OUTLIERS/OOE AND OOT DATA IN PHARMACEUTICAL MANUFACTURING AND CONTROL

INTRODUCTION

- From a very general standpoint the term ***anomalous result*** identifies any data points or observations that deviate significantly from what is considered normal or typical within a specific context.

- ***anomalous results*** include a wide range of unexpected outcomes and, in particular:
  - ***out-of-specifications*** (OOS)

  and the so-called:
  - ***unexpected results***

INTRODUCTION

In the technical lexicon, ***unexpected results*** are often identified using a wide variety of terms such as:

- *aberrant*
- *abnormal*
- *atypical*
- *irregular*
- *deviant*
- *extreme value*
- *outlier*
- *out of trend*
- *etc.*

INTRODUCTION

- Among the above terms: *aberrant*, *abnormal*, *atypical*, *irregular, extreme value* and *deviant*, are more general and can be considered somewhat colloquial in nature.

- They express the idea of something being different from what is considered *« normal »* or *« expected »*, but they do not have precise and specific definitions in statistical or technical contexts.

- *In contrast,* ***outlier*** *and* ***out of trend*** *have more precise and technical definitions and are often used in statistical and technical discussions with well-defined meanings.*

INTRODUCTION

- **Outlier**: An outlier is a data point that significantly deviates from the rest of the data and is typically defined based on statistical criteria, such as being outside a certain range or a certain number of standard deviations from the mean. Outliers are often subject to specific statistical analysis and may be treated differently in data analysis.

- **Out of Trend (OOT)** : Out of trend typically refers to a situation where data points no longer follow an established trend or pattern. It can be used to describe data that is not behaving as expected over time, and it may signal the need for further investigation into the underlying causes.

INTRODUCTION

Practically coinciding with the definition of outlier is also that of:

- **Out of Expectation (OOE)** : this is an unexpected result in a series of results obtained in a short period of time. An OOE is a result that, although within specifications, lies outside the variability expected for a given analytical procedure.

INTRODUCTION

Sometimes, in the technical literature, we also find other terms such as:

- **OOC** (Out of Control)
- **OOSC** (Out of Statistical Control)
- *etc.*

but these are nevertheless situations attributable to those described above.

INTRODUCTION

In summary, from here on we will therefore focus our attention:

- **Outliers / OOE** (Out of Expectations)
- **OOT** (Out of Trend)

results as they represent (<u>fairly</u>) clearly definable situations.

OUTLIERS / OOE

OUTLIERS

- The problem of *outliers* is one of the oldest and most widely studied by the statistics community.
- In more recent times, with the advent of large amounts of data, advanced software technology, *etc.*, the role of computer scientists in this field has been increasing.
- However, since in our field we have **small data**, **measured very precisely** and **without missing data**, we will approach the problem from a plain and classical statistical perspective.

OUTLIERS

What is an outlier?

- *An outlier is a data point that is significantly different from the remaining data, the main body of data.*

- Barnett and Lewis state that:

  "An outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data."

V. Barnett, T. Lewis, *Outliers in Statistical Data*, 2^{nd} Ed., J. Wiley and Sons (1984)

OUTLIERS

What is an outlier?

- Hawkins defined an outlier as follows:
  "An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism."

  D. Hawkins, *Identification of Outliers*, Chapman and Hall (1980)

- Alternatively, Beckman and Cook:
  "A *discordant observation* is any observation that appears surprising or discrepant to the investigator."

  R.J. Beckman, R.D. Cook, *Outlier……….s*, Technometrics, 25 (1983) 119-149

OUTLIERS

- **Why Study Outliers?**

Because they can:

- provide useful information about the data and, for example, improve the quality of the data collection process
- arise from *gross recording or measurement errors*
- result from an *incorrect distributional assumption*
- indicate that *the data contain more structure than is being used*
- *an unusual observation merely indicates that such a value is possible !*

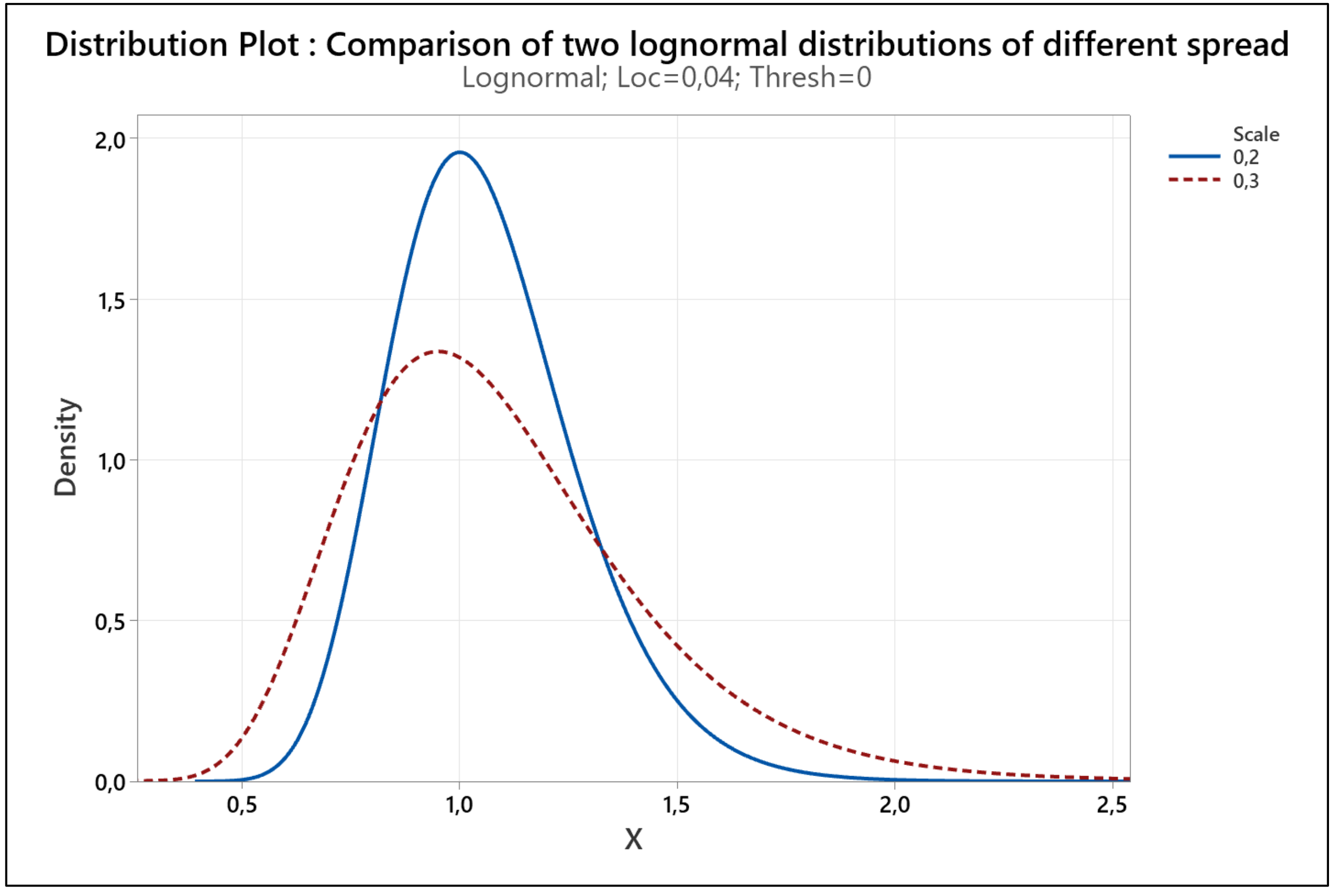
B. Iglewicz, D.C. Hoaglin, *How to Detect and Handle Outliers*, American Society for Quality Control, Vol. 16, (1993)

OUTLIERS

Genesis of outliers

Simply put, here are two basic mechanisms which give rise to samples which appear to have outliers:

Mechanism 1: The data come from some heavy tailed distribution. The "outlier-prone" families of statistical distributions have tails which go to zero slowly.

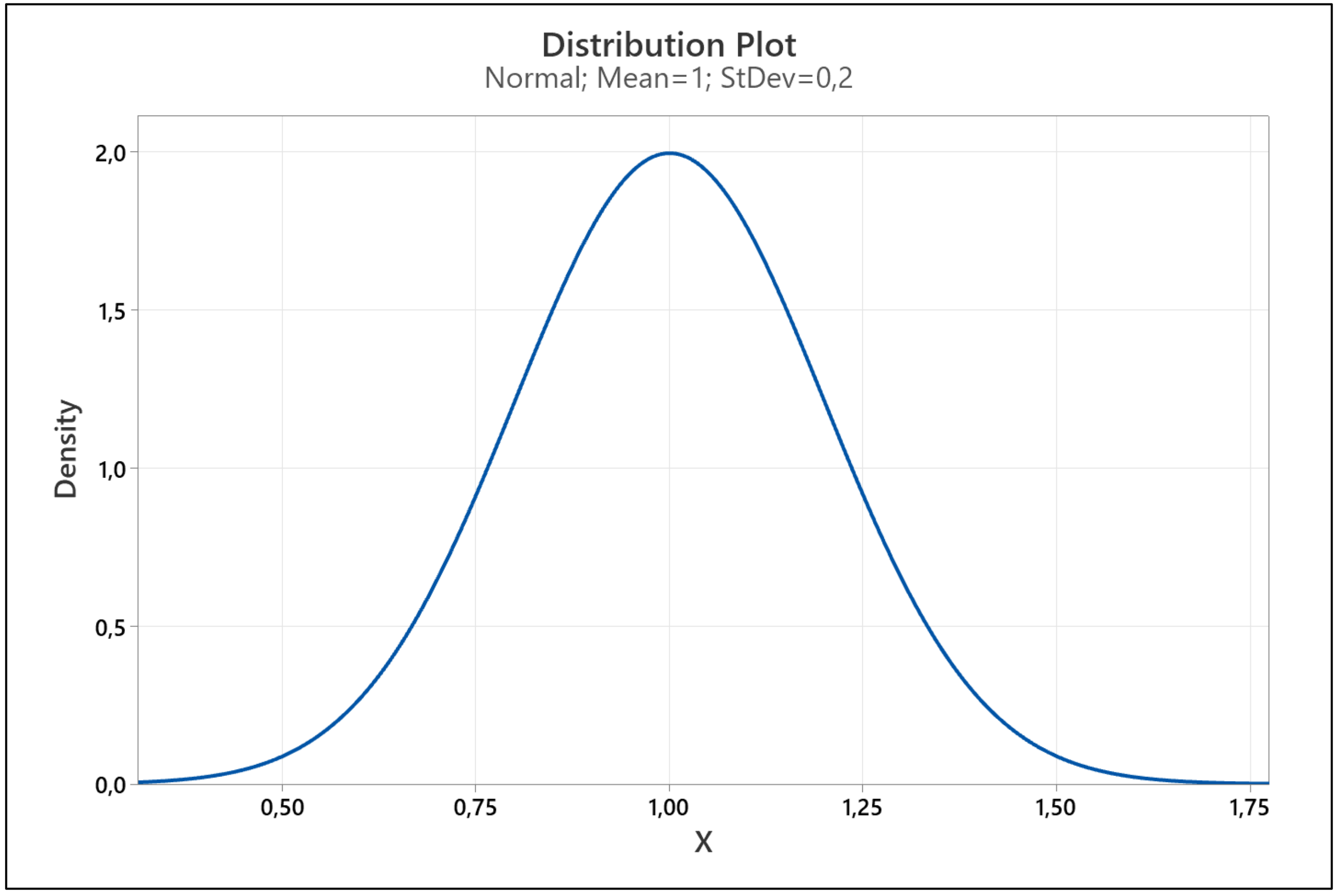


R.F. Green, *Outlier-Prone and Outlier-Resistant Distributions*, J. Am. Stat. Assoc., 71 (1976) 502-505

OUTLIERS

Genesis of outliers

According to Green's cataloging, alongside this type of distributions there are others that can be classified as “absolutely” or “relatively” outlier-resistant which are not absolutely or relatively outlier-prone, respectively. The first class includes the normal family of distributions.



R.F. Green, *Outlier-Prone and Outlier-Resistant Distributions*, J. Am. Stat. Assoc., 71 (1976) 502-505

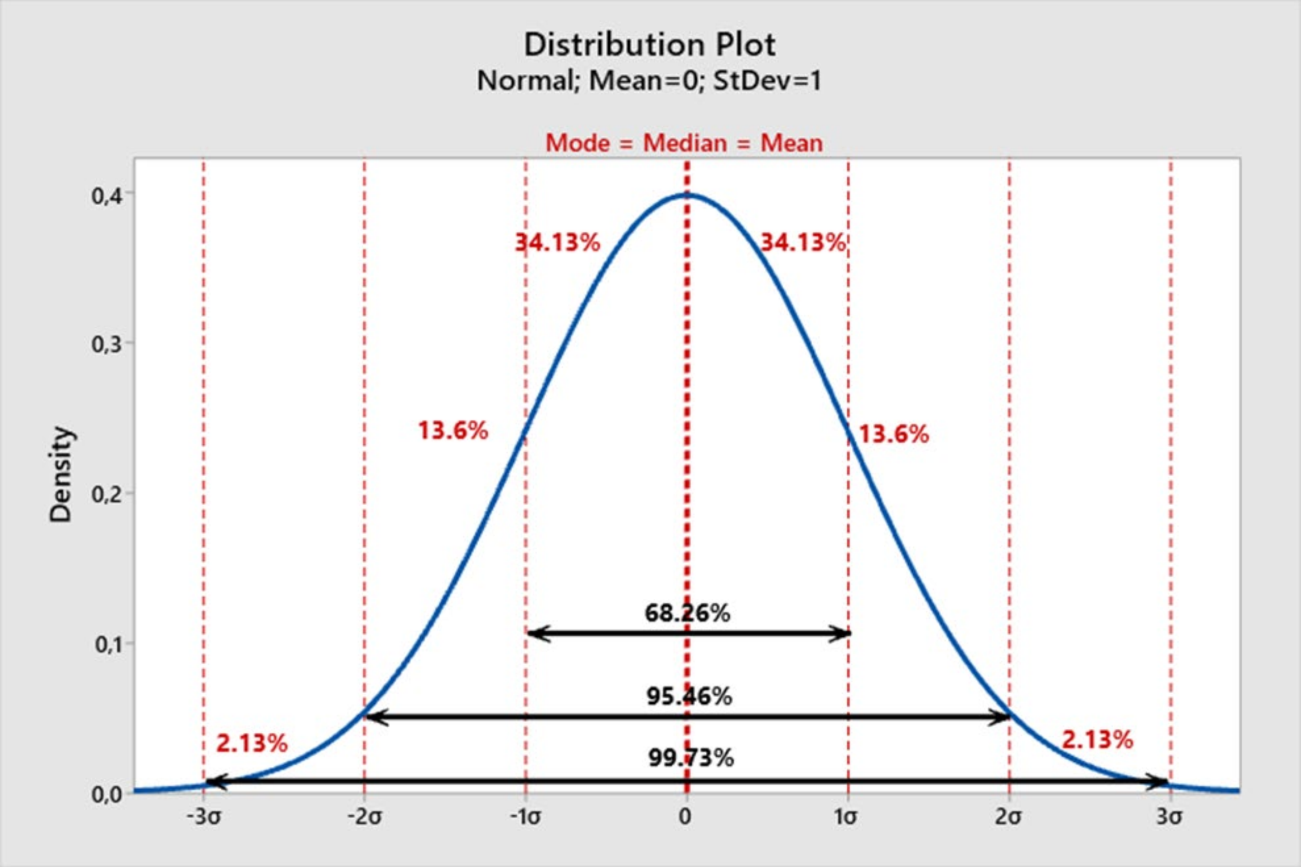
OUTLIERS

Genesis of outliers

Since the probability that an observation falls within two standard deviations of the mean is 95.4%, it is frequently seen as a rule for detecting outliers:

$$|x - \mu| > 2\sigma$$

But this may be misleading! Usually, in fact, we are dealing with samples (often even small ones) and not with the entire population!



OUTLIERS

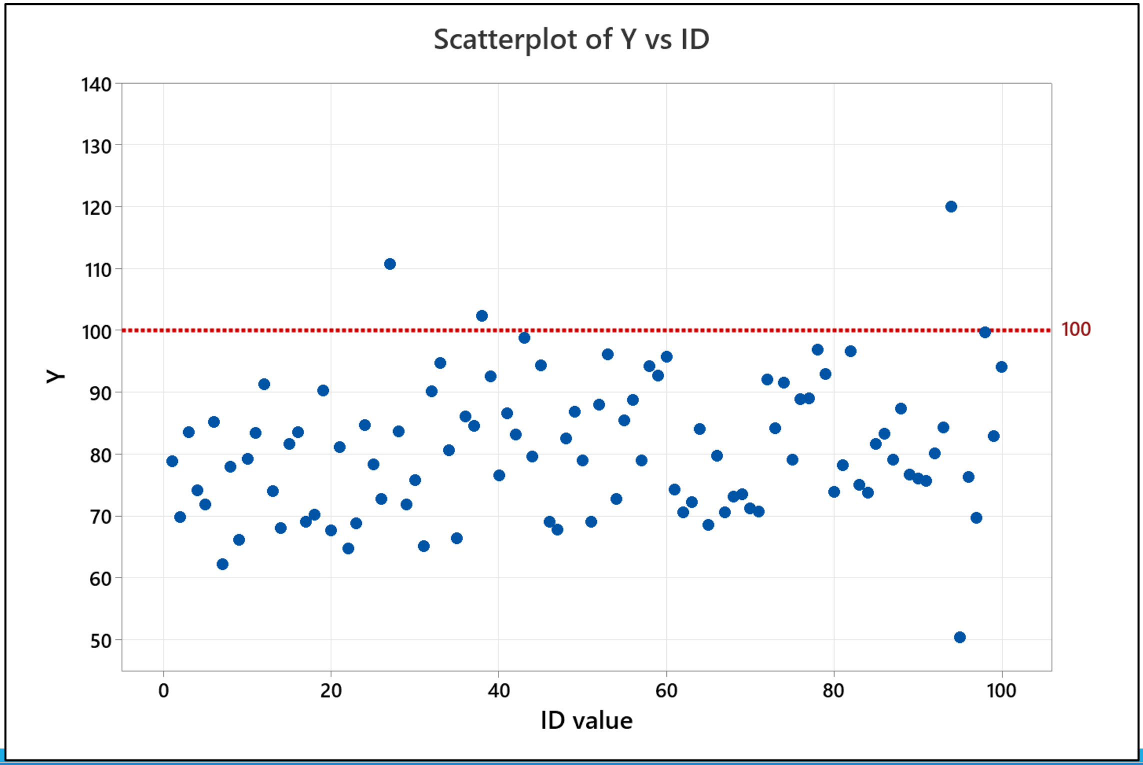
Genesis of outliers

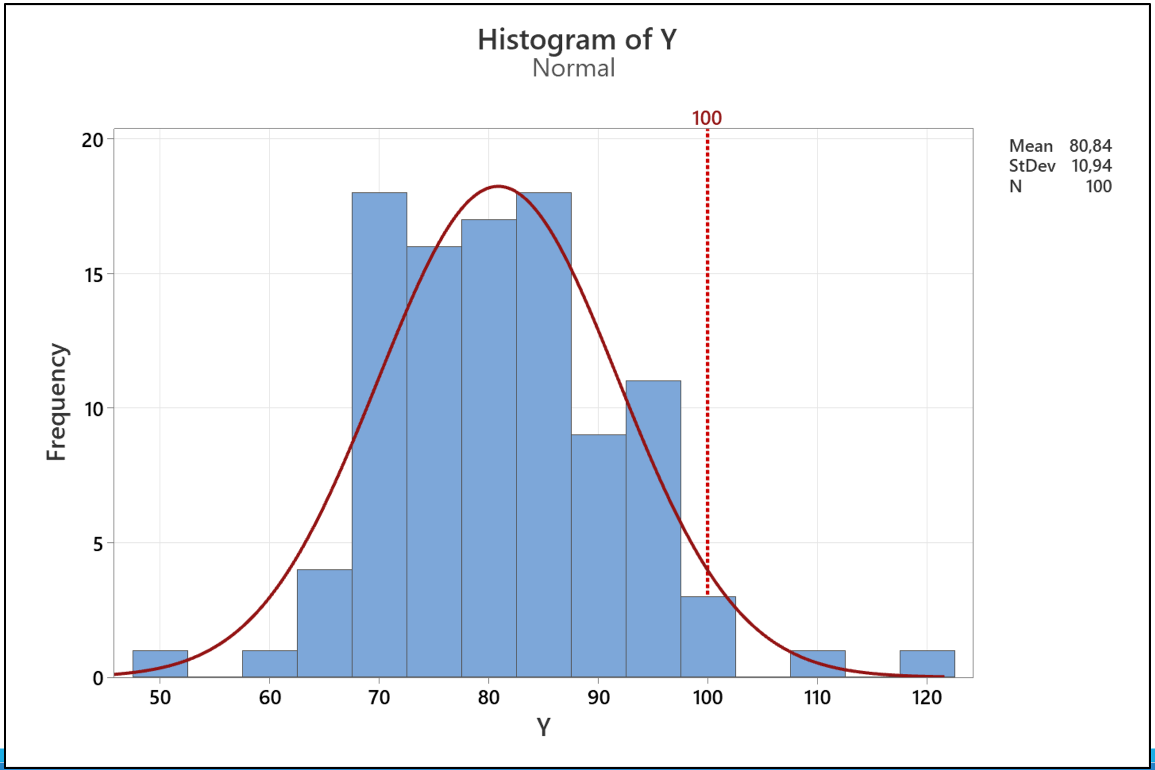
WARNING !

