{,}01$).
d) By using the control charts designed in point c) verify if the measurements of the quality characteristics reported in Table 2 are in-control or not. Does the result differ from the one obtained in point a)? Discuss the results.

Table 2

X1	X2
70,53	89,46
71,69	91,40
69,00	91,48
70,82	89,48
70,34	91,56
68,88	87,81
69,66	88,50
70,36	90,05
72,25	92,20
72,15	92,30

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

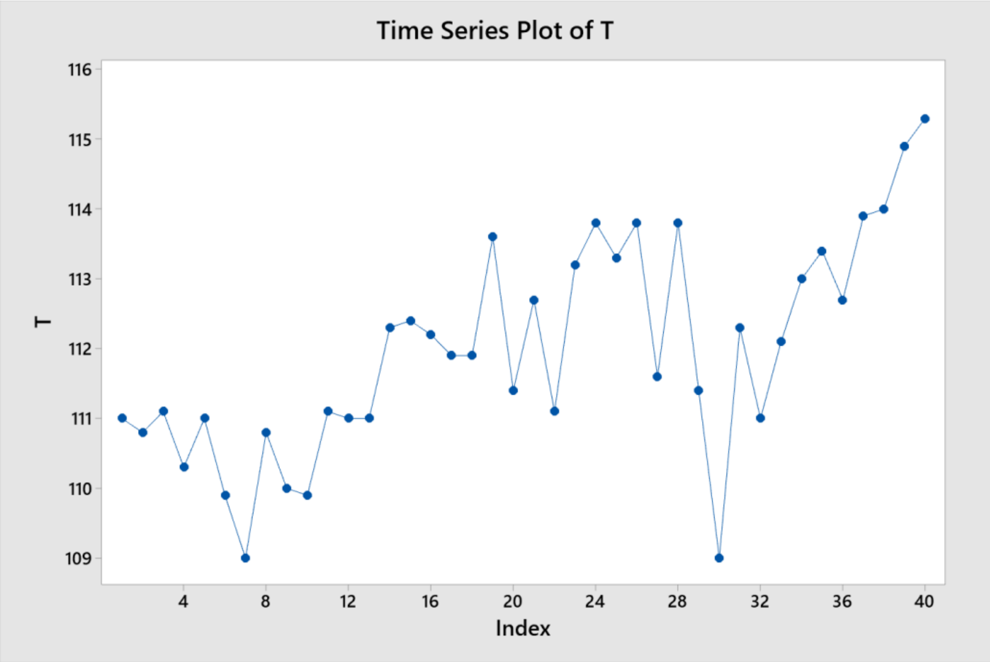
Referring to the application described in Exercise 1, it is known that a pre-heating temperature lower that 90 °C leads to parts whose quality is not conforming with respect to the product specifications. Assuming that the pre-heating temperature follows a normal distribution with mean μ and standard deviation $\sigma = 1{,}53$ °C, estimate the target mean temperature, μ, such that the probability of having non-conforming builds is 1%.

SOLUTION

Exercise 1

a)

Time series plot of the temperature series:



It is present a meandering pattern.

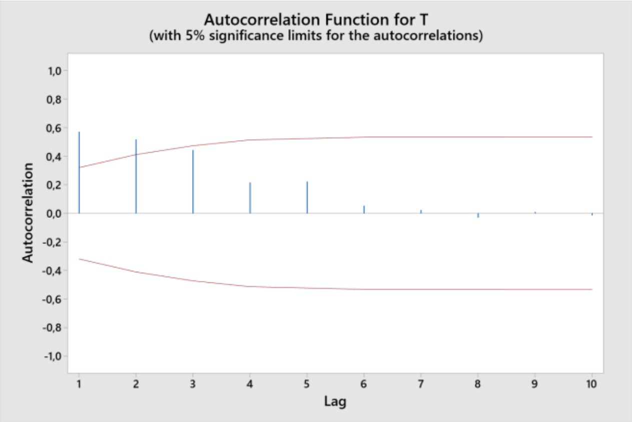
Runs test: null hypothesis is not accepted at 95% confidence:

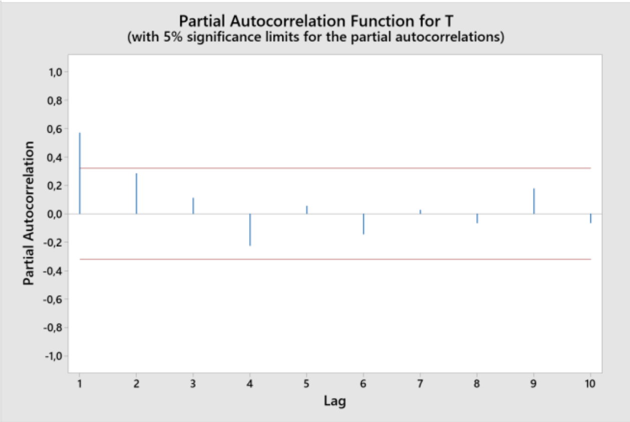
Test

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

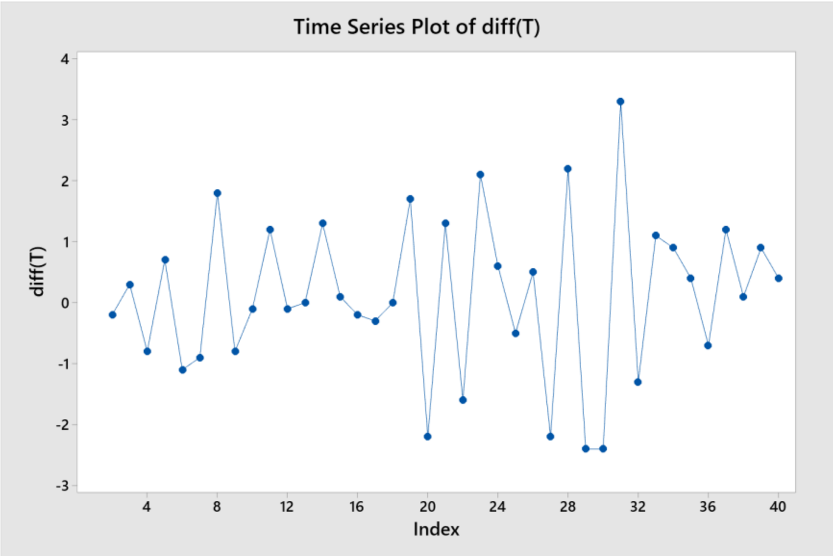
Number of Runs		
Observed	Expected	P-Value
14	20,95	0,026

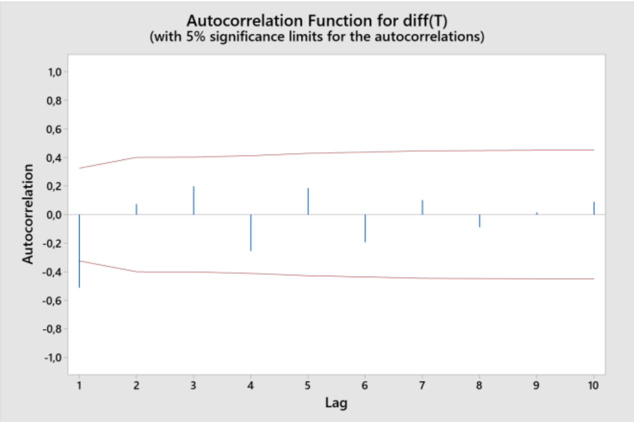
Sample autocorrelation and partial autocorrelation functions:

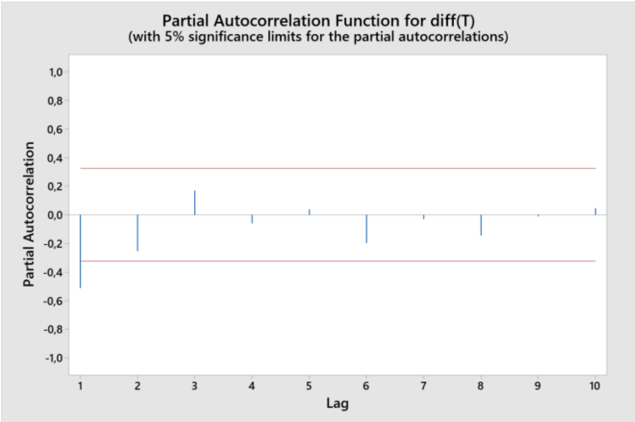




A slow decay of the SACF is present, which suggests a non-stationariety of the process. By differencing the timeseries we get:







A suitable model for the temperature time series is therefore an ARIMA(1,1,0):

WORKSHEET 1

Regression Analysis: diff(T) versus AR1diff

Method

Rows unused 2

Regression Equation

diff(T) = -0,502 AR1diff

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
AR1diff	-0,502	0,142	-3,52	0,001	1,00

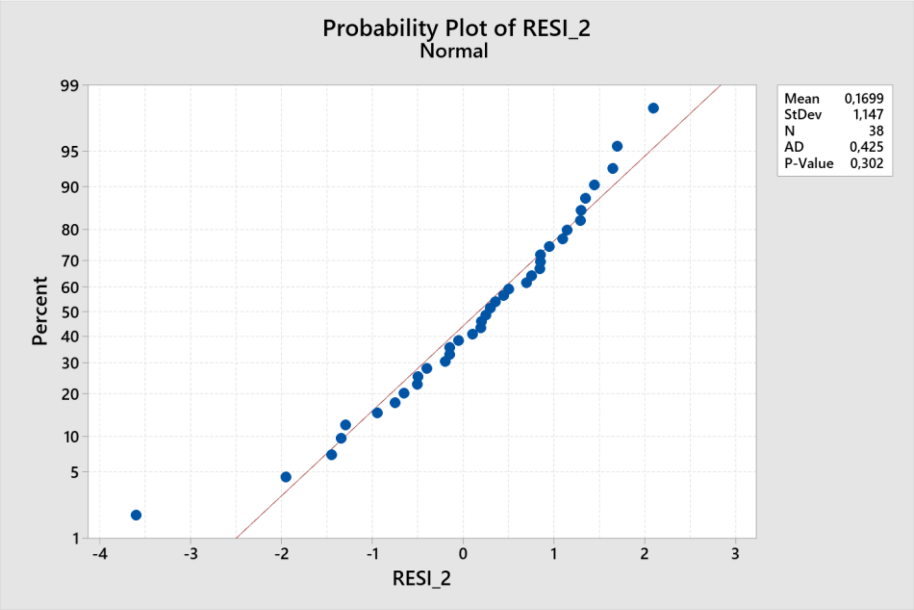
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,16029	25,13%	23,10%	17,43%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	16,718	16,7177	12,42	0,001
AR1diff	1	16,718	16,7177	12,42	0,001
Error	37	49,812	1,3463		
Lack-of-Fit	31	47,067	1,5183	3,32	0,069
Pure Error	6	2,745	0,4575		
Total	38	66,530			