If the values are distributed close to a given limit, even if distributed normally they can exceed it.

The normal distribution is a continuous distribution with tails that extend infinitely in both directions !

Hence the importance of studying how our data is distributed!



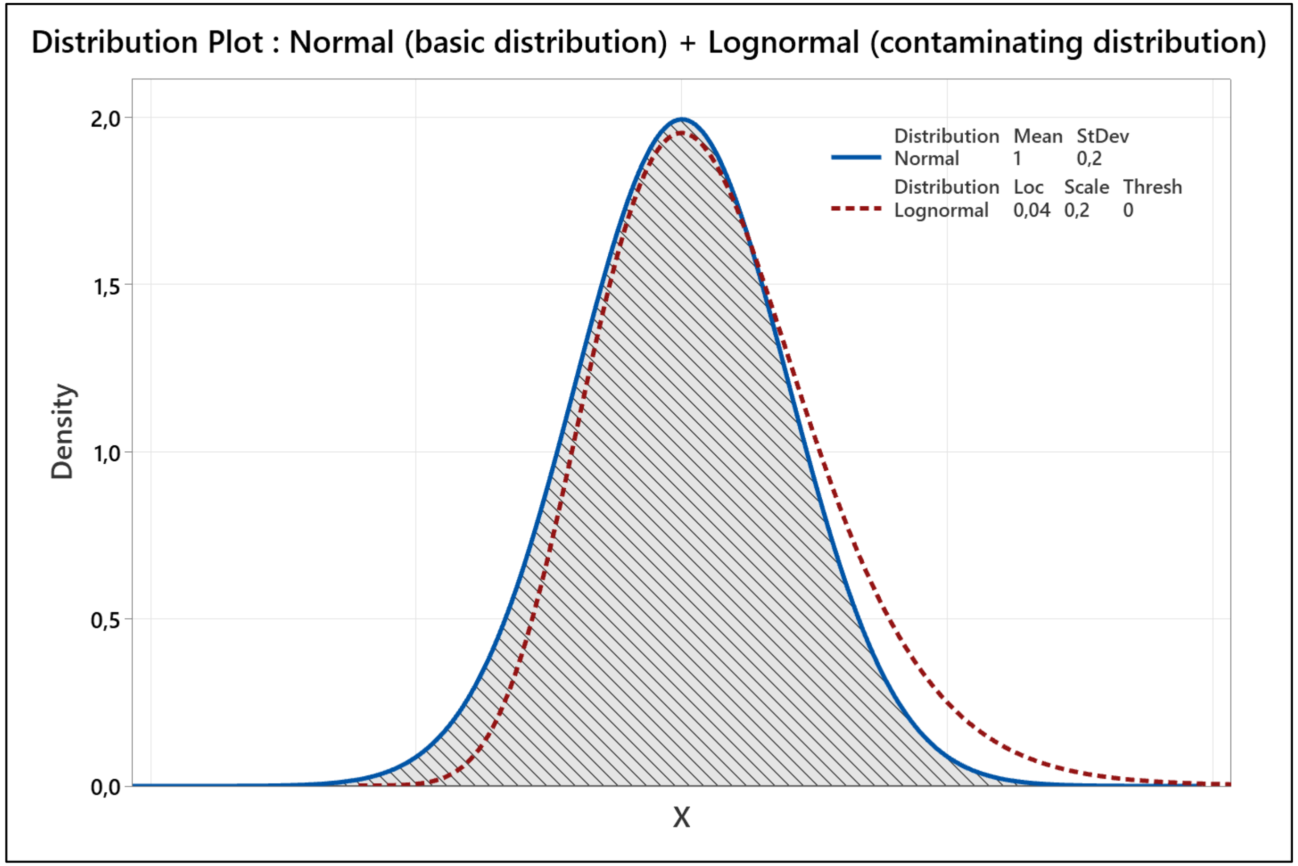


OUTLIERS

Genesis of outliers

Mechanism 2: The data arise from (at least) two distributions:

- the *basic distribution*, generates "good" observations, while
- the *contaminating distribution* generates "contaminants". If the contaminating distribution has tails heavier than the basic distribution there will be tendency for the contaminants to be outliers.



OUTLIERS

Genesis of outliers

WARNING !

- ***simply put: the presence of multiple distributions indicates the co-presence of multiple populations and therefore of multiple processes***

- ***in other words:***

the underlying manufacturing process is out of (statistical) control

OUTLIERS

- What we have just seen are not abstruse theoretical considerations, but aspects of practical and immediate impact.

- In fact, ***the effect of outliers on the analysis of a set of data depends strongly on the mechanism by which the outliers are believed to be generated.*** For instance, if Mechanism 1 is assumed, then the outliers, despite appearances, are valid observations from the distribution under study.

What is the usefulness of these considerations?

They show us the way to systematically address outliers !

In other words, outliers must be:

1. *Identified*
2. *Investigated*
3. *Treated*

OUTLIERS

How to identify Outliers

There are two main ways to identify outliers to identify outliers in a data set:

- ***graphical methods***: boxplots, histograms, scatterplots, *etc.* allow to visually inspect data and spot any unusual values.

  This approach is ***recommended*** because it is ***simple to use***, gives an immediate interpretation response and, above all, ***does not involve any assumptions on the distribution of the data*** (*e.g.*, normality).

OUTLIERS

How to identify Outliers

- ***statistical tests***: Dixon's test, Grubb's test, Modified *Z* Scores, 1-Sample *t*-test, *etc.*

  Since this approach usually relies on assumptions about the distribution of the data, it ***requires a lot of care to prevent errors***.

  It should be avoided if possible but, unfortunately, often, old procedures or agreements with Customers contain one or more tests such as those of Dixon or Grubb.

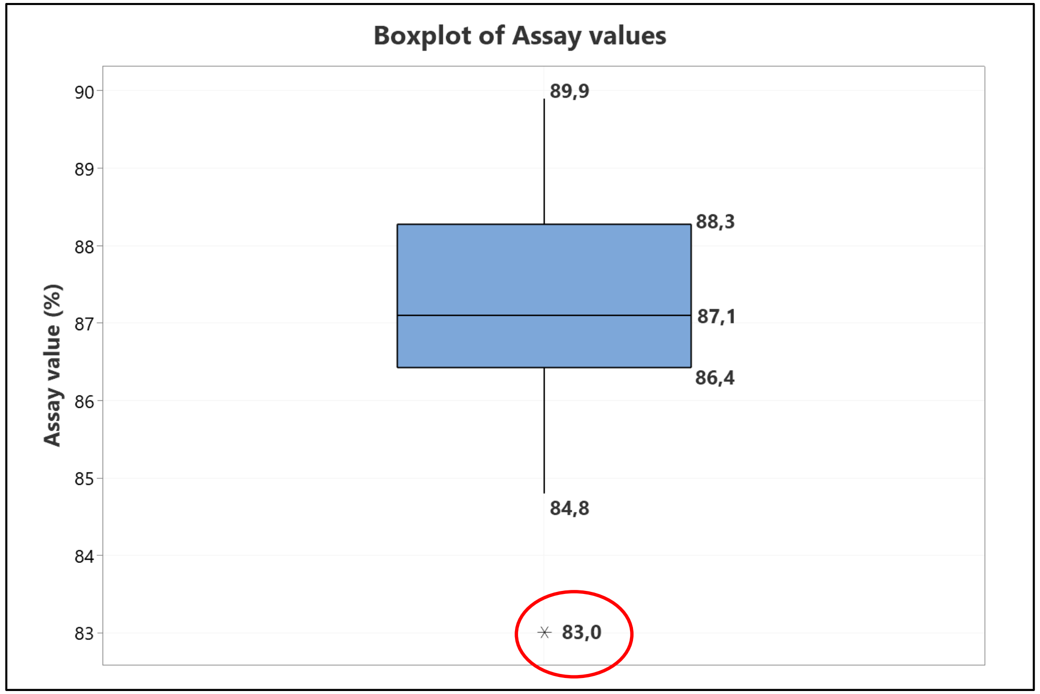
Let's start with

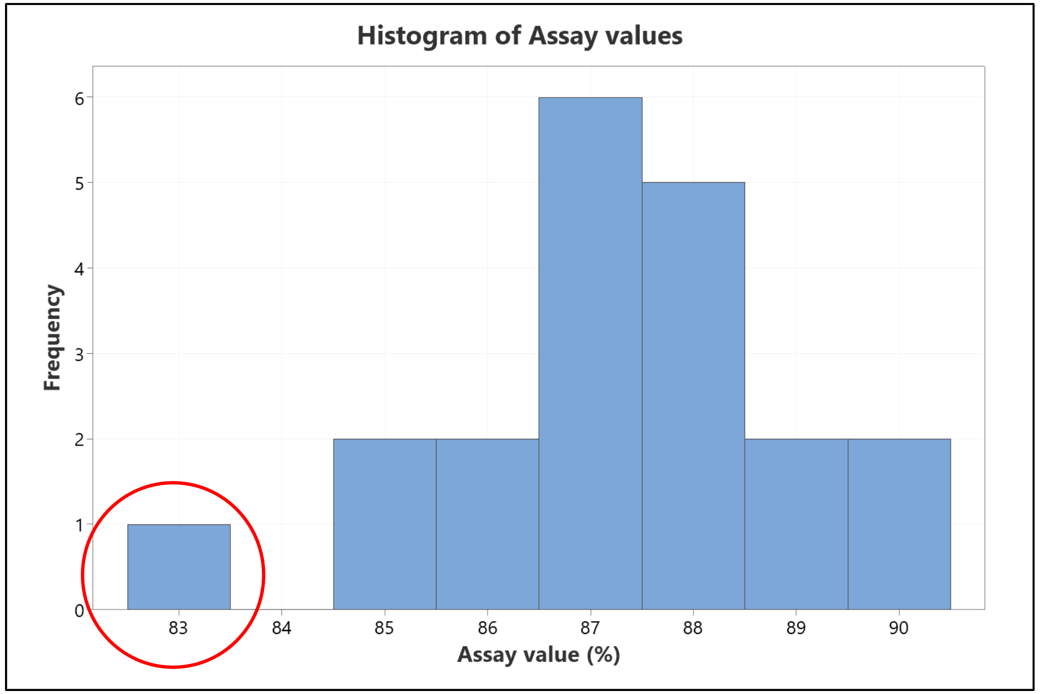
GRAPHICAL METHODS

OUTLIERS

Assay value (%)
86,6
88,2
86,4
88,3
85,4
89,9
84,8
87,0
89,6
88,8
86,1
87,9
83,0
88,5
87,2
88,0
86,5
87,5
87,0
87,0

Boxplot and Histogram

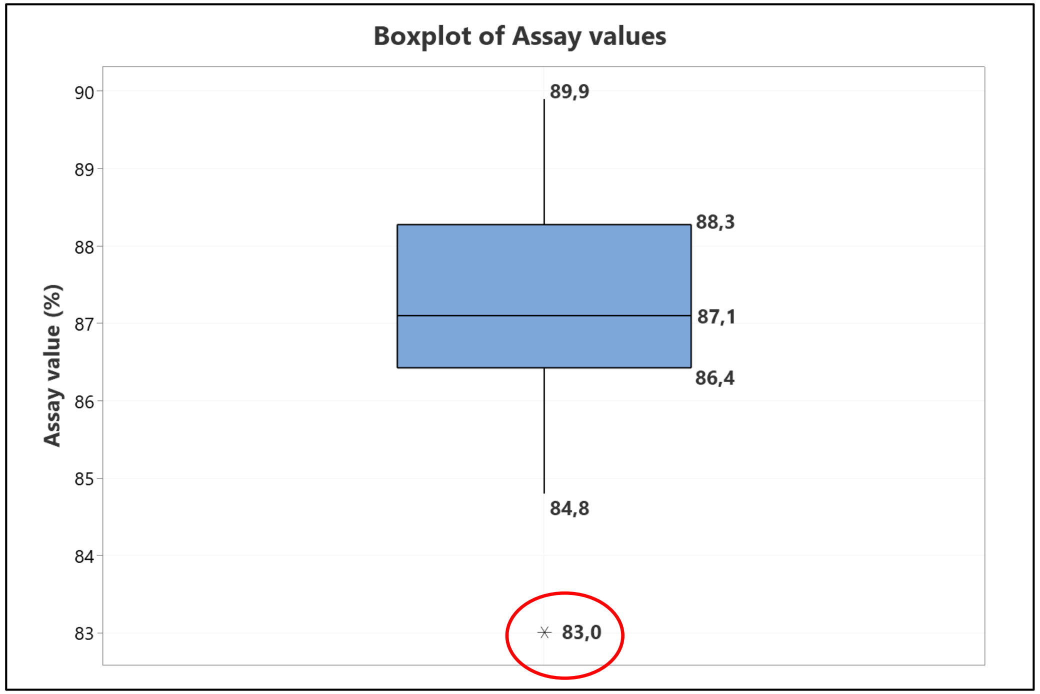


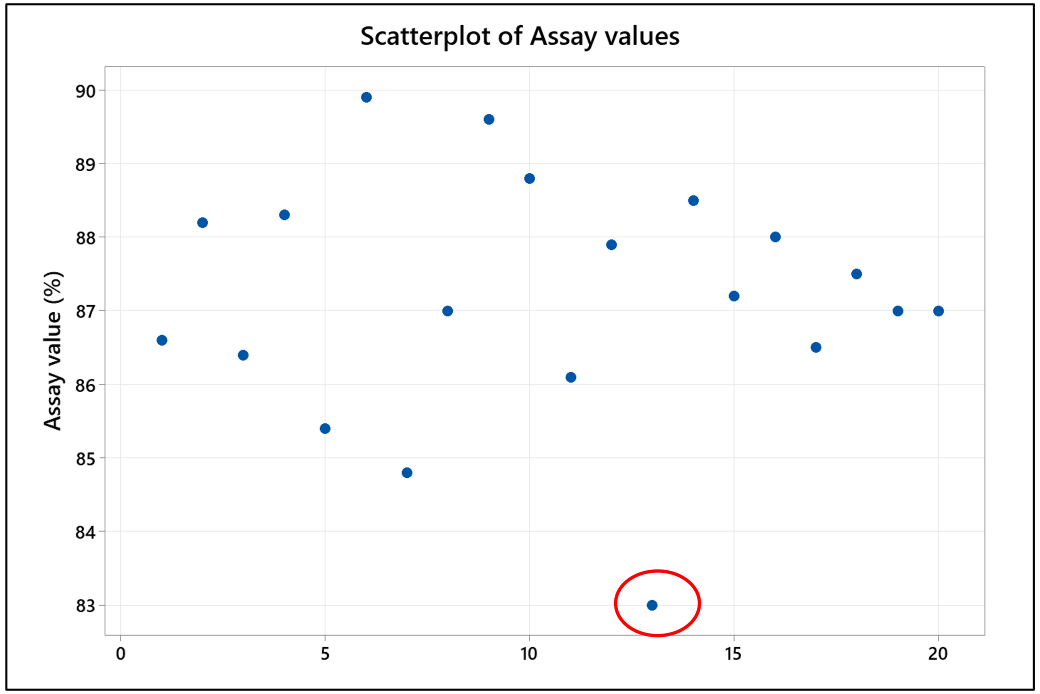


OUTLIERS

Assay value (%)
86,6
88,2
86,4
88,3
85,4
89,9
84,8
87,0
89,6
88,8
86,1
87,9
83,0
88,5
87,2
88,0
86,5
87,5
87,0
87,0

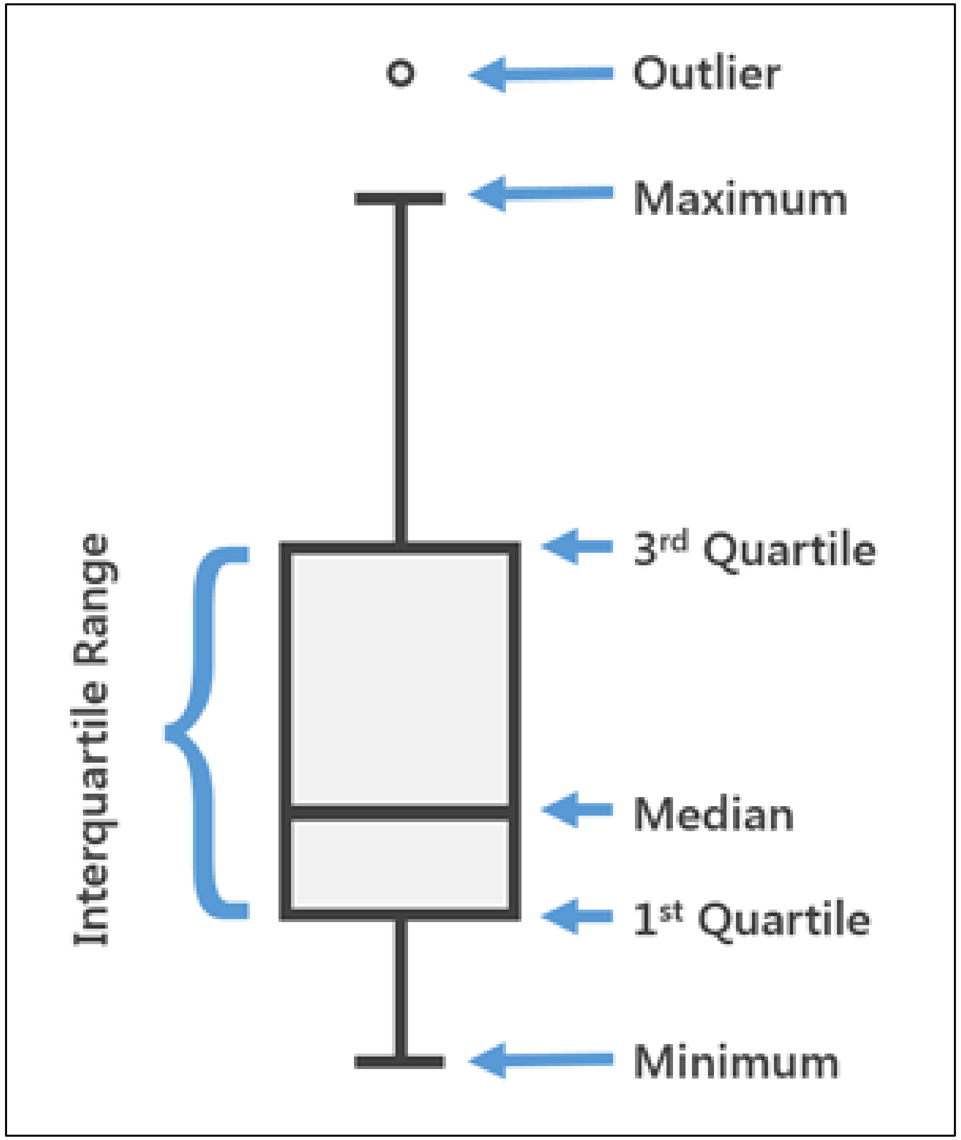
Boxplot and Scatterplot





OUTLIERS

1st Quartile, Q1:	25% of the data ≤ this value
Median, Q2:	50% of the data ≤ this value
3rd Quartile, Q3:	75% of the data ≤ this value
Interquartile range:	50% of the data
Whiskers:	extend to the minimum / maximum date point within 1.5 IQR from the bottom / top of the box
Outlier :	observation beyond upper or lower whisker, *i.e.,* over 1.5IQR



J.W. Tukey, *Exploratory Data Analysis,* Addison Wesley, 1977

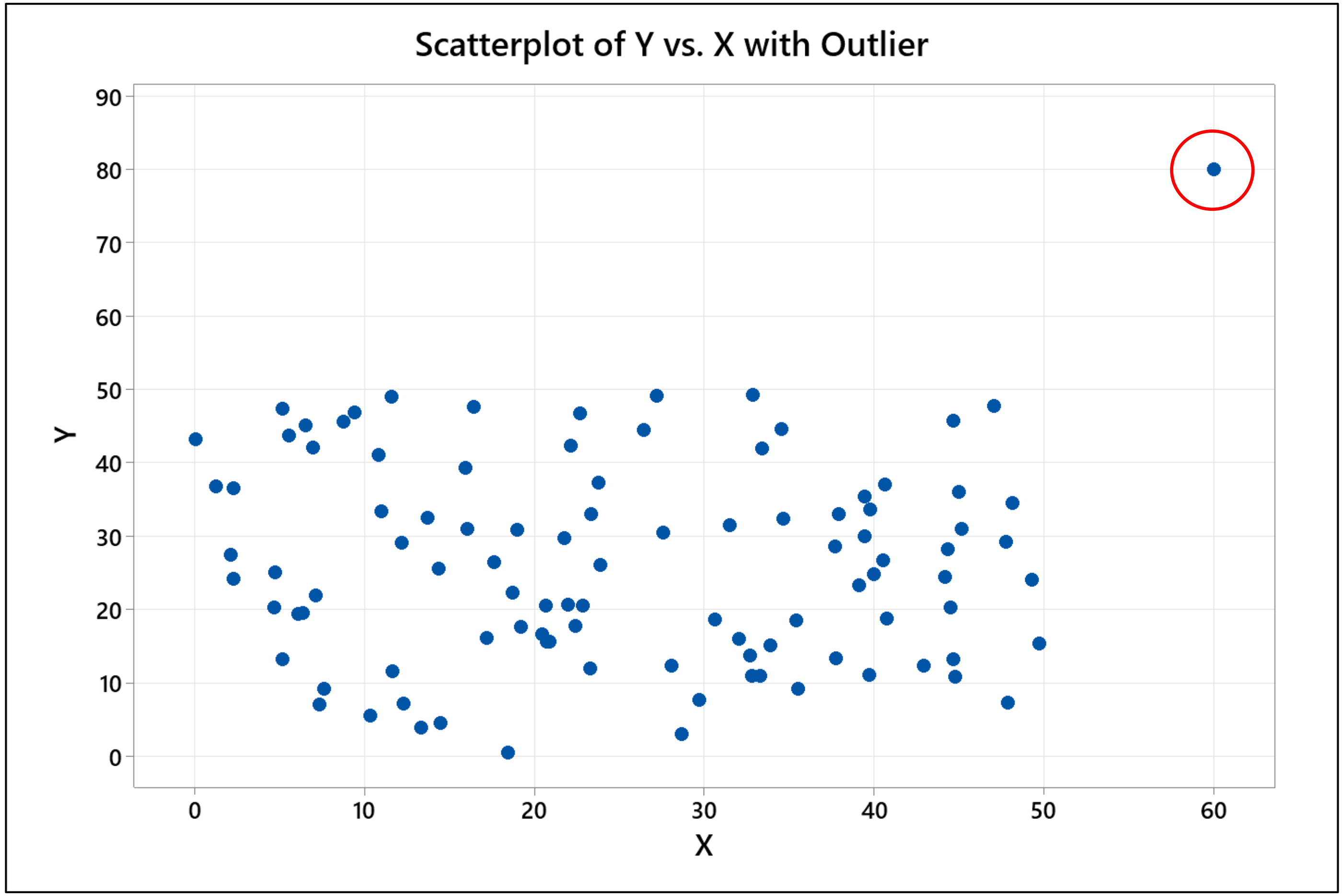
OUTLIERS

What does a Boxplot tell us at a glance?