Check of residuals:

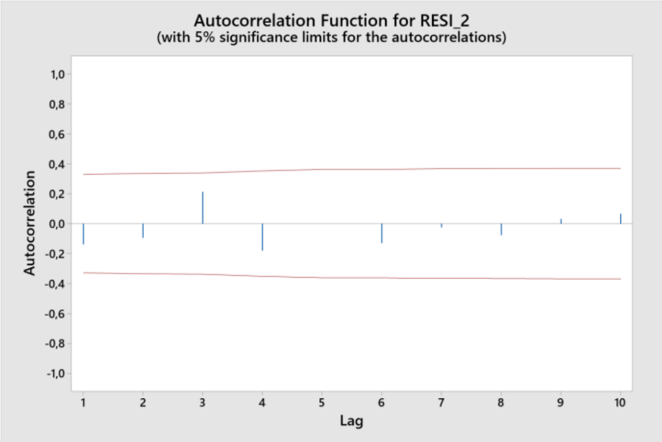


Test

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
23	19,53	0,241



The model is appropriate.

Other models deemed appropriate for these data include an AR(1) model including the constant term and possibly a trend model.

b)

95% Confidence Interval for the AR1 model term:

$$\hat{\beta}_1 - t_{\frac{\alpha}{2},n-p} se(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\frac{\alpha}{2},n-p} se(\hat{\beta}_1)$$

Where $\alpha = 0{,}05$, $p = 1$ and $n = 38$

The resulting interval is:

$$-0.790 \leq \beta_1 \leq -0{,}213$$

c)

The 95% prediction interval for the differenced time series for build 41 is the following:

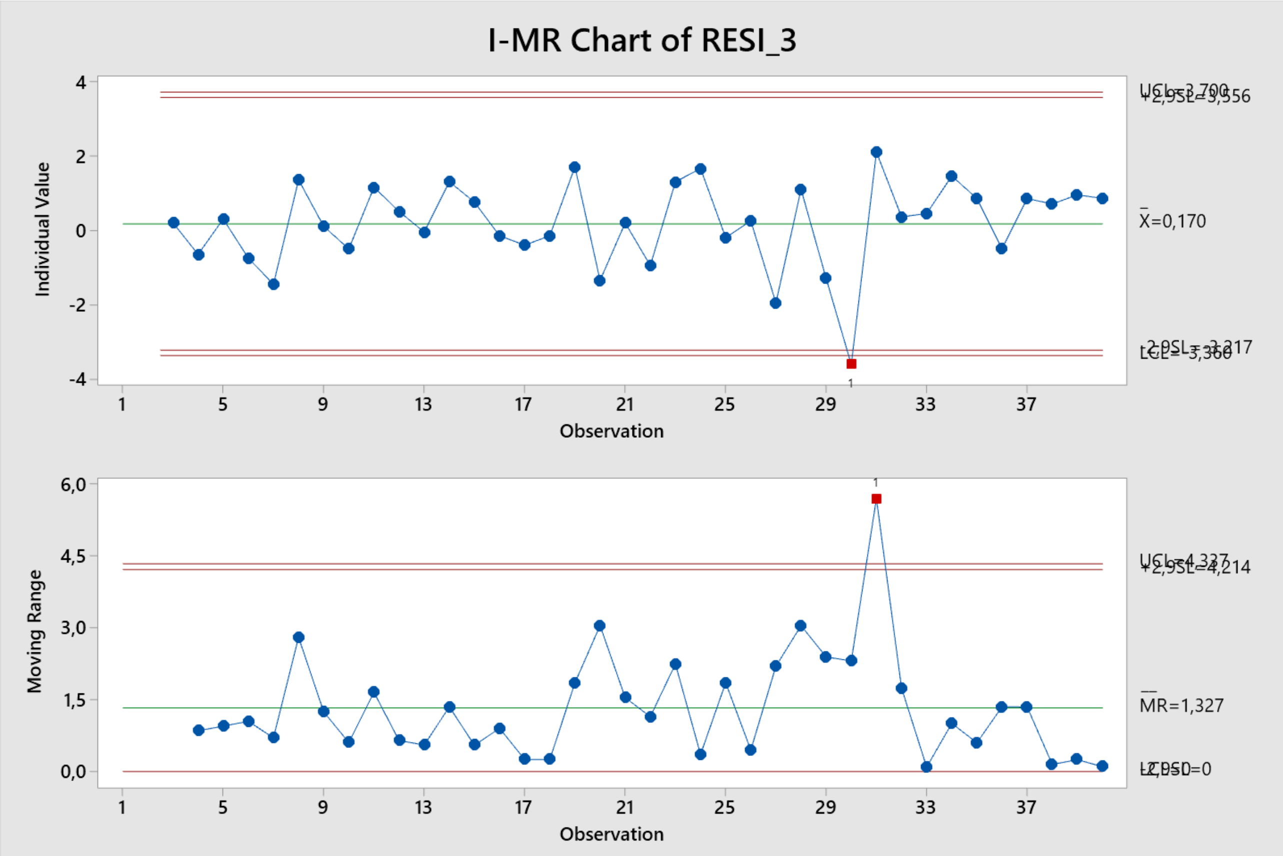
95% PI
(-2,81683; 1,91371)

Knowing that for the 40th build the temperature was 115,3 °C, the 95% prediction interval for the temperature value is:

$$113{,}39\ °C \leq T \leq 118{,}12\ °C$$

d)

The Type I error corresponding to ARL0 = 250 is $\alpha = 0{,}004$, which corresponds to $k = z_{\alpha/2} = 2{,}878$. The resulting I-MR control chart for the model residuals is the following:



An alarm is signalled for build number 30. For that observation, an assignable cause is available (partial detachment of the thermocouple used to measure the temperature). Thus, it is possible to introduce a dummy variable in the model that is equal to 1 for that observation and 0 for all other observations.