- **If it looks «compact»** : most of the data are like each other since there are so many values in a narrow range
- **If it looks «stretched»** : most of the data are quite different from each other, as the values spread over a wide range
- **If the median is close to the bottom**: most of the data will have the lower range values
- **If the median is close to the top**: most of the data will have the higher values of the range
- **If the median is not in the center** data distribution will be « tailed »

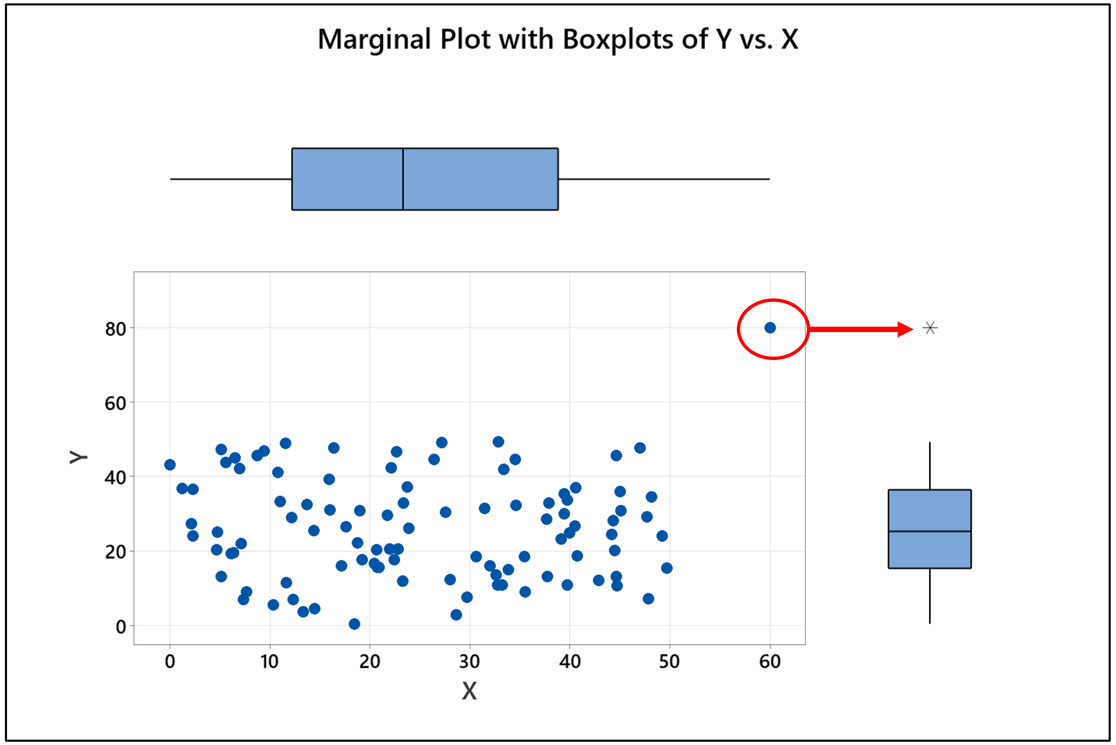
OUTLIERS

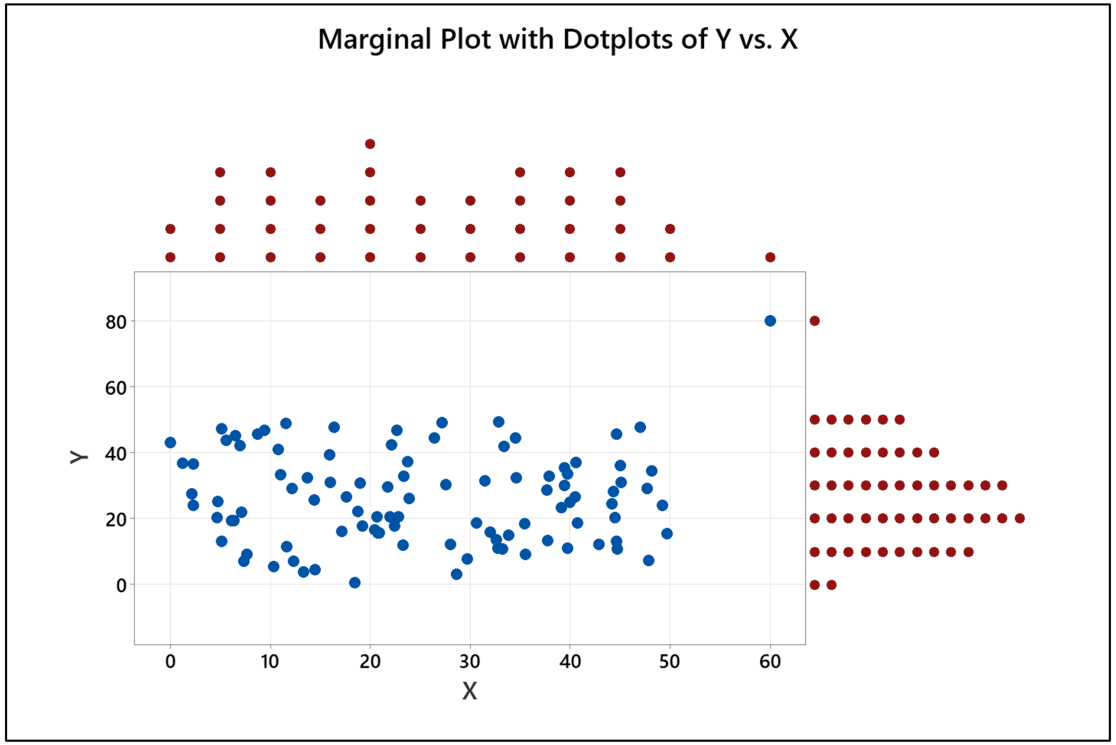
X	Y
14,38	25,58
39,42	30,00
20,45	16,64
44,15	24,43
47,02	47,72
2,28	24,15
26,41	44,52
44,62	45,72
27,57	30,44
22,83	20,53
47,84	7,35
22,67	46,76
33,88	15,06
28,63	3,04
5,15	47,39
44,99	36,03
…	…
…	…
39,11	23,33
4,68	20,34
23,34	32,96
60	80



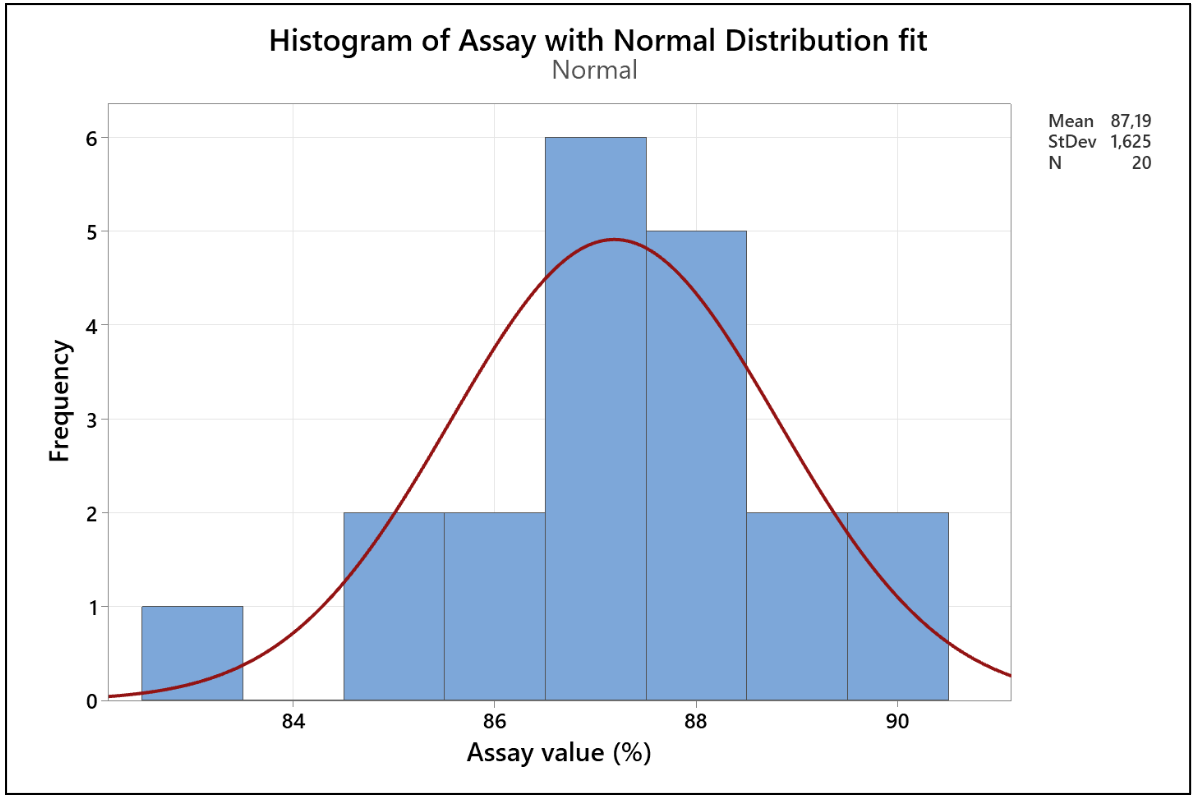
OUTLIERS

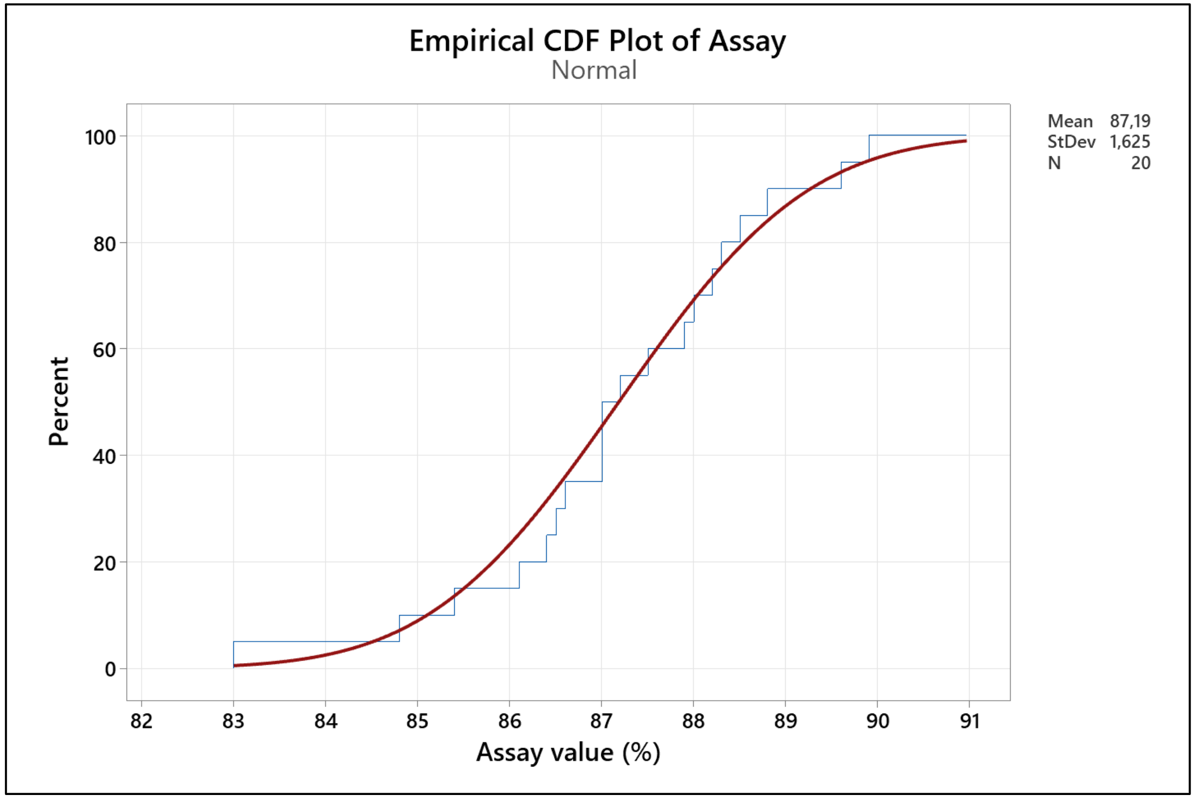
Marginal Plots (*i.e.*, scatterplots with, boxplots, dotplots or histograms, in the margins of the x- and y-axes) are also useful to assess the relationship between two variables and examine their distributions.





Empirical Cumulative Distribution Function (eCDF)





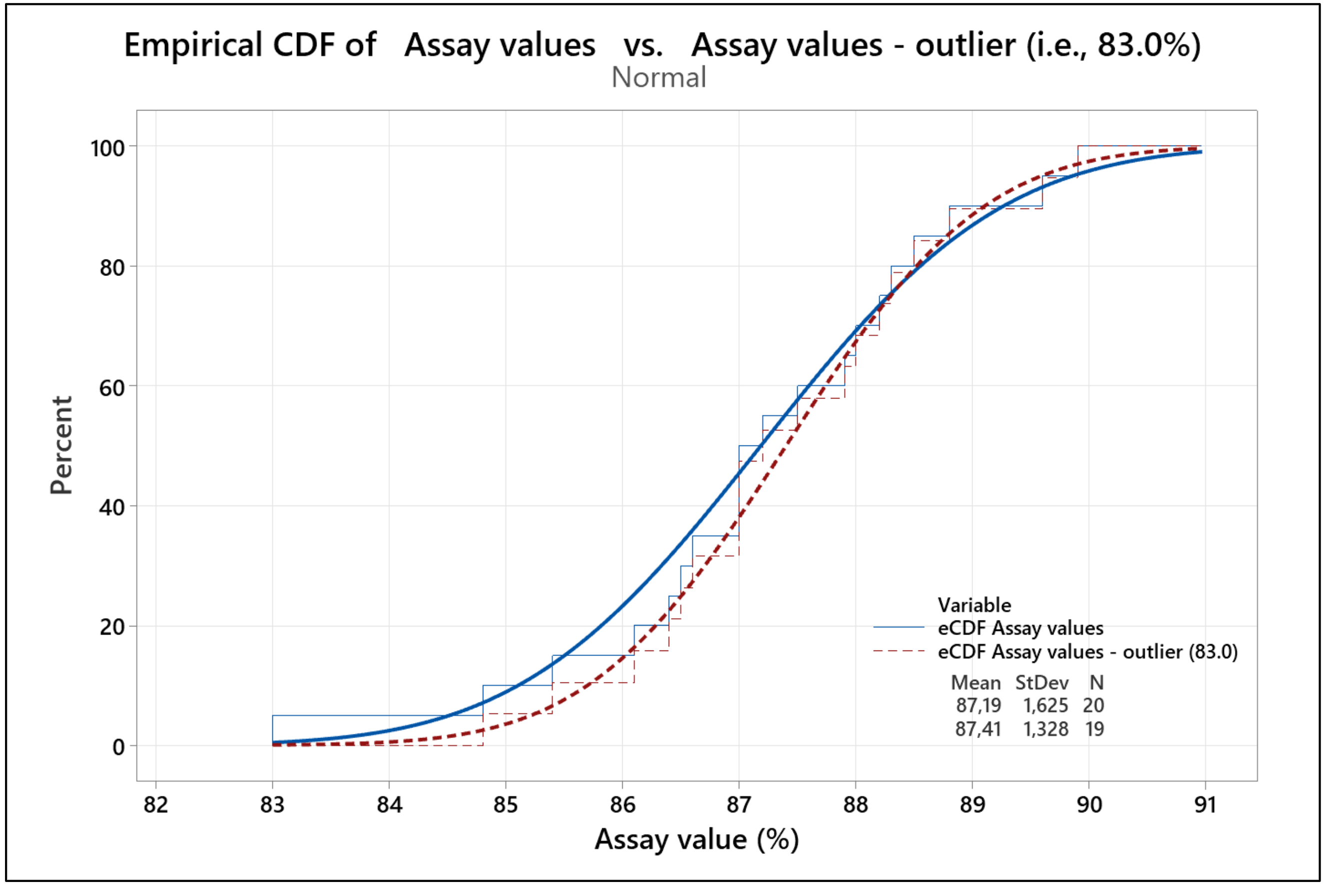
eCDF, *i.e.* the Cumulative Distribution Function, is constructed using experimental data.

OUTLIERS

Empirical Cumulative Distribution Function (eCDF)

The graph opposite compares the eCDFs calculated using all the available data (and therefore also the outlier) and only the data without the outlier value (*i.e.*, 83.0%).