WORKSHEET 1

Regression Analysis: diff(T) versus AR1diff; dummy

Method

Rows unused 2

Regression Equation

diff(T) = -0,644 AR1diff - 3,95 dummy

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
AR1diff	-0,644	0,128	(-0,903; -0,385)	-5,05	0,000	1,09
dummy	-3,95	1,04	(-6,06; -1,84)	-3,79	0,001	1,09

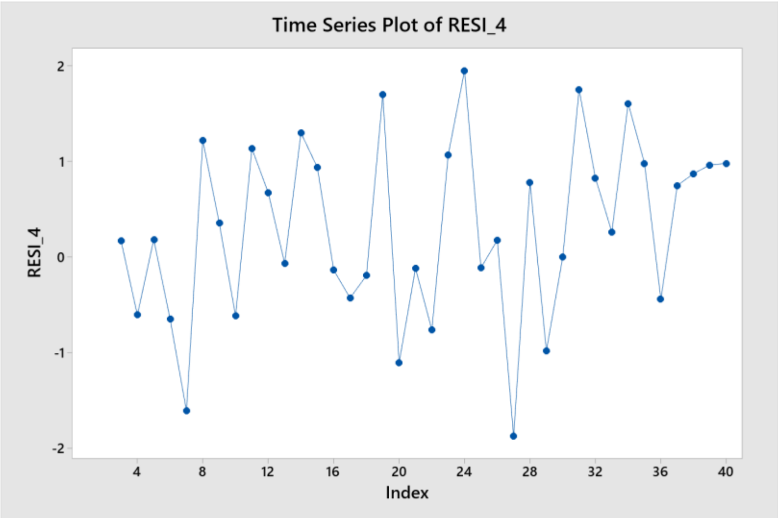
Model Summary

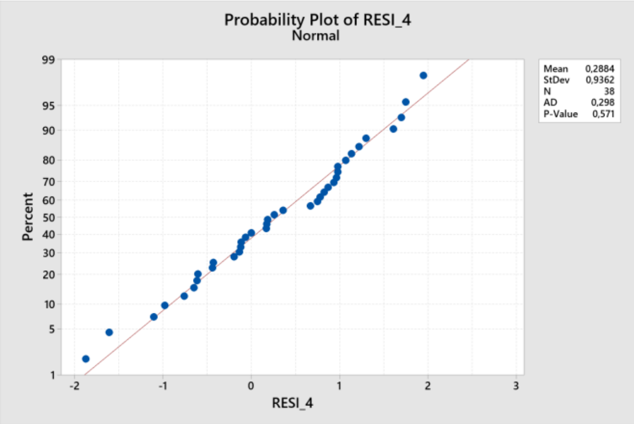
S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0,994271	46,51%	43,54%	*	*	112,05	116,26

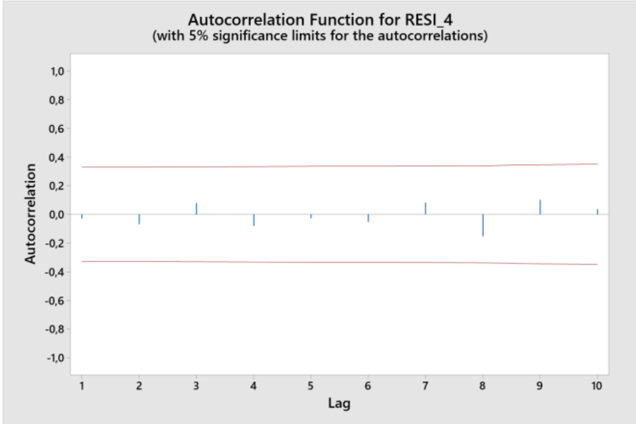
Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	2	30,941	46,51%	30,941	15,4707	15,65	0,000
AR1diff	1	16,718	25,13%	25,181	25,1813	25,47	0,000
dummy	1	14,224	21,38%	14,224	14,2236	14,39	0,001
Error	36	35,589	53,49%	35,589	0,9886		
Lack-of-Fit	30	32,844	49,37%	32,844	1,0948	2,39	0,139
Pure Error	6	2,745	4,13%	2,745	0,4575		
Total	38	66,530	100,00%				