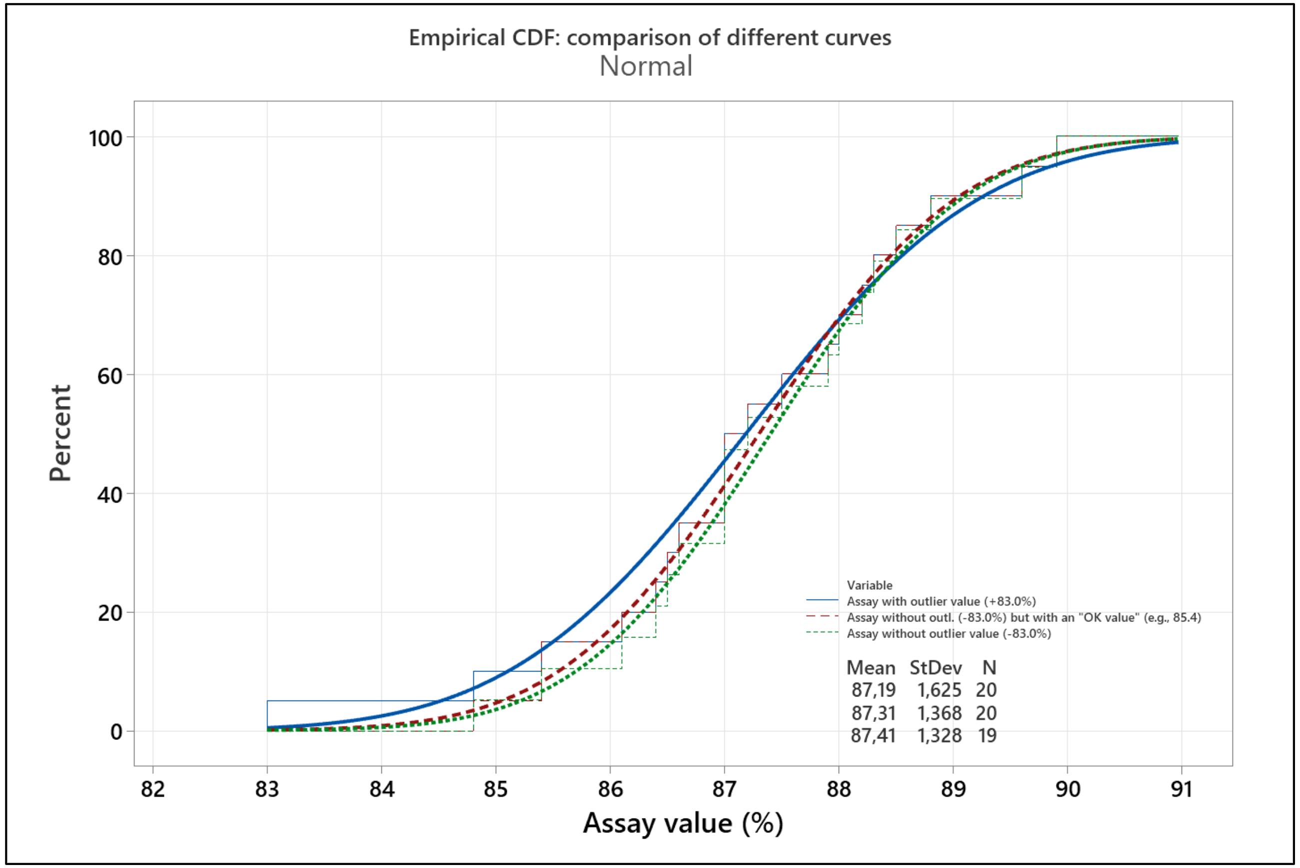
The shift in the curve resulting from the removal of the outlier is evident!



OUTLIERS

Empirical Cumulative Distribution Function (eCDF)

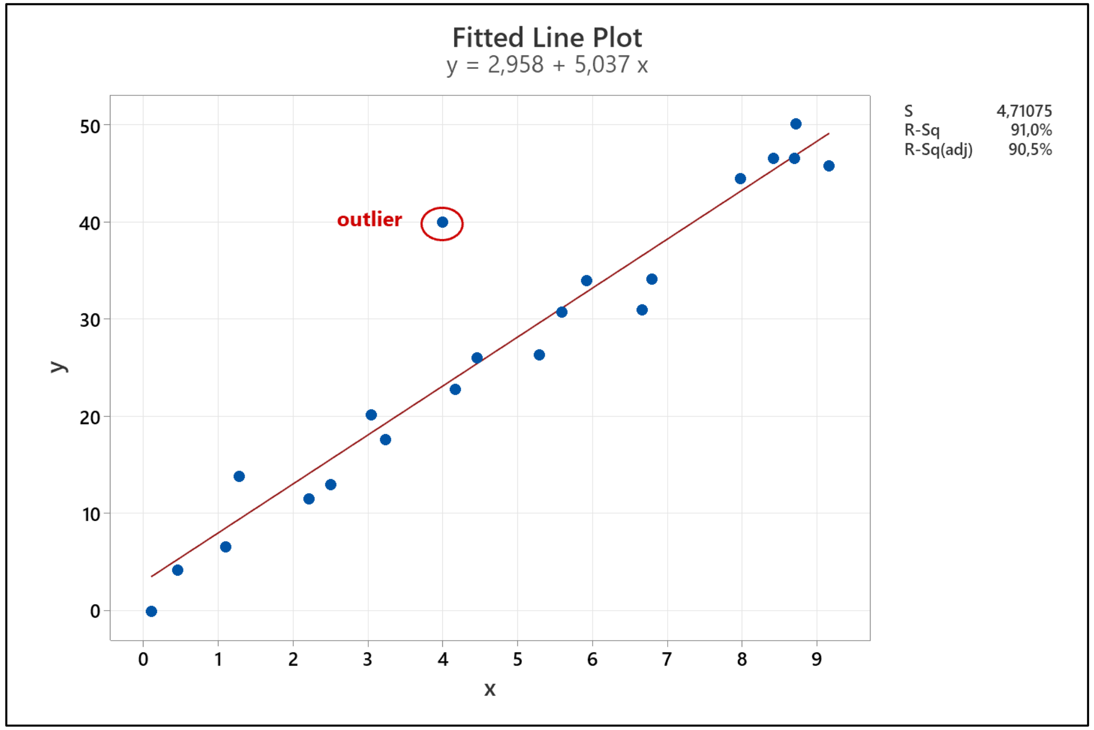
That the shift is due to the outlier is evident because by replacing this data (*i.e.*, 83.0%) with another value in line with the remaining values, the eCDF curves are practically superimposable, as shown alongside.

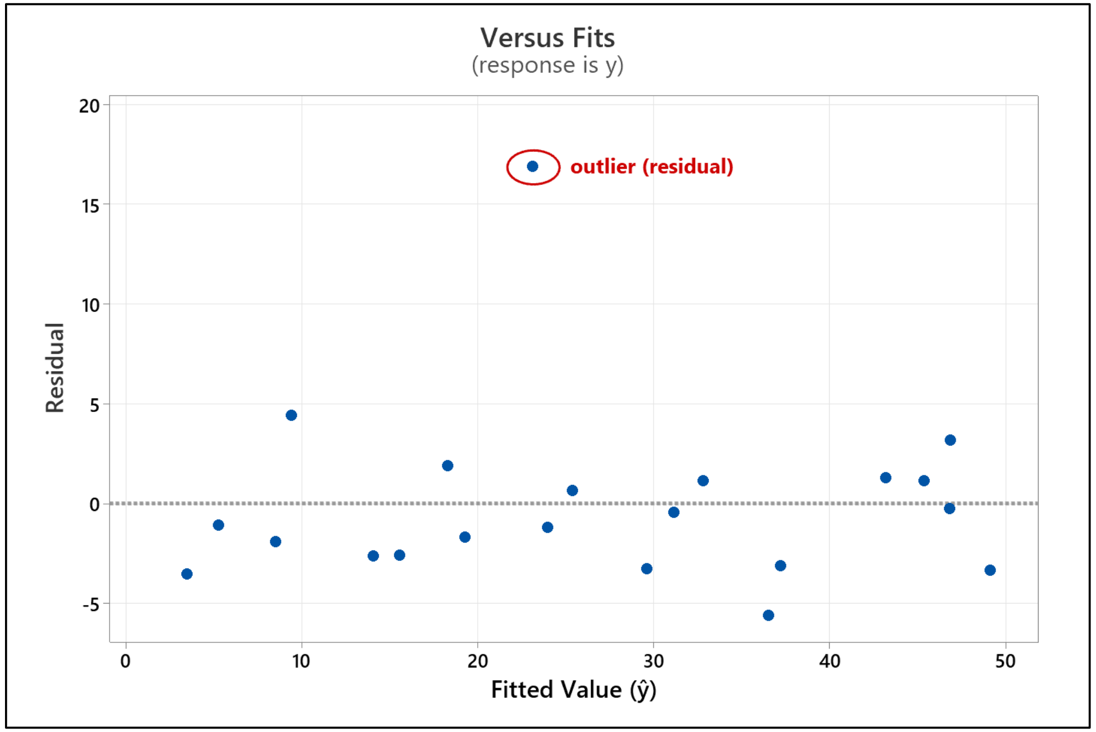


OUTLIERS

In Regression Analysis outliers are extreme observations that do not fit the model well.

Simply put, they can be identified in many ways, mainly from *scatterplots* (or *fitted line plots*) and the corresponding *residual plots*.

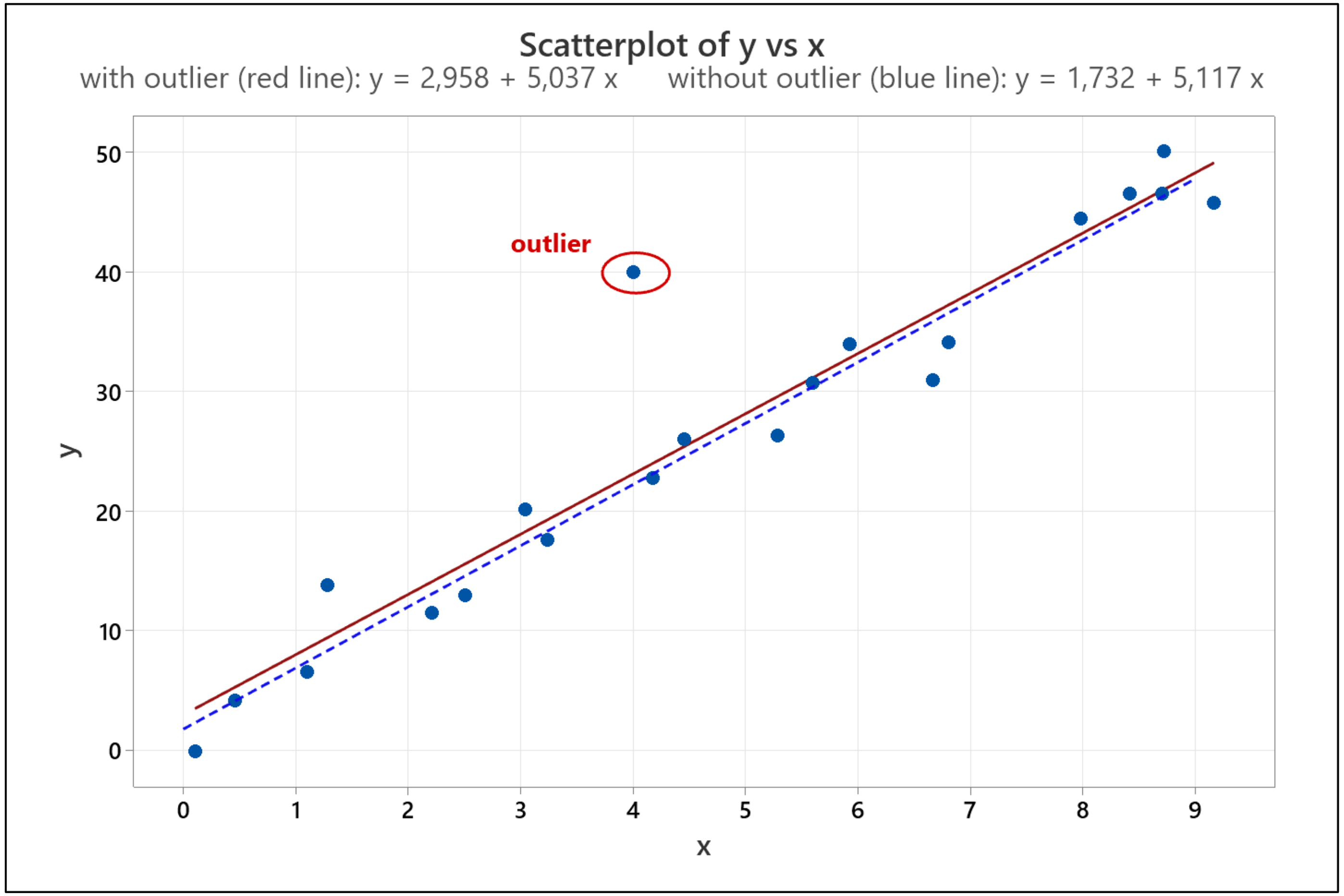




OUTLIERS

In this specific context, *outliers* are extreme values in the *y* direction relative to the fitted regression line.

By removing the outlier and recalculating the regression line, it is parallel to those of the observations that also included the outlier.



OUTLIERS

In Regression Analysis it is always important to keep in mind that:

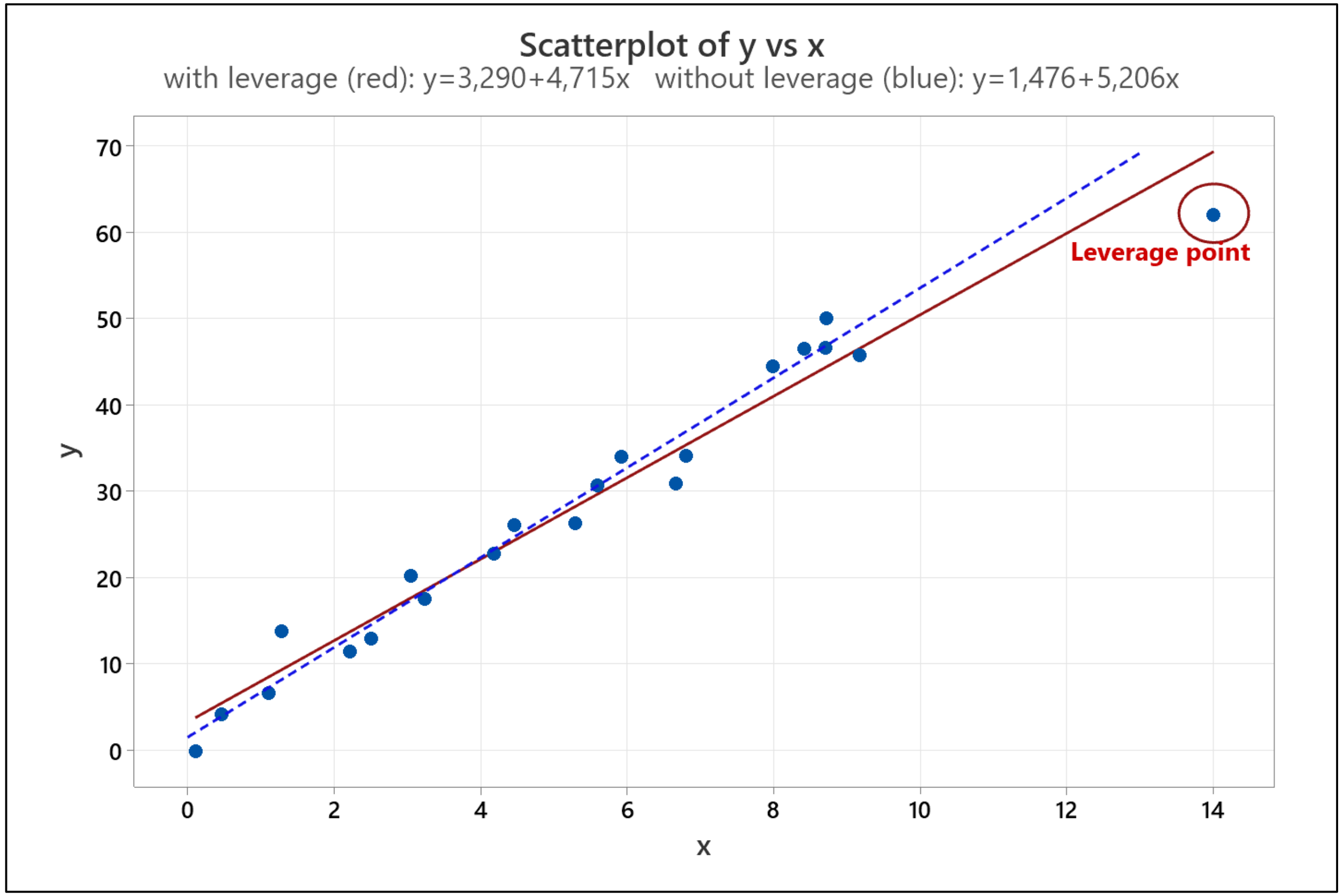
Not all outlying cases have the same influence in the fitted regression function!

The example we just saw appears not to be “too influential” because other data points have similar *x* values that will prevent the fitted regression function from being displaced too far by the outlying case.

However, there are other situations (like the one illustrated on the next slide) where the outlying value can have a lot of influence in affecting the fit of the regression function.

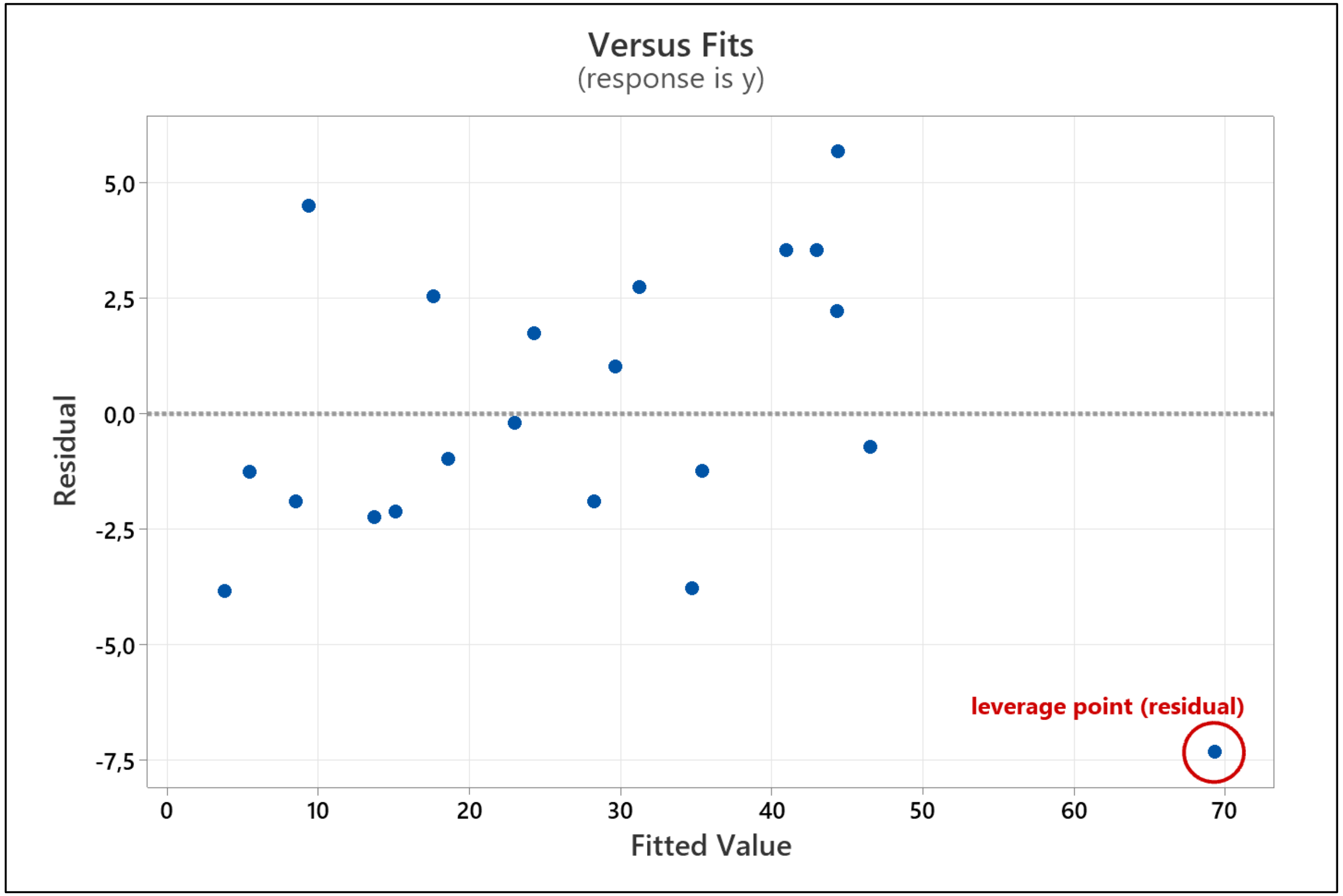
OUTLIERS

In the case of the so-called *leverage points*, by removing the outlier and recalculating the regression line, this turns out to be "rotated around a point" which acts as a fulcrum compared to that of the observations which also included the *leverage point*.



OUTLIERS

Even in the case of *leverage points*, examining the residual graph helps.



STATISTICAL TESTS

The situation is more complex and requires maximum attention!

OUTLIERS

- The most used statistical tests for outliers assume that the **data is normally distributed**.
- In general, **tests are designed to detect a single outlier in a sample**. If a sample contains more than one potential outlier, tests may not be effective.
- Therefore, applying these tests to data known to be, for example, highly biased (*Mechanism 1*) could result in the rejection of legitimate data.

Once again, we return to the concept that

the distribution of data must first be statistically investigated!

OUTLIERS

*«… Many statistical methods used for assessing validation characteristics rely on population normality, and **it is important to determine whether or not to reject this assumption**. There are many techniques, such as **histograms**, **normality tests**, and **probability plots** that can be used to evaluate the observed distribution. It may be appropriate to **transform the data** to better fit the normal distribution **or** **apply distribution-free (nonparametric) approaches** when the observed data are not normally distributed… »*

FDA Guidance for Industry, *Analytical Procedures and Methods Validation for Drugs and Biologics* (July 2015)

OUTLIERS

- Therefore, first of all we need to verify the *normality of the data*.
- How can we do it ?
- And if the data is not normal, what do we do?