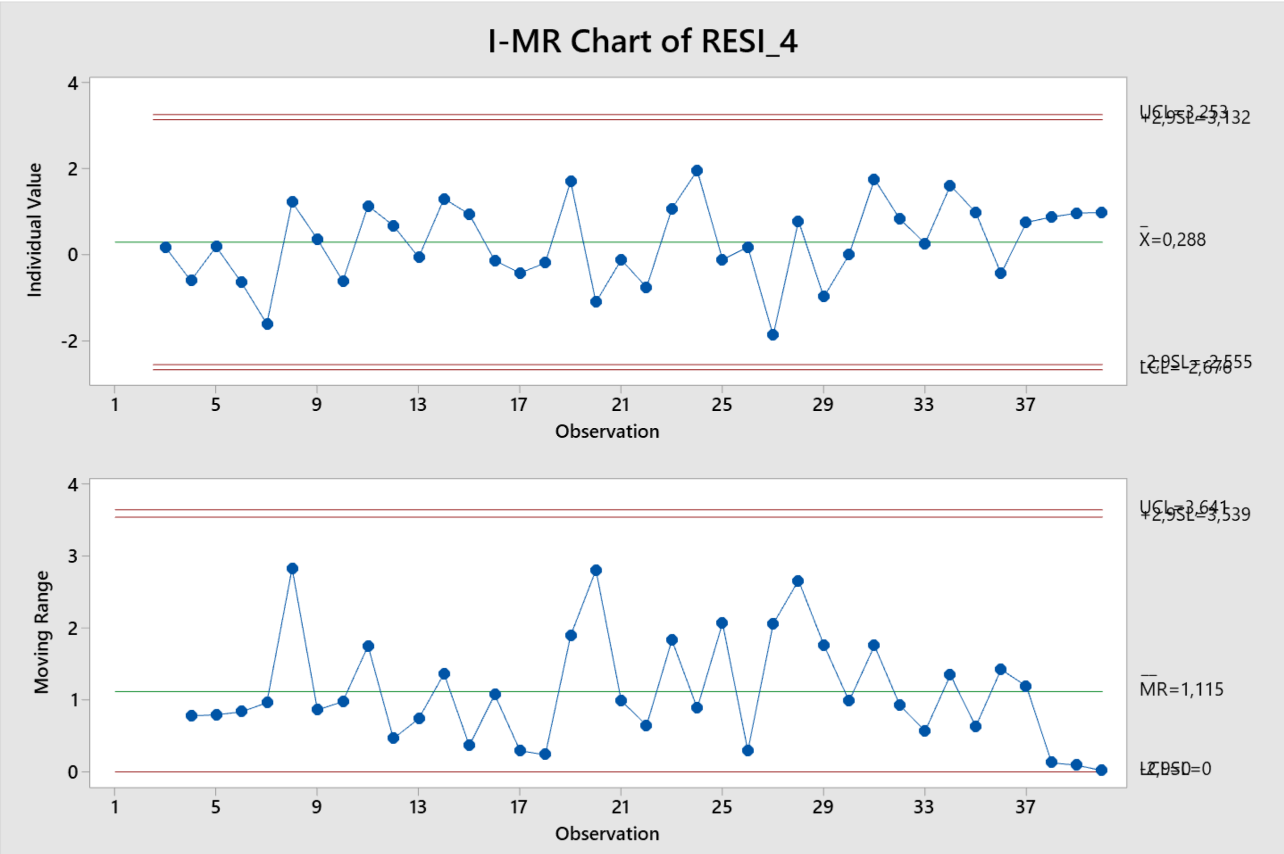
Residuals:







The resulting control chart is the following:



No other violation of the control limits is present. The design of the control chart is over.

Exercise 2

a)

By using the Bonferroni's correction for a target familywise type I error $\alpha = 0{,}01$, each control chart can be designed using $k = z_{\alpha'/2} = 2{,}807$ where $\alpha' = \alpha/2$.

Being known the distribution of the monitored quantities, with $\mu_1 = 70{,}05$ mm, $\mu_2 = 89{,}89$ mm and $\sigma_1 = 0{,}89$ mm, $\sigma_2 = 0{,}95$ mm, the two control charts can be designed as follows:

Control chart for X1:

$$UCL = \mu_1 + k\sigma_1 = 72{,}54\ mm$$

$$CL = \mu_1 = 70{,}05\ mm$$

$$LCL = \mu_1 - k\sigma_1 = 67{,}55\ mm$$

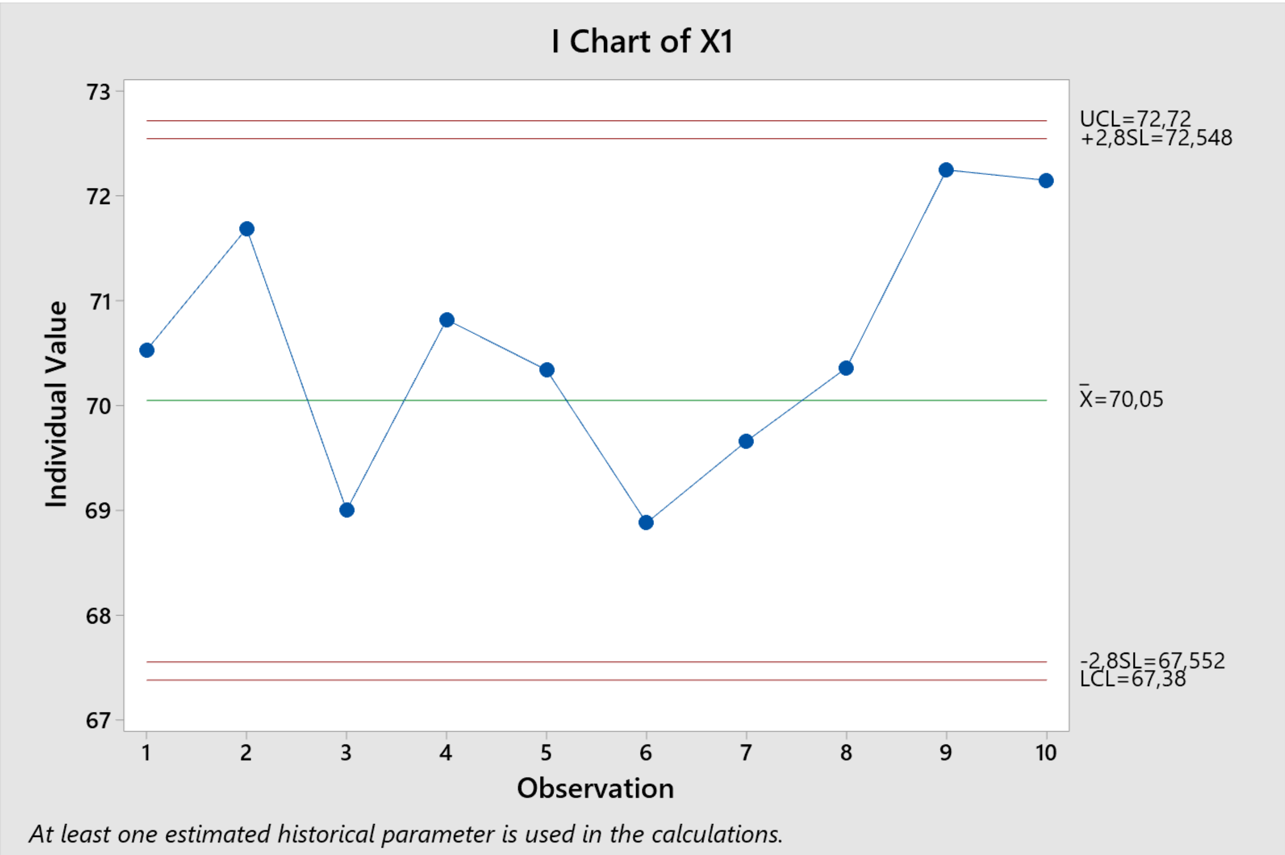
Control chart for X2:

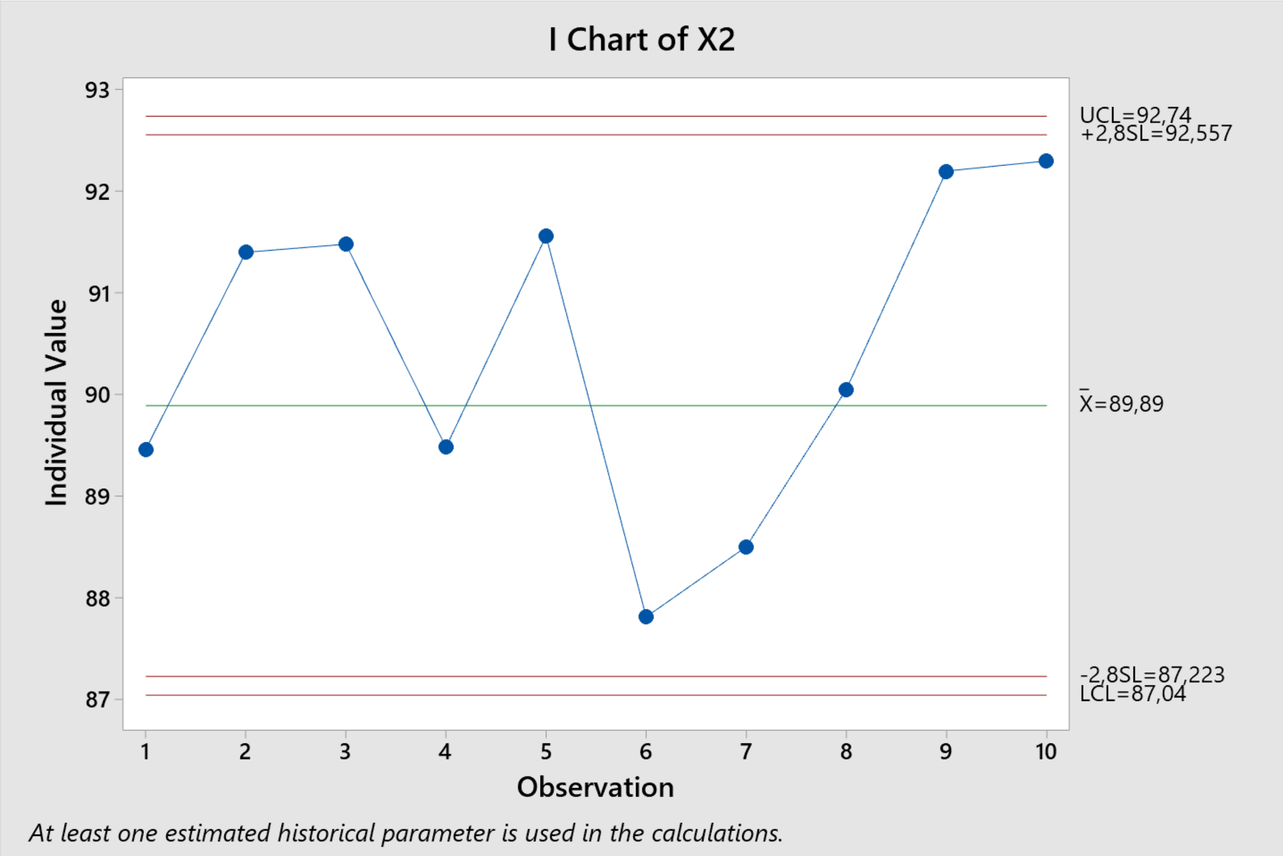
$$UCL = \mu_2 + k\sigma_2 = 92{,}56\ mm$$

$$CL = \mu_2 = 89{,}89\ mm$$

$$LCL = \mu_2 - k\sigma_2 = 87{,}22\ mm$$

The Phase II control chart for data in Table 2 is the following:





No alarm is signalled, the process is judged in-control.

b)

The Principal Components can be computed though the eigendecomposition of the variance covariance matrix:

$$z_1 = 0{,}6795X_1 + 0{,}733722X_2$$

$$z_2 = 0{,}733722X_1 - 0{,}6795X_2$$

The loadings are $\lambda_1 = 1{,}50192$, $\lambda_2 = 0{,}19808$, thus the first principal component explains about 88,3% of the overall variance and the second the remaining 12,7%.

c)

The scores of the first two PCs for the data in Table 2 are:

X1	X2	z_1	z_2
70,53	89,46	113,564	-9,0387
71,69	91,40	115,776	-9,5058
69,00	91,48	114,006	-11,5338
70,82	89,48	113,776	-8,8395

70,34	91,56	114,976	-10,6050
68,88	87,81	111,232	-9,1281
69,66	88,50	112,268	-9,0247
70,36	90,05	113,881	-9,5643
72,25	92,20	116,743	-9,6385
72,15	92,30	116,748	-9,7798

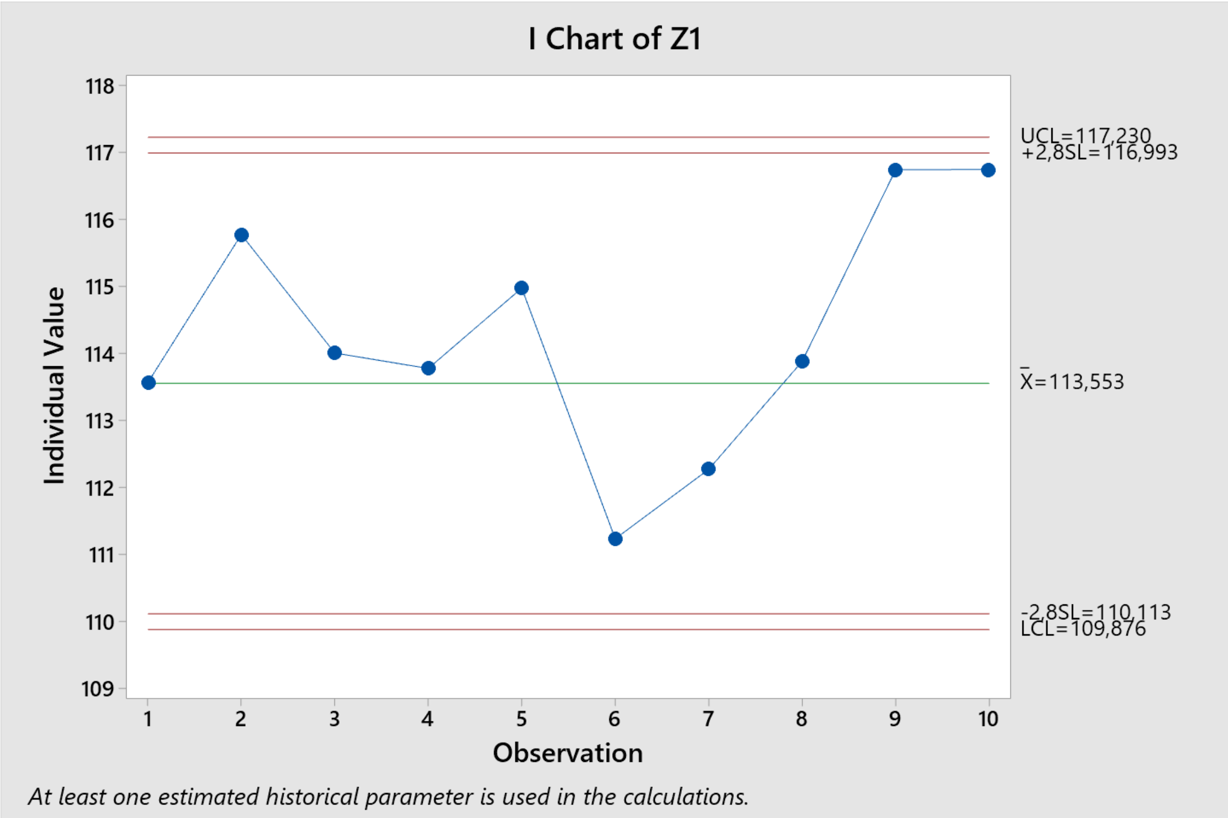
The resulting control charts can be designed with the following parameters:

$k = z_{\alpha'/2} = 2{,}807$,

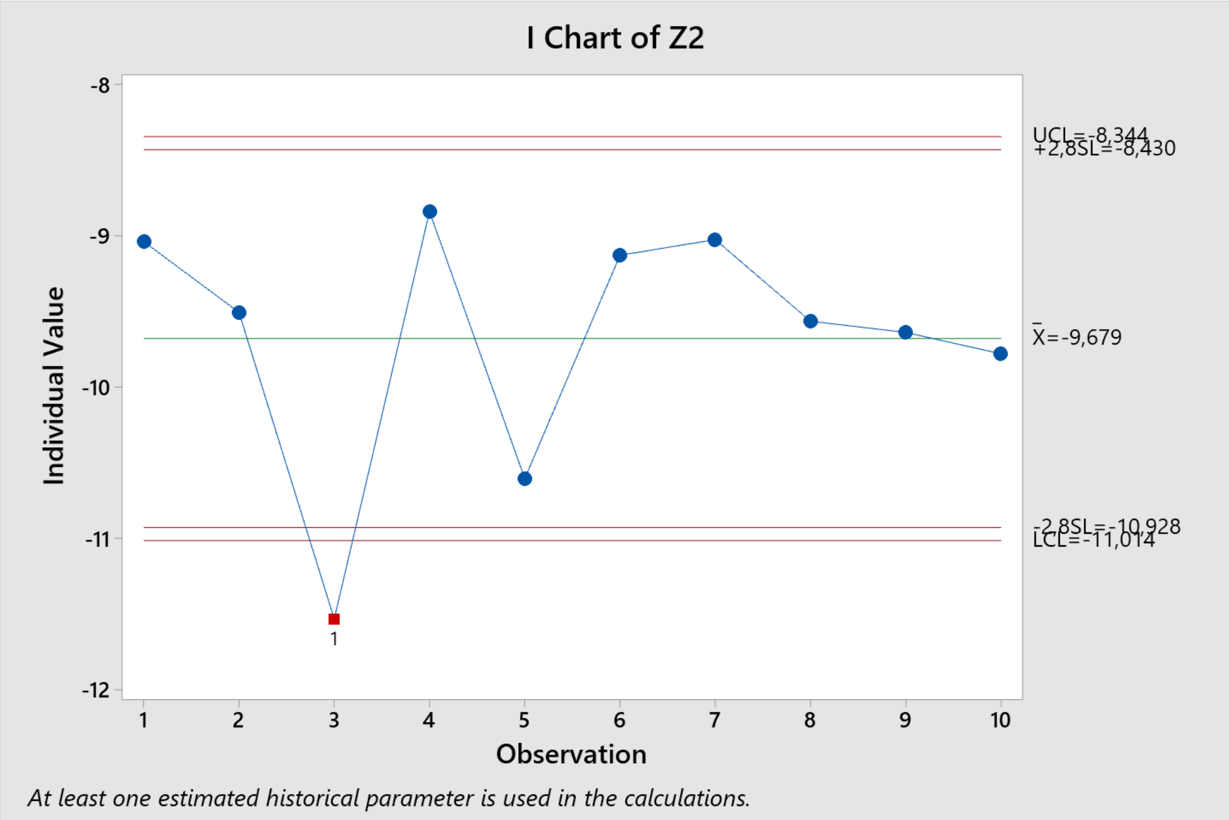
$\mu_1 = 0{,}6795 \cdot 70{,}05 + 0{,}733722 \cdot 89{,}89 = 113{,}553$ and $\sigma_1 = \sqrt{\lambda_1} = 1{,}2255$

$\mu_1 = 0{,}733722 \cdot 70{,}05 - 0{,}6795 \cdot 89{,}89 = -9{,}679$ and $\sigma_2 = \sqrt{\lambda_2} = 0{,}4451$

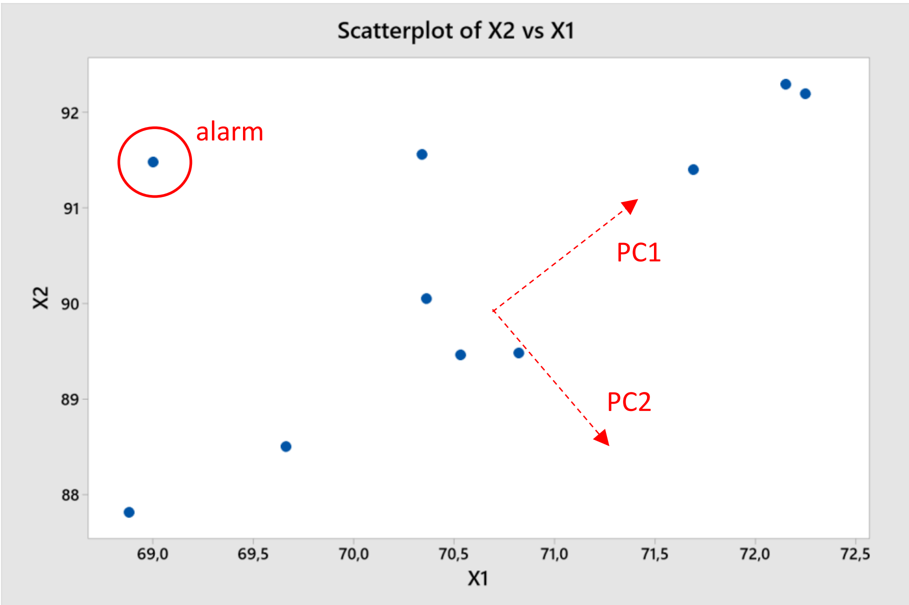
Control chart for the first PC:



Control chart for the second PC:



The control chart on the second PC signals an alarm for the third observation. The nature of this alarm can be highlighted by looking at the scatter plot of the data:



The third observation results to be an outlying one when projected along the second PC. By monitoring the PCs the correlation between the variables is properly taken into account, which allows signalling an alarm for the third observation, which can not be captured by the control charts on the original variables as they neglect their correlation structure. The same observation could be signalled also by using an Hotelling's T^2 control chart, which exploits an elliptical control region.

Exercise 3

If the one-sided specification limit LSL = 90 °C corresponds to a non-conforming probability $\gamma = 0{,}01$, it is possible to estimate the mean temperature μ as follows:

$$z_\gamma = \frac{LSL - \mu}{\sigma} = \frac{90 - \mu}{1{,}53} = -2{,}326$$

Which results into: $\mu = 93{,}56\,°C$.