OUTLIERS

Verification of data normality

- There are several ***statistical tests of normality*** such as:
  - Anderson Darling
  - Shapiro Wilk
  - Kolmogorov Smirnov, *etc*.

and

- Two reference ***graphical methods***, namely:
  - Normal Probability Plot
  - QQ Plot

Statistical tests for verifying data normality

- **Shapiro-Wilk Test**:
  assesses whether a dataset follows a normal distribution by comparing the observed data to what would be expected in a normal distribution.
- **Anderson-Darling Test:**
  is similar to the Shapiro-Wilk test but is more sensitive to deviations from normality in the tails of the distribution.
- **Kolmogorov-Smirnov Test:**
  assesses the goodness of fit between the sample data and a theoretical normal distribution.

OUTLIERS

Statistical tests for verifying data normality

- In summary, all three tests estimate whether a dataset deviates from a normal distribution.

- Each test provides a test statistic that is compared to a critical value to make a determination about the normality of the data.

- The choice of which test to use depends on factors like the nature of your data and your specific requirements for sensitivity to deviations from normality.

Test for Normality	*Statistics*	*P-value*
Anderson- Darling (A)	0.26017	0.6731
Kolmogorov Smirnov (D)	0.11454	0.9556
Shapiro Wilk (W)	0.96455	0.6382

In all cases *P-value* > 0.05

OUTLIERS

Graphical methods for verifying data normality

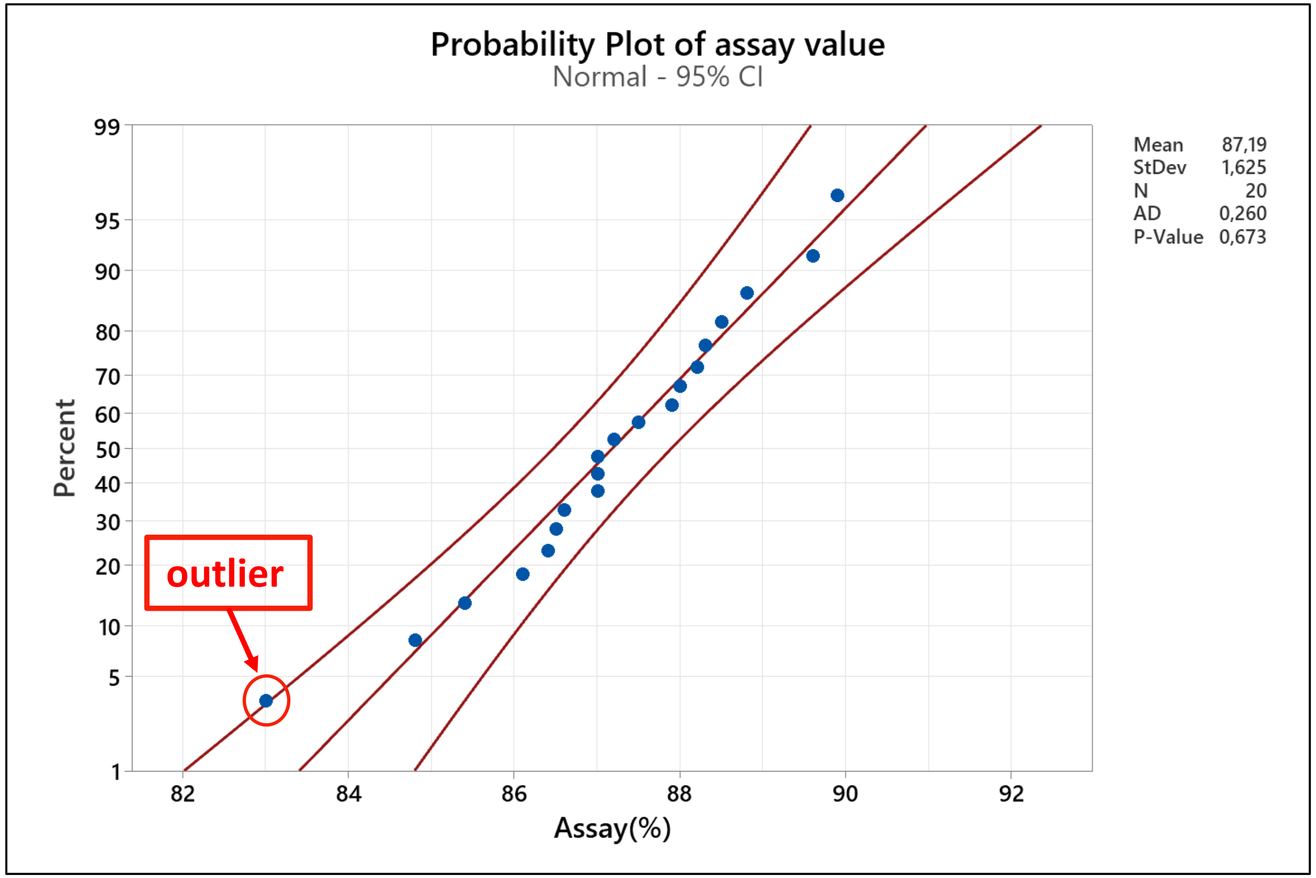
- **Normal Probability Plot** : is a graph that shows data relative to its theoretical position in a normal distribution. In other words, if the data follows a normal distribution, the points in the normal probability graph should follow a straight line. It can be used for a visual assessment of the normality of the data.

OUTLIERS

Graphical methods for verifying data normality

- From this graph it can be seen that the assay data follow the normal distribution as they appear arranged along a straight line.

- The data values are on the abscissa, while the estimated cumulative probability is on the ordinate.



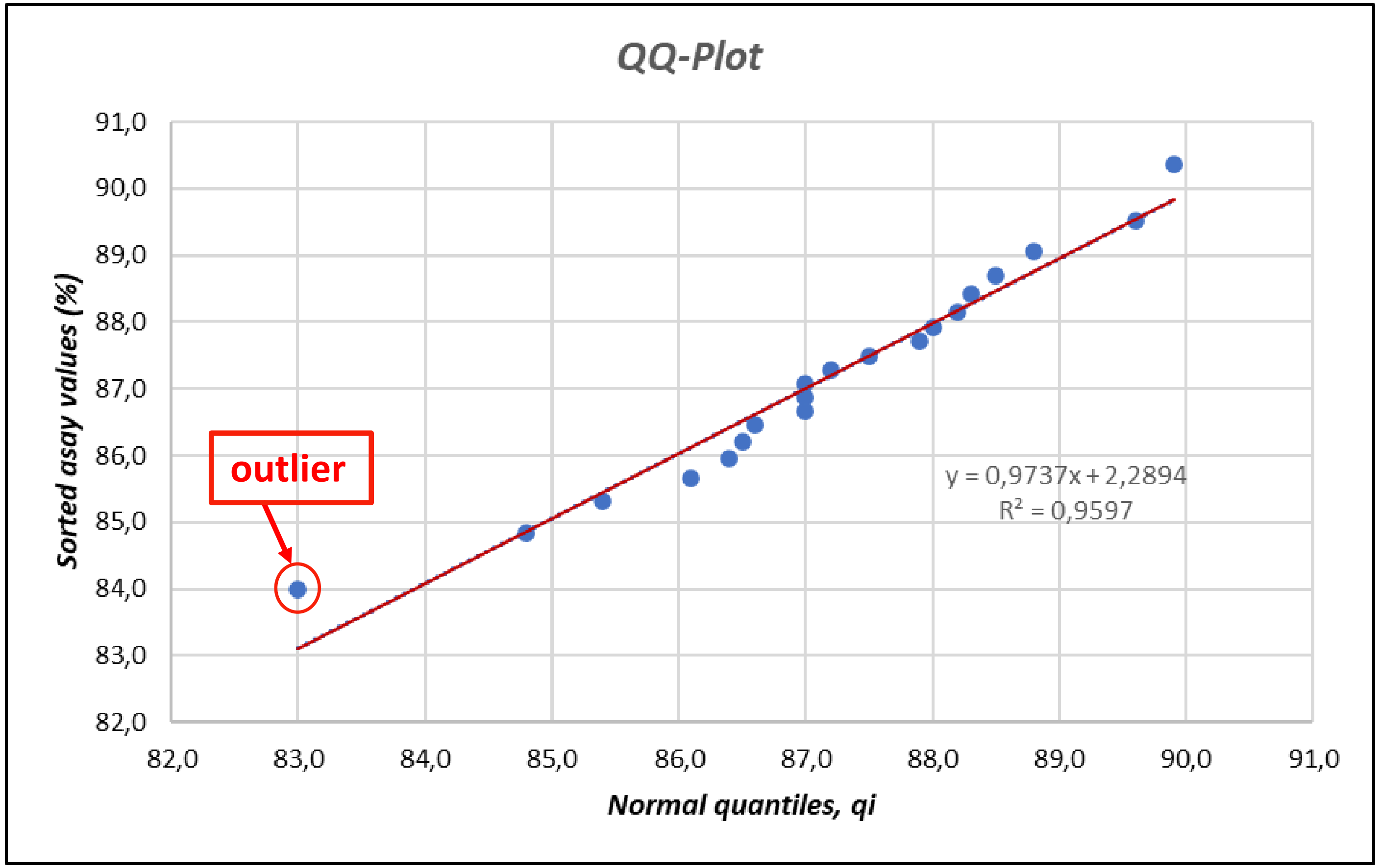
OUTLIERS

Graphical methods for verifying data normality

- **QQ Plot (Quantile-Quantile Plot)**: is a graph that compares the quantiles of the data with the quantiles of a theoretical normal distribution. Quantiles are values that divide the data into specific percentiles.
- In the QQ plot, if the data follows a normal distribution, the points should follow a diagonal straight line.
- It is also a detailed and precise method for evaluating the normality of the data, as it compares specific quantiles rather than just a general view.

OUTLIERS

Graphical methods for verifying data normality



OUTLIERS

Data not normally distributed

- When dealing with data that does not follow a normal distribution, but we wish to apply statistical tests that assume normality (*e.g.*, Dixon's test), one approach is to mathematically transform the data.
- By employing straightforward mathematical operations such as *logarithmic transformation* or *square root extraction*, we can convert the original data into a form that approximates a normal distribution.
- The choice of the most suitable transformation method depends on the nature of the data and the size of the dataset. ***In any case, it's not easy and should be the last choice!***

OUTLIERS

Let's now look at the main statistical tests for outliers!

- *Dixon's test*
- *Grubbs' test*
- *1-Sample t-test*
- *Z-Scores / Modified Z-Scores*
- *Cochran's test**

Dixon's Test for Extreme Values or Q-test

- Dixon's Q test (or rather, series of tests) for single outliers was proposed by the American statistician Wilfried John Dixon in the 1950s and applies to small ($\leq$ 30) and normally distributed data sets.

- The test requires that the observations be sorted in order of magnitude (increasing or decreasing).
  The ratio of the difference between the extreme value and its immediate neighbor to the range of observations is then calculated.

W.J. Dixon, F.J. Massey, *Introduction to Statistical Analysis*, 3rd Ed., McGraw-Hill, 1969

OUTLIERS

Dixon’s Test for Extreme Values or Q-test

- The ratios thus calculated are compared with tabulated values (Q_{crit}) which vary depending on the number of observations and the chosen significance level (*e.g.*, α = 0.05 or C = 95%).

$$Q = \left|\frac{x_2 - x_1}{x_n - x_1}\right| \qquad Q = \left|\frac{x_n - x_{n-1}}{x_n - x_1}\right|$$

lower value higher value

n	3	4	5	6	7	8
Q_{crit}	0,94	0,77	0,64	0,56	0,51	0,48

C = 95%

Dixon's Test for Extreme Values or Q-test - Example

- Let's consider the assay values listed here on the side.
- 83.0, the lowest value in the series, is suspected to be the outlier.
- The data is sorted in ascending order and since n = 20 the ratio is calculated:

$$r_{20} = \left|\frac{x_2 - x_1}{x_{20} - x_1}\right| = \left|\frac{84.8 - 83.0}{89.9 - 83.0}\right| = \left|\frac{1.8}{6.9}\right| = 0.261$$

- In the table of critical values for the Dixon Test r_{10} we find that the critical value for n=20 is equal to 0.392 for α = 0.01 and therefore since 0.261 < 0.392, the value 83.0 is not considered an outlier at 99% confidence.

Ordered assay data
83,0
84,8
85,4
86,1
86,4
86,5
86,6
87,0
87,0
87,0
87,2
87,5
87,9
88,0
88,2
88,3
88,5
88,8
89,6
89,9

Dixon's Test for Extreme Values or Q-test – Example using Minitab

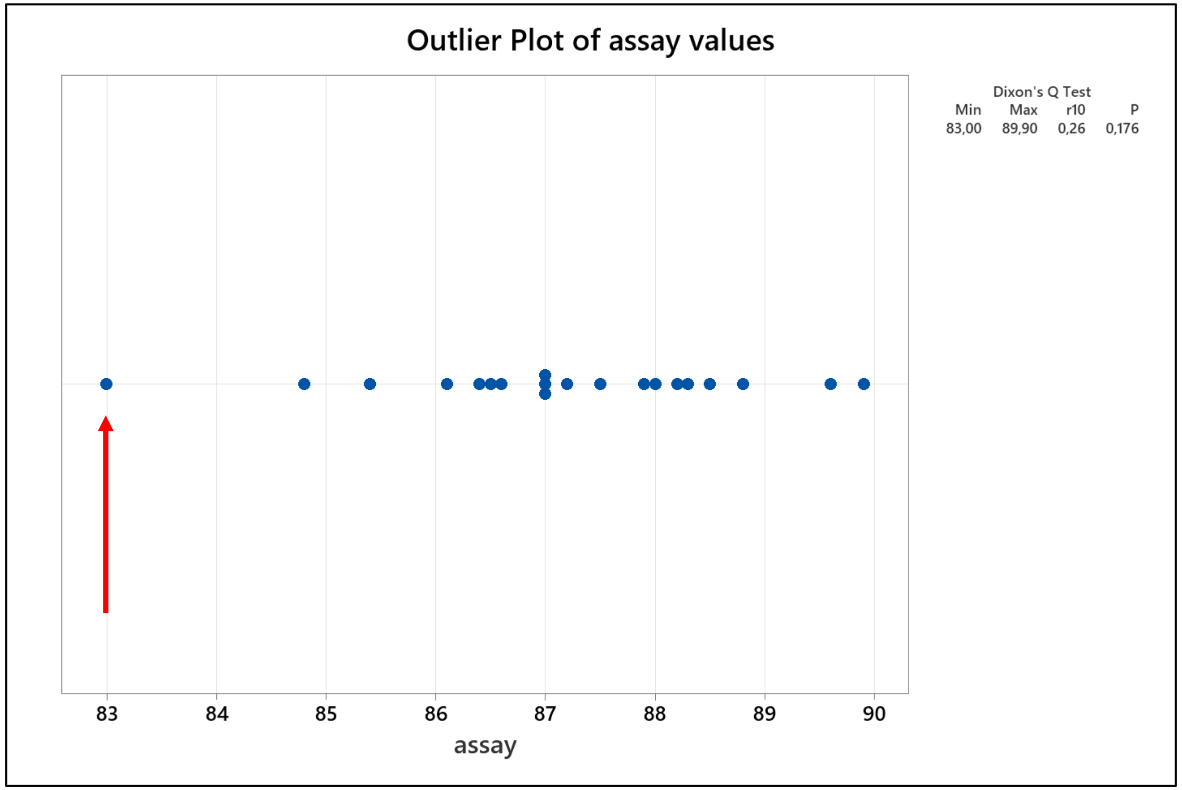
Method

Null hypothesis	All data values come from the same normal population
Alternative hypothesis	Smallest or largest data value is an outlier
Significance level	α = 0,01

Dixon's Q Test

Variable	N	Min	x[2]	x[N-1]	Max	r10	P
assay	20	83,000	84,800	89,600	89,900	0,26	0,176

x[i] denotes the ith smallest observation.



As *P-value* > 0.05 not enough evidence to reject H_0 →

* NOTE * No outlier at the 1% level of significance

OUTLIERS

Grubbs' Test

- It is usually recommended when it is not known whether the data includes outliers.
- Also, in this case a statistics, *G*, is calculated in which x_n can be the largest or smallest value.

$$G_n = \left| \frac{x_n - \bar{x}}{s} \right|$$

$\bar{x}$ is the sample mean and *s* the sample standard deviation.

- If the calculated value of *G* equals or exceeds the tabulated value, the outlier is rejected as an extreme value.

OUTLIERS

Grubbs' Test

- Recommended by the EPA

- Mentioned in the ISO 5725-2 guideline *"Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method"*

OUTLIERS

Grubbs' Test - Example

- Let's consider the assay values listed on the side: $\bar{x}$ = 87.2 s = 1.625
- 83.0, the lowest value in the series, is suspected to be the outlier.
- The value of G_n statistics is:

$$G_n = \left|\frac{83.0 - 87.2}{1.625}\right| = 2.58$$

- In the table of critical values for the Grubbs' Test we find that the critical value for n=20 is equal to 2.88 for α = 0.01 and therefore since 2.58 < 2.88, the value 83.0 is not considered an outlier at 99% confidence.

Ordered assay data
83,0
84,8
85,4
86,1
86,4
86,5
86,6
87,0
87,0
87,0
87,2
87,5
87,9
88,0
88,2
88,3
88,5
88,8
89,6
89,9

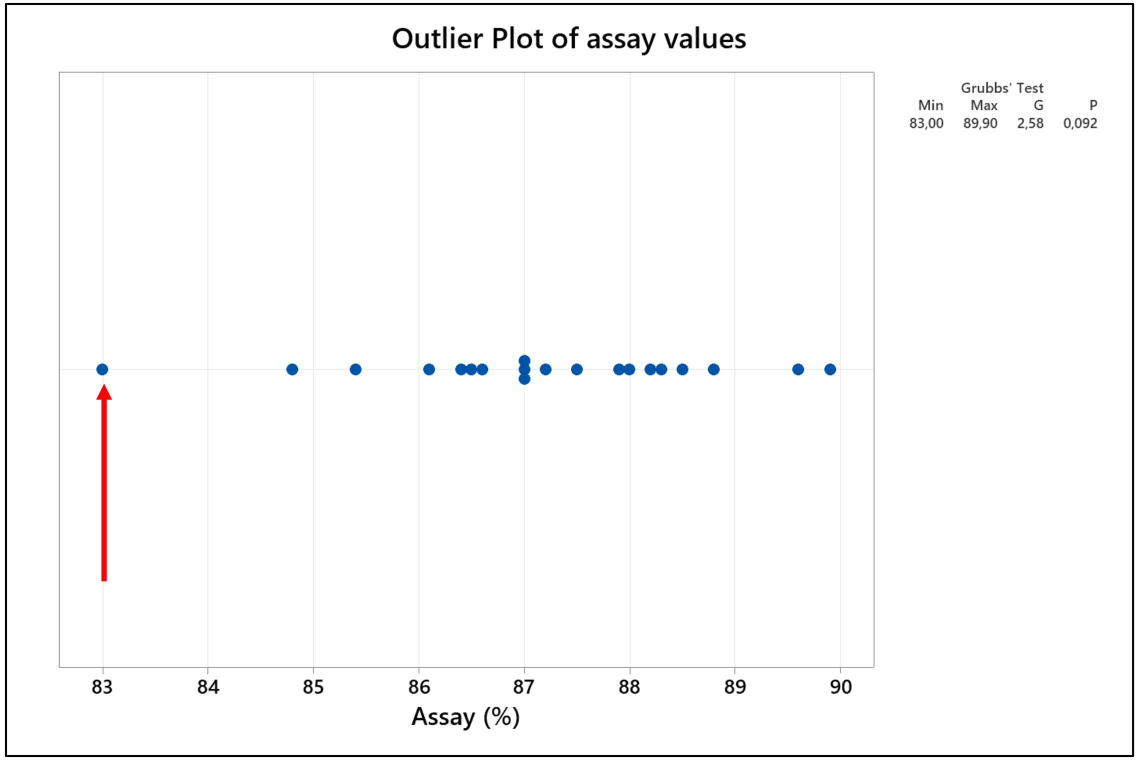
Grubbs' Test – Example using Minitab

Method

Null hypothesis	All data values come from the same normal population
Alternative hypothesis	Smallest or largest data value is an outlier
Significance level	α = 0,01

Grubbs' Test

Variable	N	Mean	StDev	Min	Max	G	P
assay	20	87,185	1,625	83,000	89,900	2,58	0,092



As *P-value* > 0.05 not enough evidence to reject H_0 →

* NOTE * No outlier at the 1% level of significance

OUTLIERS

1-Sample *t*-test

- Is a "difference test" which *allows us to determine whether the population average differs from a target value or a reference value* when we do not know the standard deviation of the population.

- It can be useful with chemical data to establish the anomalous nature of a piece of data because it does not belong to the population from which the other results come.

OUTLIERS

1-Sample t-test: Example using Minitab

Determines whether the population mean differs significantly from an assumed mean value. In this case it was verified whether *the assay value of 83.0 differs from the average of the population from which the other values were extracted.*

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
19	87,405	1,328	0,305	(86,765; 88,045)

μ: population mean of assay - 83

Test

Null hypothesis H_0: $\mu = 83$

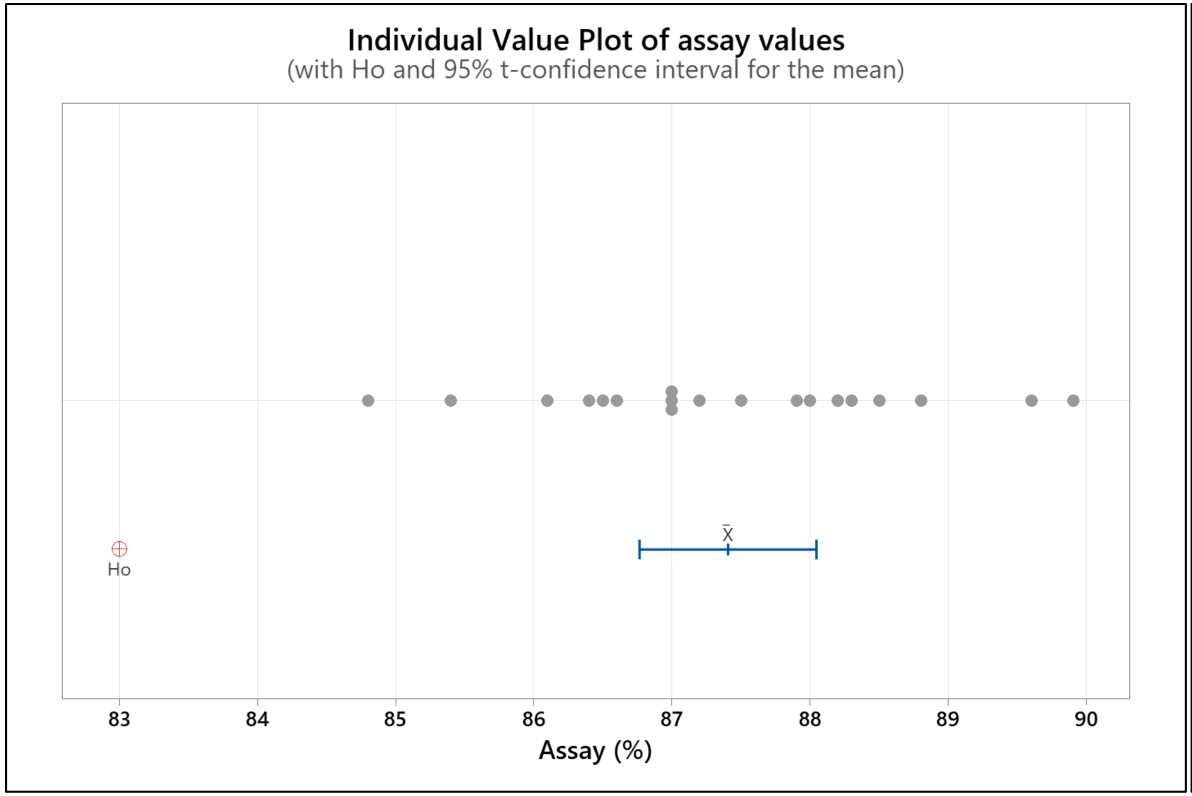
Alternative hypothesis H_1: $\mu \neq 83$

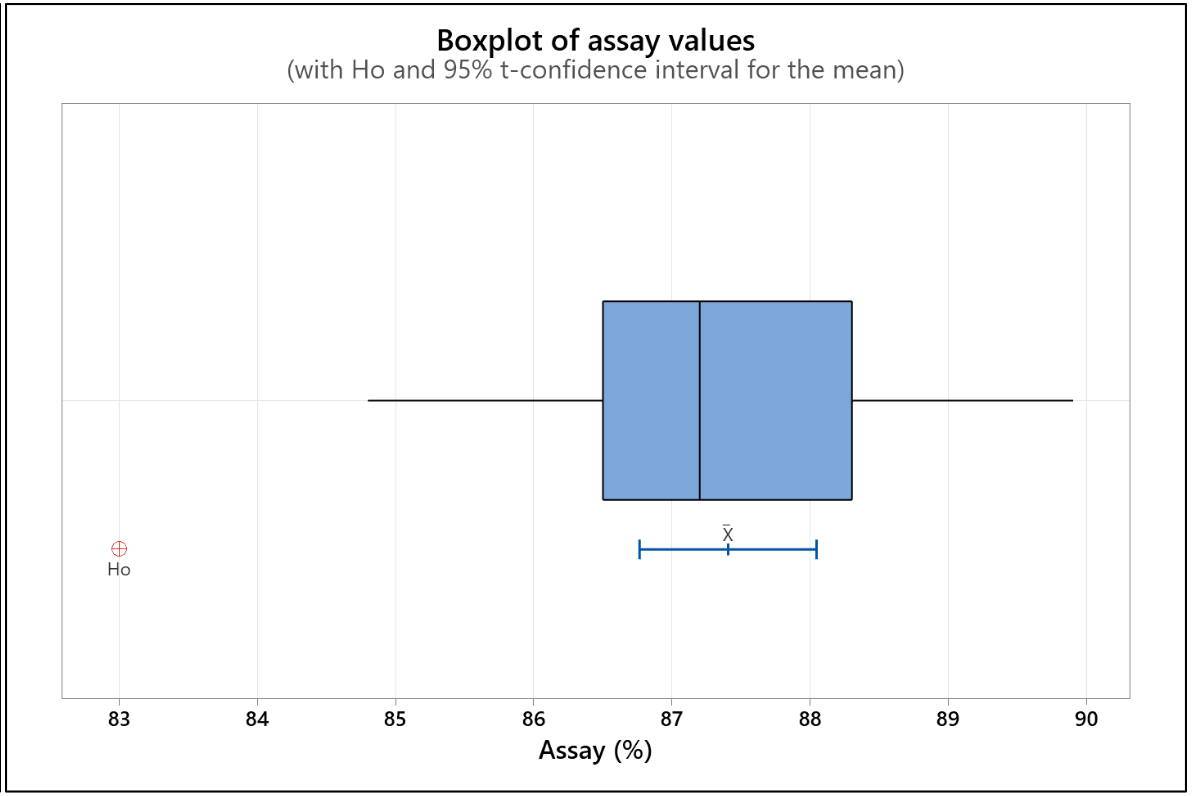
T-Value	P-Value
14,46	0,000

As *P-value* < 0.05 there is enough evidence to reject H_0

OUTLIERS

1-Sample t-test: Example using Minitab

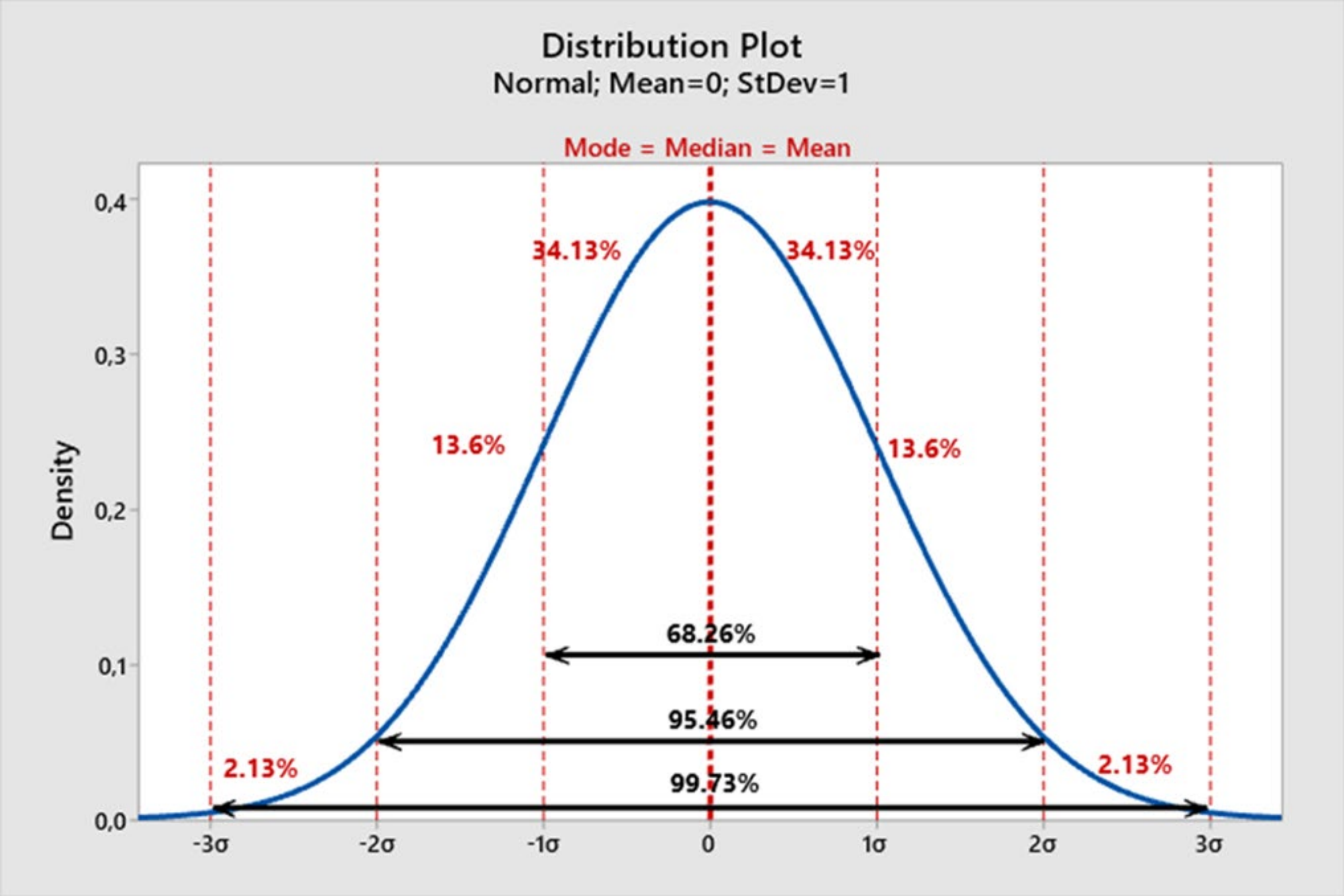




OUTLIERS

Z-Scores

- The Z-scores are based on the well-known property of the normal distribution that if X is distributed as $N(\mu, \sigma^2)$, then: $Z = (X - \mu)/\sigma$ is distributed as $N(0,1)$
- A Z-score tells us approximately where the point lies on a standard normal distribution.



Z-Scores

- One popular rule labels Z-scores that exceed 3 in absolute value as outliers, however Shiffler showed that the value of a Z-score is at most: $(n - 1)/\sqrt{n}$.
- This formula says that Z_{max} does not depend on the data values, but only on the number of observations, *n*.
- For very large samples, the idea of considering a Z-score > 3 as an outlier becomes less appropriate. In fact, in a sample of 10,000 observations, the largest possible Z-score according to this formula would be only about 3.16, which is very close to 3.

R.E. Shiffler, *Maximum Z Scores and Outliers*, American Statistician, 42 (1988) 79-80

OUTLIERS

Z-Scores

There’s a few drawbacks to using Z-Scores to detect outliers:

- it does not detect outliers when the sample size is small (n<12)
- since the arithmetic mean is sensitive to outliers in the data, Z-scores will be also. In case the dataset where contains a few very large values, this will skew the mean and standard deviation and thus, the Z-Scores will be skewed as well. More likely than not, only the largest data points will be marked as outliers, and while this is true, the other, smaller scale outliers may not be flagged.

Modified Z-Scores addresses this issue replacing the mean with the median to calculate a score

Modified Z-Scores

The use of *resistant estimators* (MAD, *i.e.*, the Median of the Absolute Deviations about the median) leads to Modified Z-scores, M_i, defined as follows:

$$M_i = \frac{0.6745 \times (x_i - \tilde{x})}{MAD}$$

where:

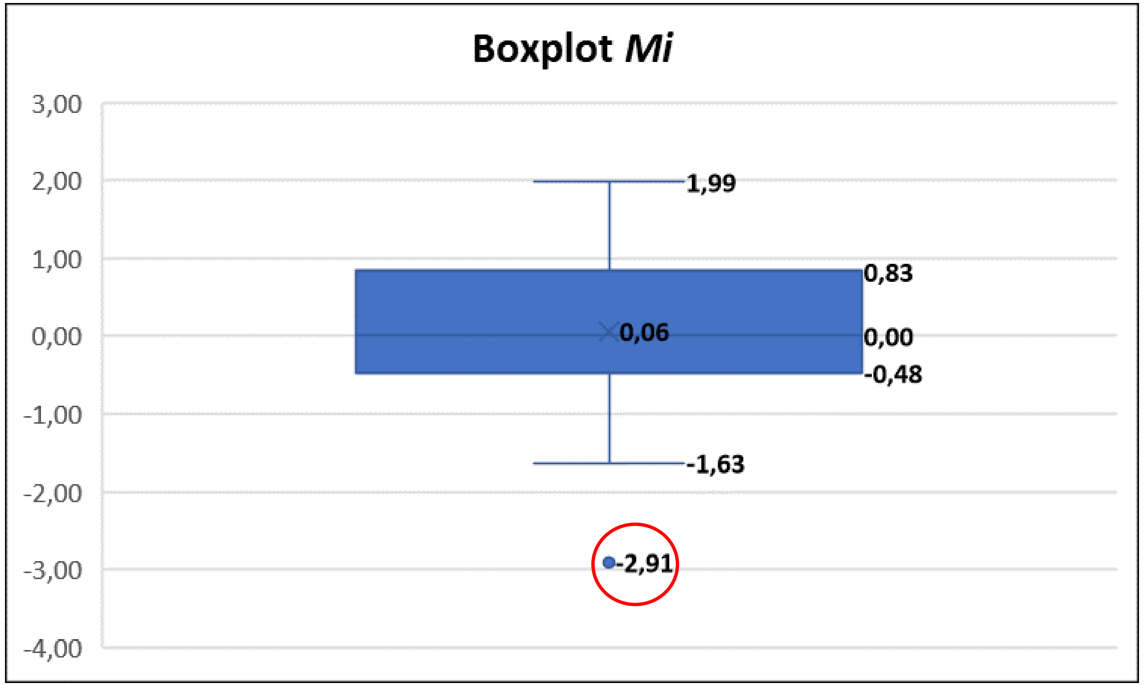
$$MAD = median_i\{| x_i - \tilde{x} |\}$$

Iglewicz and Hoaglin suggest to label observations as outliers when $|Mi| > 3.5$

B. Iglewicz, D.C. Hoaglin, *How to Detect and Handle Outliers*, American Society for Quality Control, Vol. 16, (1993)

OUTLIERS

In this example the above criterion is not satisfied, however….



ID	Assay data	Ordered Assay data	Ordered |xi-median|	*Mi*
1	86,6	**83,0**	4,10	**-2,91**
2	88,2	84,8	2,30	-1,63
3	86,4	85,4	1,70	-1,21
4	88,3	86,1	1,00	-0,71
5	85,4	86,4	0,70	-0,50
6	89,9	86,5	0,60	-0,43
7	84,8	86,6	0,50	-0,35
8	87,0	87,0	0,10	-0,07
9	89,6	87,0	0,10	-0,07
10	88,8	87,0	0,10	-0,07
11	86,1	87,2	0,10	0,07
12	87,9	87,5	0,40	0,28
13	83,0	87,9	0,80	0,57
14	88,5	88,0	0,90	0,64
15	87,2	88,2	1,10	0,78
16	88,0	88,3	1,20	0,85
17	86,5	88,5	1,40	0,99
18	87,5	88,8	1,70	1,21
19	87,0	89,6	2,50	1,77
20	87,0	89,9	2,80	1,99
	Median	87,10	0,95	

Before we proceed, let me briefly sidetrack

Cochran's Test

Cochran's Test

- It is a slightly different test from the previous ones as it does not refer to the data themselves but to the variability of series of dates compared to each other.
- It is used to decide whether, in a series of variance estimates (or standard deviations), one is significantly greater than the others with which it should be comparable.

W.G. Cochran, *The distribution of the largest of a set of estimated variances as a fraction of their total*, Annals of Human Genetics, 1941, 11(1), 47–52

OUTLIERS

Cochran's Test

- The data groups on which the variances (or standard deviations) to be compared are calculated must contain the same number of replications (*balanced design*)
- This test, like the previous ones, also assumes that each individual data series is normally distributed
- Because it tests only the highest value in a series of variances (or standard deviations), it is a *one-sided outlier test*.

W.G. Cochran, *The distribution of the largest of a set of estimated variances as a fraction of their total*, Annals of Human Genetics, 1941, 11(1), 47–52

OUTLIERS

Cochran's Test

- Given a set of *p* standard deviations s_i, all calculated on the same number of measurements *n*, the Cochran test is based on the calculation of the ***C*** statistic thus defined:

$$C = \frac{s_{max}^2}{\sum_{i=1}^{p} s_i^2}$$

where s_{max} is the largest standard deviation in the set.

- The ***C*** value calculated in this way is compared with a tabled value and if it is greater than the latter it cannot be assumed that there is homogeneity of the deviations.

OUTLIERS

Cochran's Test – Example using Excel

- Let's consider, for example, the case of an inter-laboratory test such as the one summarized below:

	LAB 1	LAB 2	LAB 3	LAB4	LAB5	LAB6			
Rep 1	100,04	100,05	99,93	100,25	100,47	99,83			
Rep 2	99,97	99,88	100,11	99,91	99,25	99,49			
Rep 3	100,03	100,1	100,14	99,97	100,46	100,33			
Rep 4	100,09	99,95	99,51	100	101,34	100,28			
Rep 5	100,02	100	99,95	100,06	98,29	99,51			
Rep 6	99,99	99,98	99,9	99,9	99,83	100,2			
Mean =	100,02	99,99	99,92	100,02	99,94	99,94			
Std. Dev. =	0,04	0,08	0,23	0,13	1,07	0,38			
Variance =	0,00	0,01	0,05	0,02	**1,14**	0,15		Sum =	1,37
Cochran's C =	**0,84**								

OUTLIERS

Cochran's Test

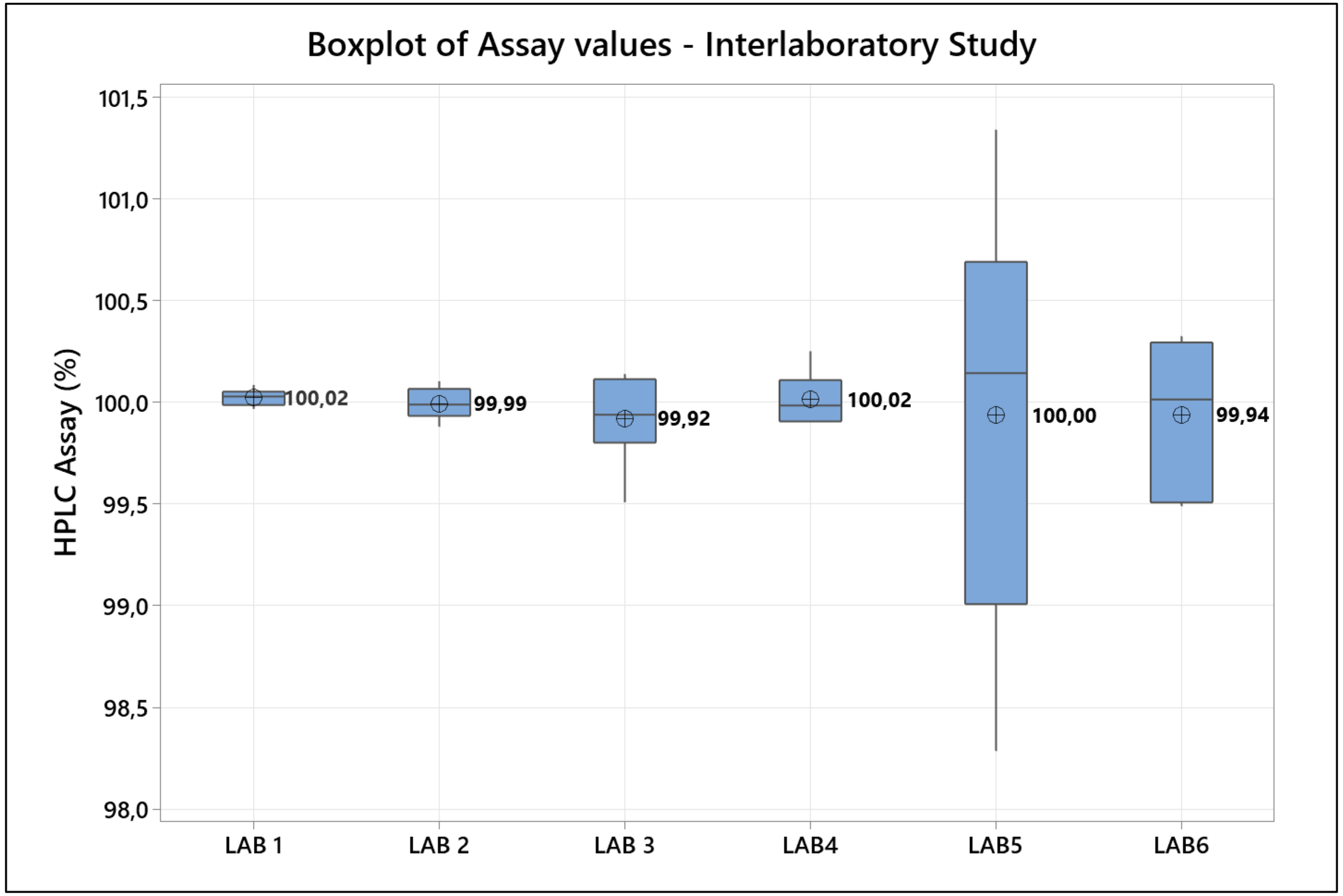
- The ***C*** statistic calculated is equal to 0.84
- In the reference tables* for N = 6 and dof = 5 at α = 0.05, the upper critical limit for ***C*** is 0.44
- Since $C_{calc} > C_{tab}$, the null hypothesis of equal variances can be rejected with a confidence of 95% and therefore the data relating to LAB 5 removed from the comparison as they are outliers
- According to some guidelines (*e.g.*, ISO 5725-2) Cochran's test can be repeated until excessive variances disappear, but this is inadvisable

*S. Kokoska, C. Nevison, *Statistical Tables and Formulae*, Springer Verlag, 1988, p. 74

OUTLIERS

Cochran's Test

Using a graphical method also allows you to determine which standard deviation (or variance) is extreme.



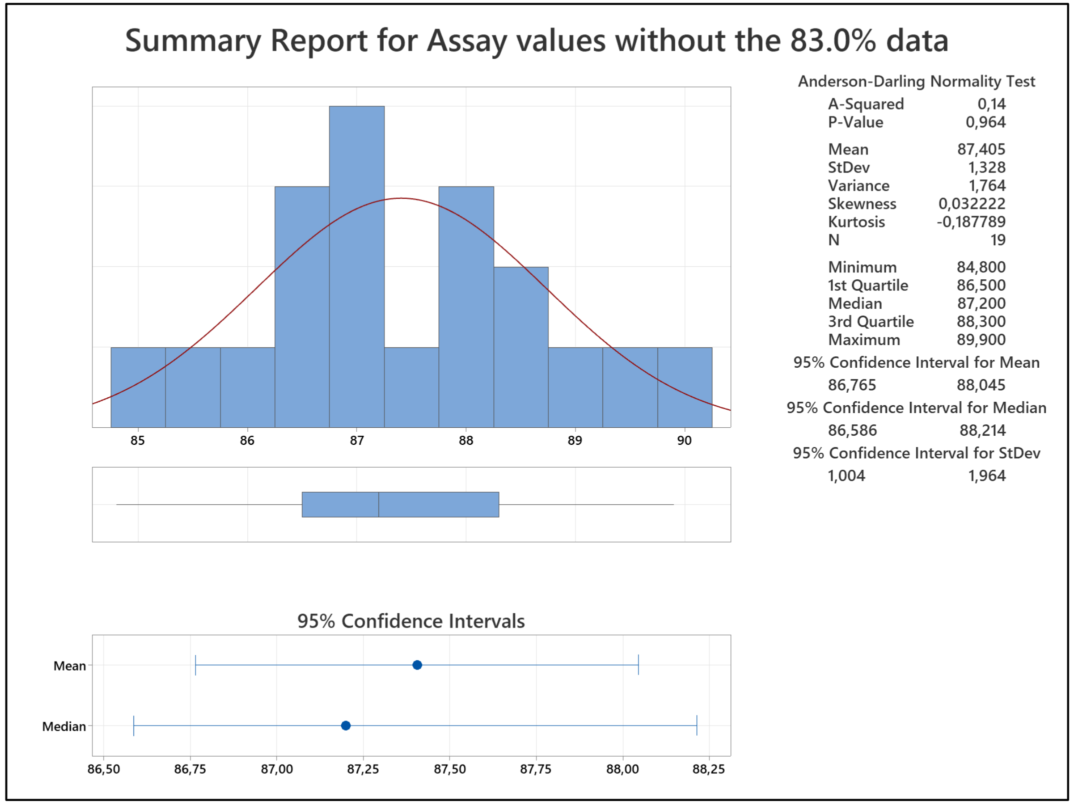
Now let’s go back to our original discussion

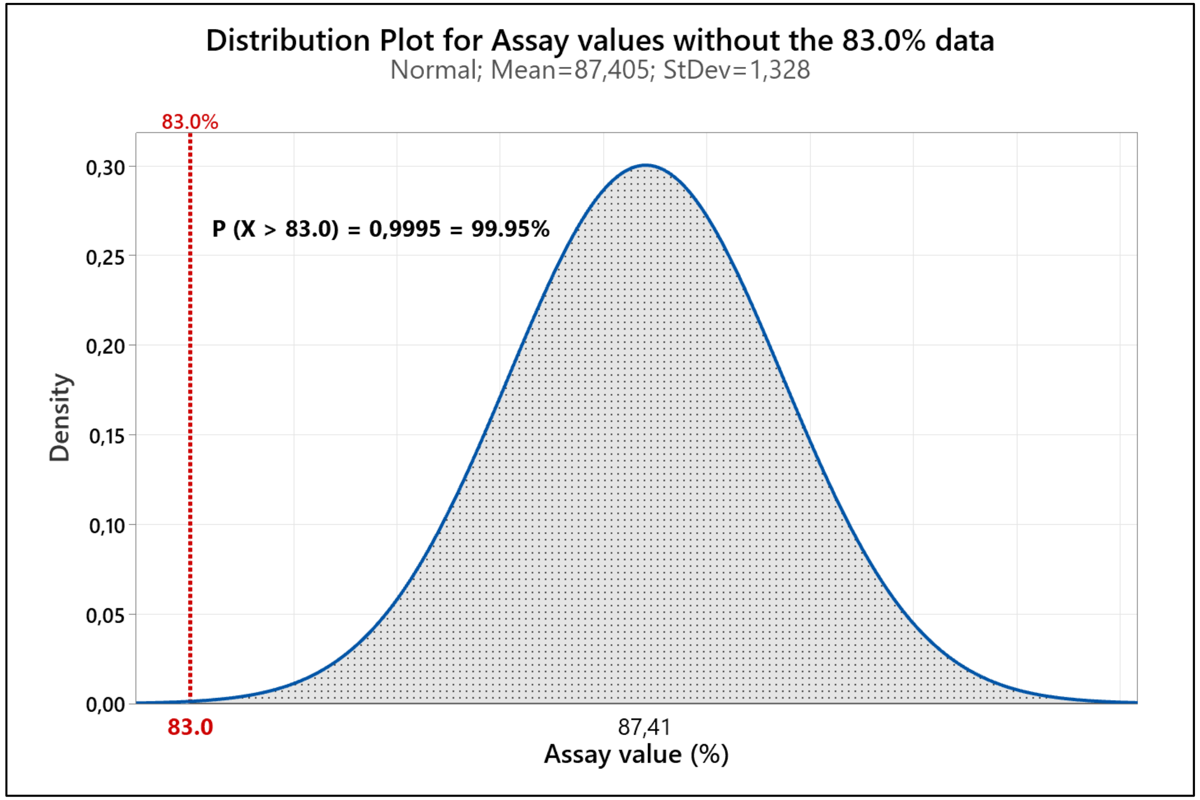
OUTLIERS

In summary, based on statistical tests, the value of 83.0 is an outlier when evaluated using the 1-sample t-test, but is not an outlier when using the Dixon, Grubbs, and modified Z-Score tests. So:

is this assay value of 83.0% an outlier or not?

OUTLIERS





Given the other values, the probability of an outcome of 83.0% is less than 0.05%, therefore….

OUTLIERS

STATISTICAL TESTS FOR OUTLIERS IN REGRESSION ANALYSIS

- To verify whether the outliers shown by fitted line/residual plots are influential or not, metrics are used that are based on the concept of ***distance*** (***Cook's Distance***, ***DFFITS***, *etc.*)

- Particular attention should be paid in the case of two nearly coincident influential outliers as the described methods may be ineffective.

OUTLIERS

In summary:

- From everything we've seen so far, ***there is no single tool that allows us to say whether a given observation is an outlier or not***.

- However, if the joint use of multiple tools (*e.g.*, boxplots, modified *Z* scores, …, fitted line plots, residual plots, *etc.*) gives mutually agreeable results, there is evidence of different behavior for that or those observations.

How to handle Outliers

What Bolton and Bon say in this regard is absolutely clear:

In general, aberrant observation ***should not be arbitrarily discarded*** *only because they look too large or too small, perhaps only for the reason of making the experimental data look "better".*

Therefore, ***the question of what to do with outliers is not an easy one to answer****. The error of either incorrectly including or excluding outlying observations will distort the validity of interpretation and conclusions of the experiment.*

S. Bolton, C. Bon, *Pharmaceutical Statistics – Practical and Clinical Applications,* CRC Press (2010)

How to handle Outliers

- The Authors conclude that, in the end, the choice that must be made is based on an overall "***good judgment***" that considers a deep knowledge of the process under study and the statistical consequences.
- ***A good general strategy is to analyze the conclusions by considering and omitting the outlier and studying/comparing the two situations.***

➢ ***Whatever the final decision is made, this must be technically argued in depth and not just in words, but based above all on an in-depth analysis of the data that uses valid and correctly applied statistical criteria.***

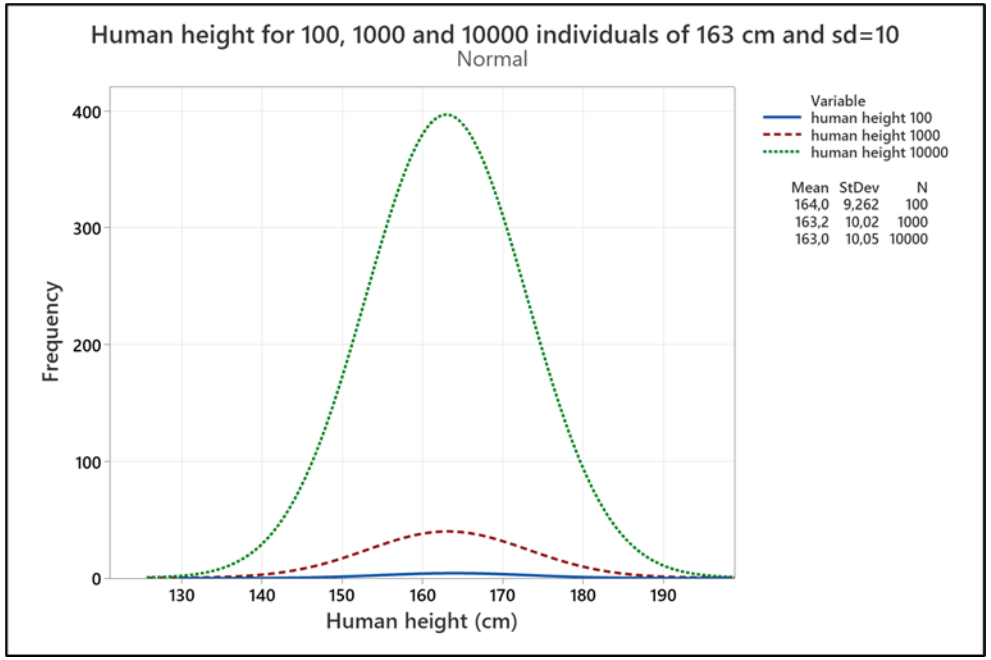
OUTLIERS

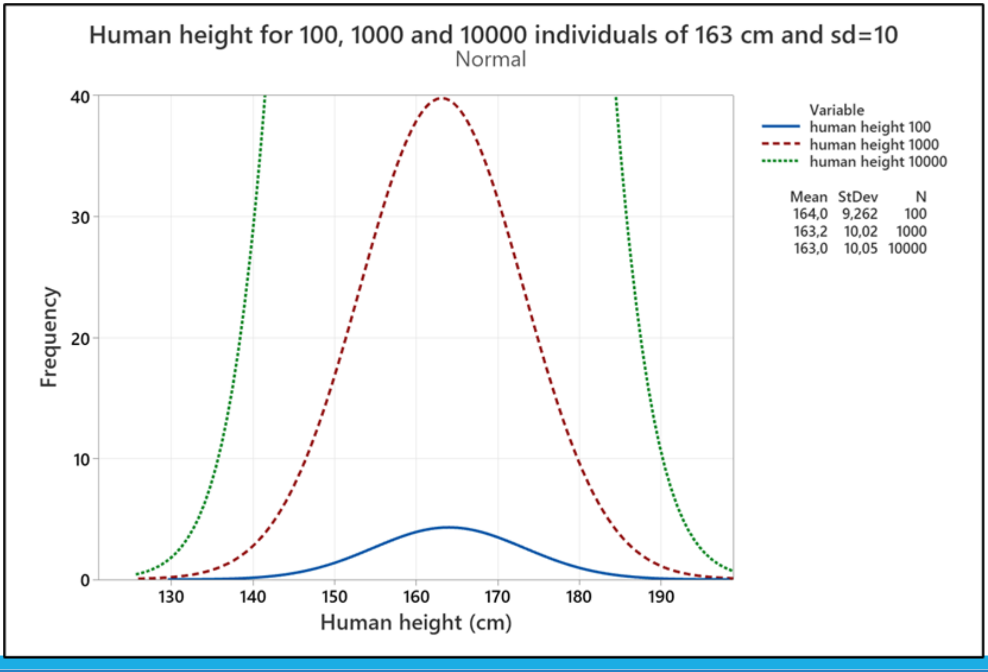
- **How to handle Outliers**

WARNING !

It may happen that an “unexpected data" is not an outlier, but only a "new data" perhaps resulting from a larger sample base as in the example opposite.

It is therefore always good practice to also consider this aspect ☺





OUT OF TREND

The term "trend" occurs very often in many regulatory documents!

Below are some examples

OUT OF TREND

- An ongoing program to collect and analyze product and process data that relate to product quality must be established and … the data should be statistically trended and reviewed by trained personnel.

FDA Guidance for Industry – *Process Validation: General Principles and Practices* (January 2011)

- 5.29. Manufacturers should monitor product quality to ensure that a state of control is maintained throughout the product lifecycle with the relevant process trends evaluated.

EU Guidelines for GMP – *Annex 15 : Qualification and Validation*, Eudralex, Volume 4 (March 2015)

OUT OF TREND

- Although the CGMP regulations (§ 211.180(e)) require product review on at least an annual basis, **a quality systems approach calls for trending on a more frequent basis** as determined by risk. Trending enables the detection of potential problems as early as possible to plan corrective and preventive actions. Another important concept of modern quality systems is the use of trending to examine processes as a whole; this is consistent with the annual review approach. Trending analyses can help focus internal audits (see IV.D.2.).

FDA Guidance for Industry - *Quality Systems Approach to Pharmaceutical CGMP Regulations* (September 2006)

OUT OF TREND

- Process performance can be monitored to ensure that it is working as anticipated to deliver product quality attributes as predicted by the design space. This monitoring could include trend analysis of the manufacturing process as additional experience is gained during routine manufacture.

ICH Guideline Q8(R2), *Pharmaceutical Development* – August 2009

- The pharmaceutical company should have a system for implementing corrective actions and preventive actions resulting from the investigation of complaints, product rejections, non-conformances, recalls, deviations, audits, regulatory inspections and findings, and trends from process performance and product quality monitoring.

ICH Guideline Q10, *Pharmaceutical Quality System* – June 2008

OUT OF TREND

- The quality control unit should provide routine oversight of near-term (*e.g.*, daily, weekly, monthly, quarterly) and long-term trends in environmental and personnel monitoring data. Trend reports should include data generated by location, shift, room, operator, or other parameters …. Written procedures should define the system whereby the most responsible managers are regularly informed and updated on trends and investigations.

FDA Guidance for Industry - *Sterile Drug Products Produced by Aseptic Processing - Current Good Manufacturing Practice* (2004)

But what does "trend" really mean?

Too often, in fact, this concept is left to intuition!

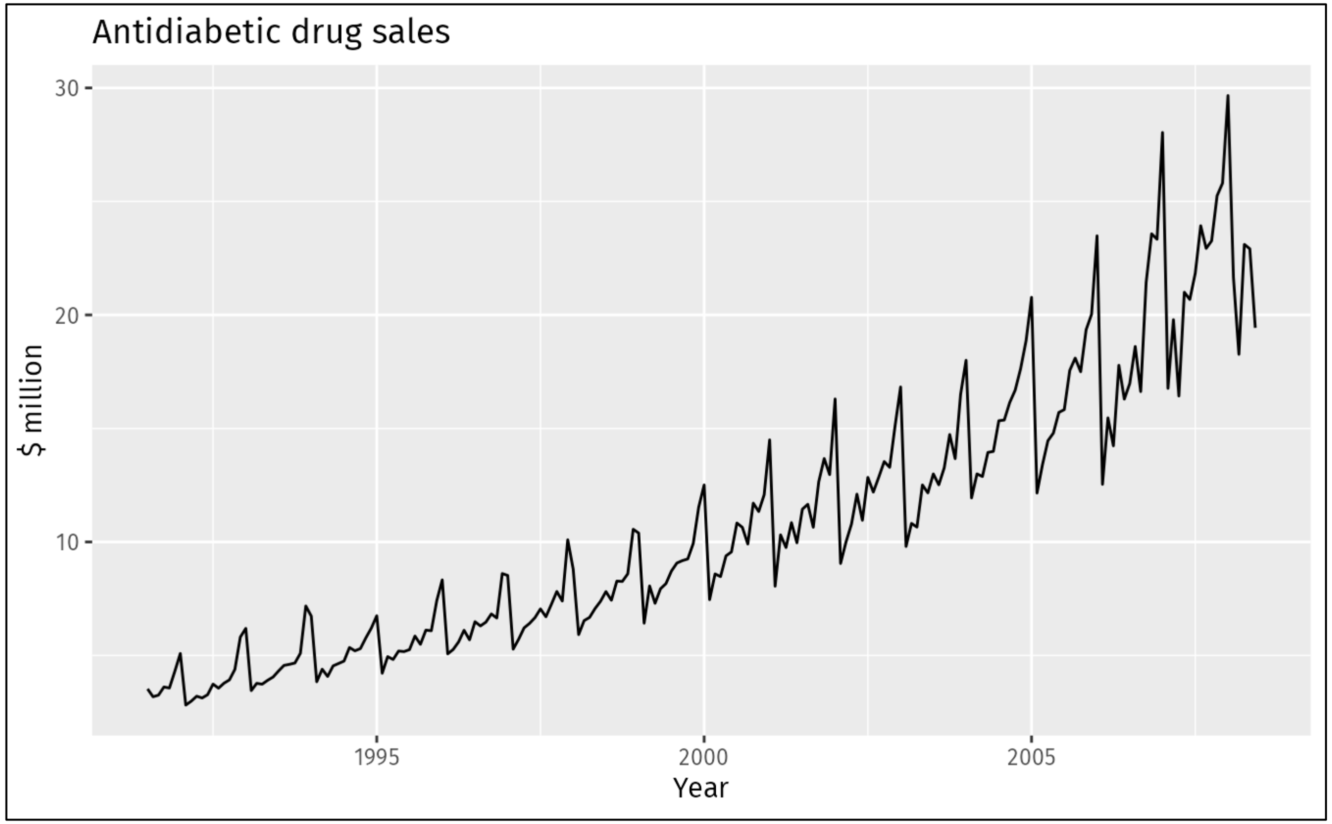
OUT OF TREND

Professor Hyndman clearly states that:

« A ***trend*** exists when there is a **long-term increase or decrease in the data.** It does not have to be linear.

Sometimes we will refer to a trend as "changing direction", when it might go from an increasing trend to a decreasing trend.

There is a trend in the antidiabetic drug sales shown in Figure 2.2 »



R.J. Hyndman, G. Athanasopoulos, *Forecasting – Principles and Practice,* OTexts (2018) 31

OUT OF TREND

Professor Montgomery states:

« A ***trend***, or **continuous movement in one direction**, is shown on the control chart of Figure 6.11. Trends are usually due to a gradual wearing out or deterioration of a tool or some other critical process component. In chemical processes they often occurs because of settling or separation of the components of a mixture….»

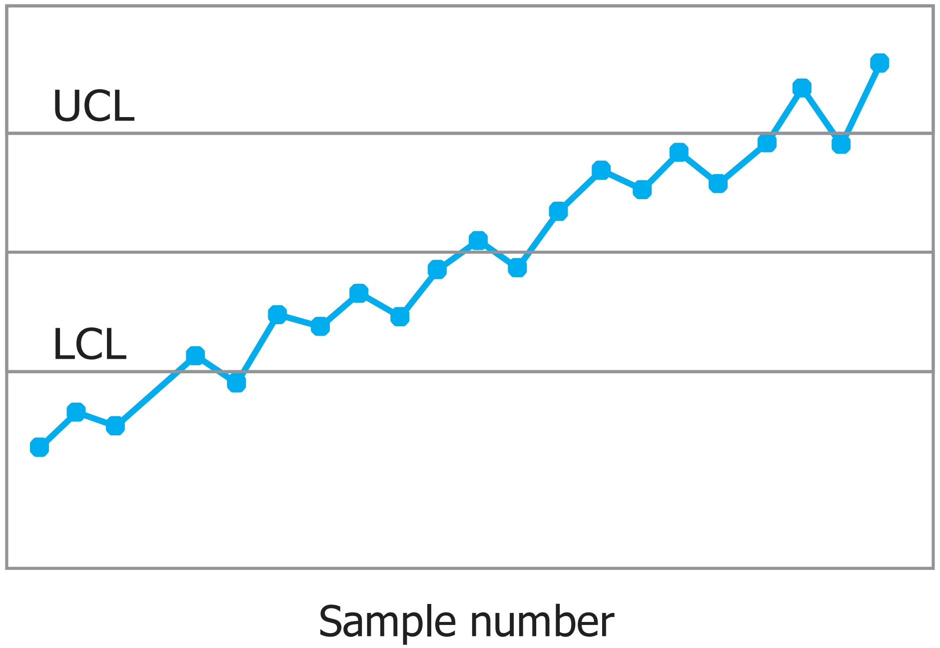


FIGURE 6.11 A trend in process level.

D.G. Montgomery, *Statistical Quality Control – A Modern Introduction, Wiley* (2013) 252-253

OUT OF TREND

Still **Professor Montgomery**, but elsewhere in the book:

« … an assignable cause can result in many different types of shifts in the process parameters. For example, the mean could shift instantaneously to a new value and remain there (this is sometimes called a **sustained shift**); or it could shift abruptly; but the assignable cause could be short-lived and the mean could then return to its nominal or in-control value; or the assignable cause could result in a ***steady drift or trend in the value of the mean***….»

D.G. Montgomery, *Statistical Quality Control – A Modern Introduction, Wiley* (2013) 191

OUT OF TREND

In the **Statistical Quality Control Handbook by Western Electric Co. Inc.**, it is reported :

« A trend is defined as continuous movement up or down; x's on one side of the chart followed by x's on the other; ***a long series of points*** without a change of direction »

Western Electric Co. Inc., *Statistical Quality Control Handbook, 10th Printing, Delmar* (1984), page 177

OUT OF TREND

- from all this (*i.e.*, long-term, continued, sustained shift, etc.) emerges a first and very important characteristic that must be defined to establish or not the presence of a trend and that is

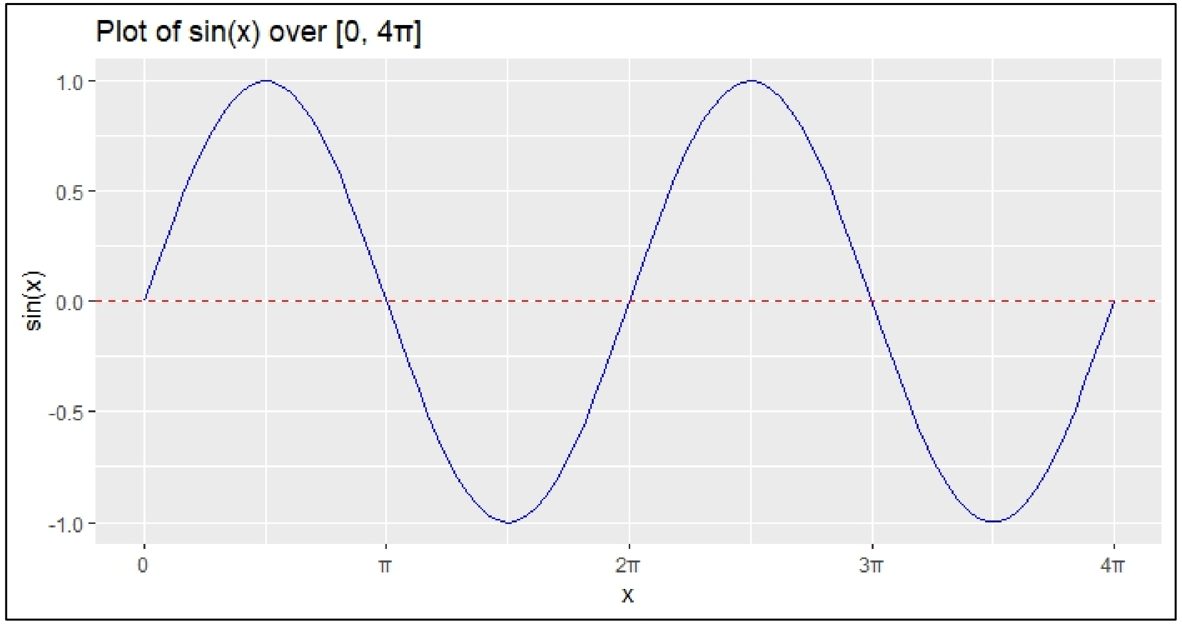
time !

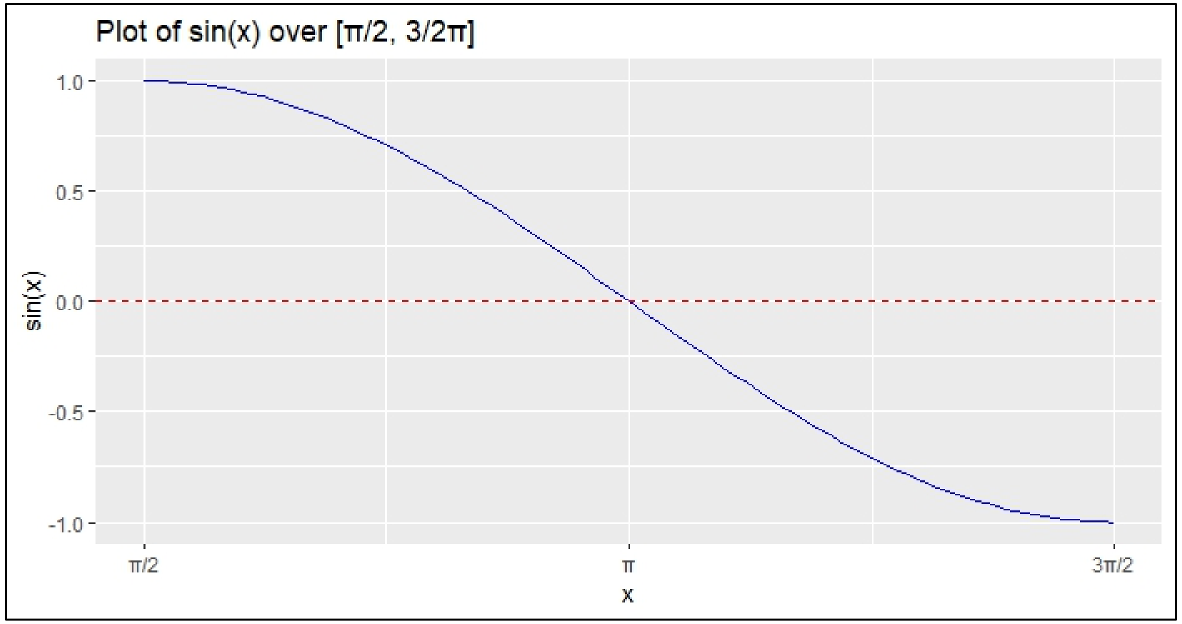
- **Depending on the time frame selected, we may or may not be in the presence of a trend !**

- An intuitive example is that of the *sin(x)* function

OUT OF TREND

the function is the same in both cases, but...





cyclicality, but no specific trend

a decreasing trend

OUT OF TREND

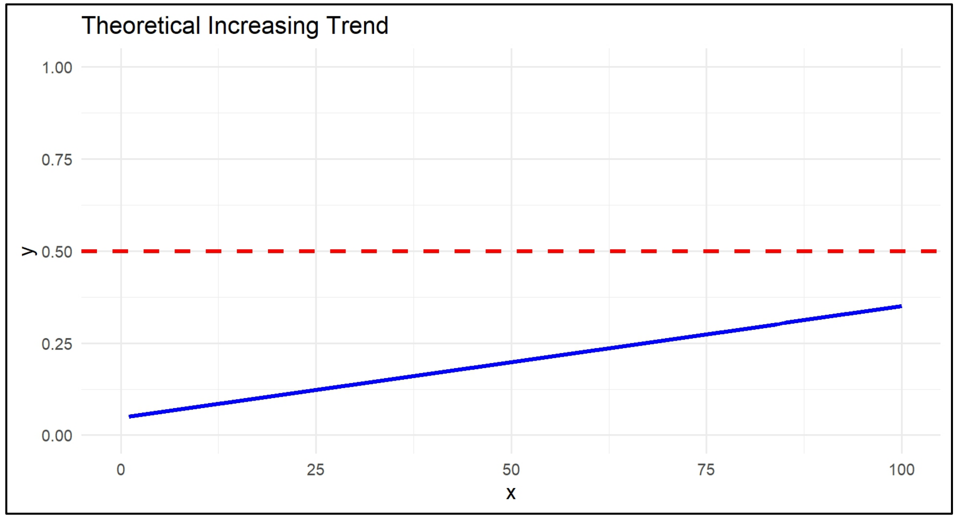
- alongside the time frame, there is also another important aspect that must be defined, namely the

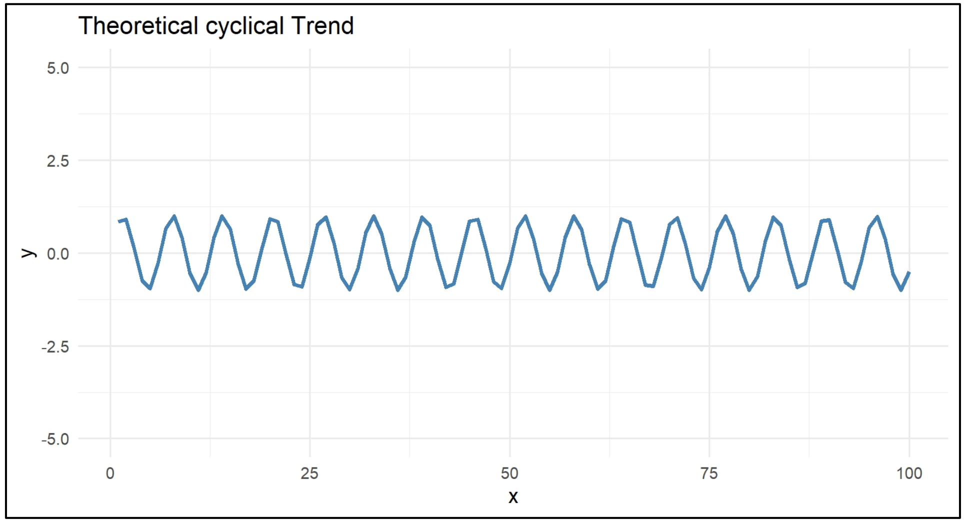
reference limits

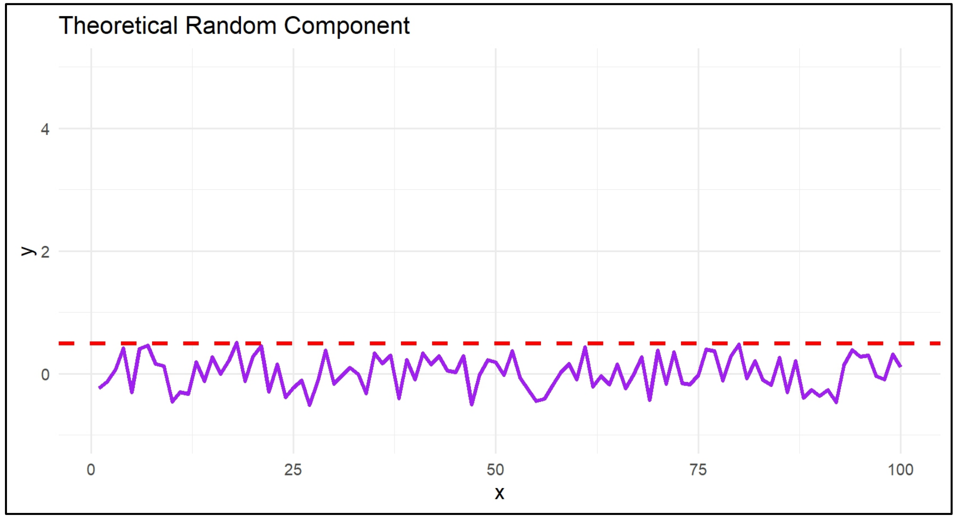
- a trend, in fact, may not be practically significant if it does not give evidence of approaching pre-established reference limits and, potentially, exceeding them.

OUT OF TREND

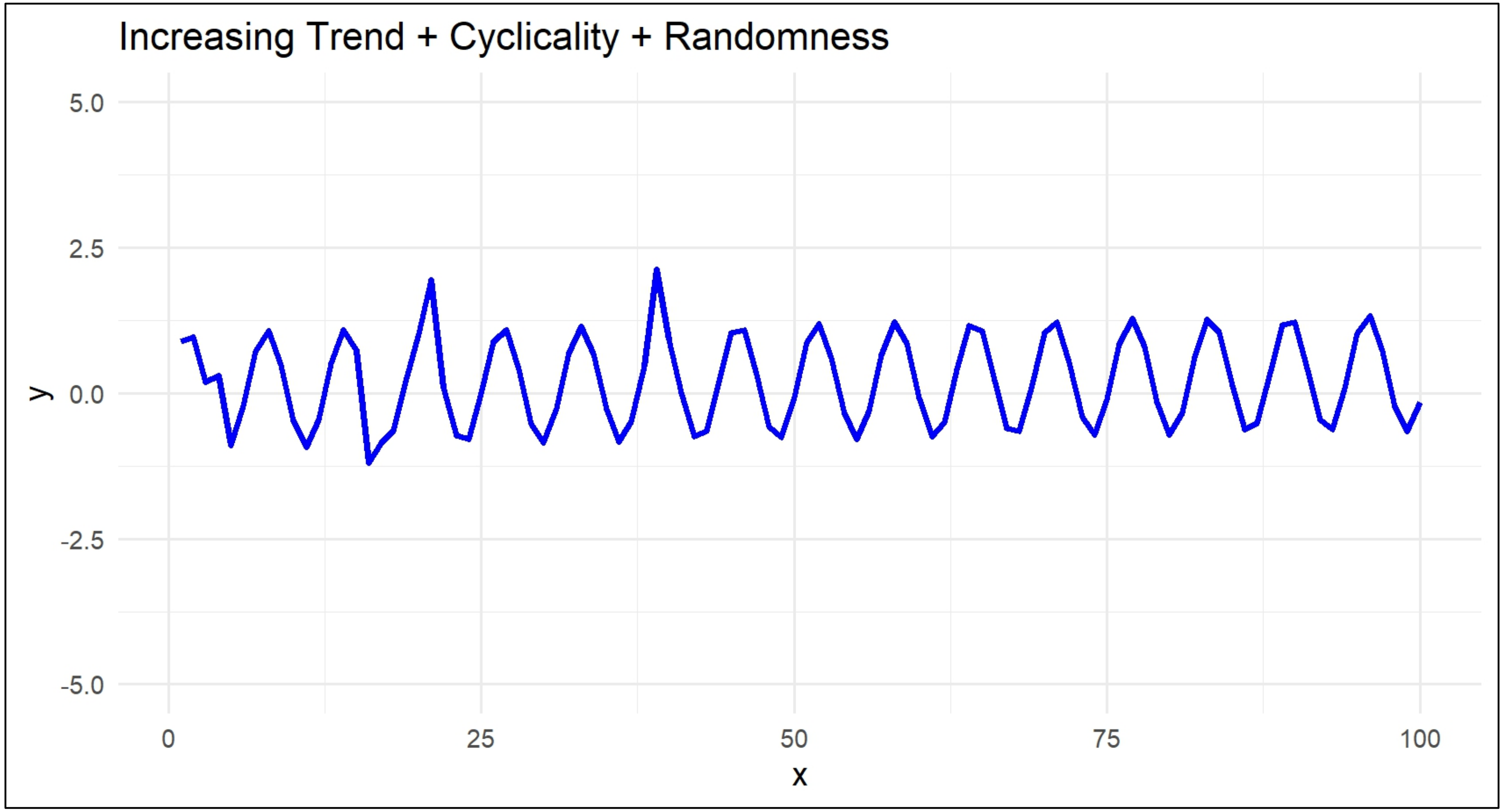
In practice, things are more complicated. In fact, a ***time series*** is made up of different components:







OUT OF TREND



Trend is therefore a type of pattern in data that must be assessed

How can we handle all this in a simple way?

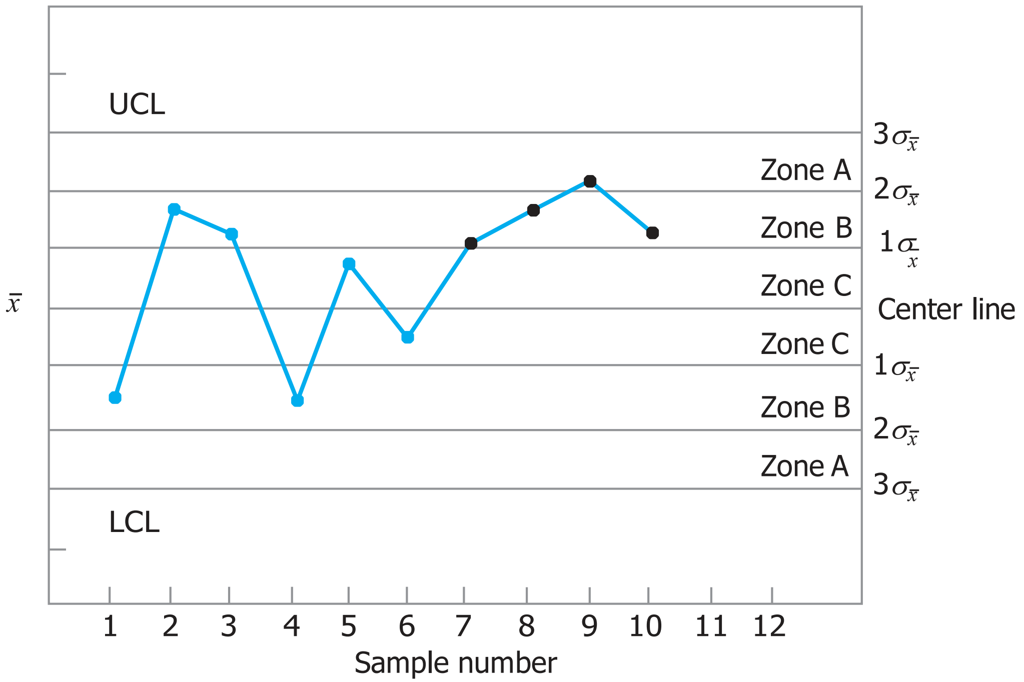
OUT OF TREND

- The ***Western Electric Statistical Quality Control Handbook*** (1956) suggests a set of decision rules for detecting « nonrandom patterns » (and trend among them) on control charts.
- These rules, sometimes called ***zone rules*** for control charts, apply to one side of the center line at a time.

OUT OF TREND

Western Electric Rules:

1. One or more points outside of the control limits
2. Two of three consecutive points outside the two-sigma warning limits but still inside the control limits
3. Four of five consecutive points beyond the one-sigma limits
4. A run of eight consecutive points on one side of the center line



Example with violation of Rule 3

D.G. Montgomery, *Statistical Quality Control – A Modern Introduction, Wiley* (2013) 191

OUT OF TREND

Subsequently, further rules were added to the first four:

5. Six points in a row steadily increasing or decreasing
6. Fifteen points in a row in zone C (both above and below the center line)
7. Fourteen points in a row alternating up and down
8. Eight points in a row on both sides of the center line with none in zone C
9. An unusual or nonrandom pattern in the data
10. One or more points near a warning or control limit

L.S. Nelson, *The Shewhart Control Chart—Tests for Special Causes,* J. Quality Technology, Vol. 16, 4 (*1984*) 237-239

OUT OF TREND

- ***In general, a lot of care should be exercised when using several decision rules simultaneously.***
- ***Some of the individual Western Electric rules are particularly troublesome*.**
- An illustration is the rule of several (usually seven or eight) consecutive points that either increase or decrease. **This rule is very ineffective in detecting a trend**, the situation for which it was designed. It does, however, **greatly increase the false alarm rate***

R.B. Davis, W.H. Woodall, *Performance of the Control Chart Trend Rule under Linear Shift,* J. Quality Technology, Vol. 20, 4 (*1988*) 260-262

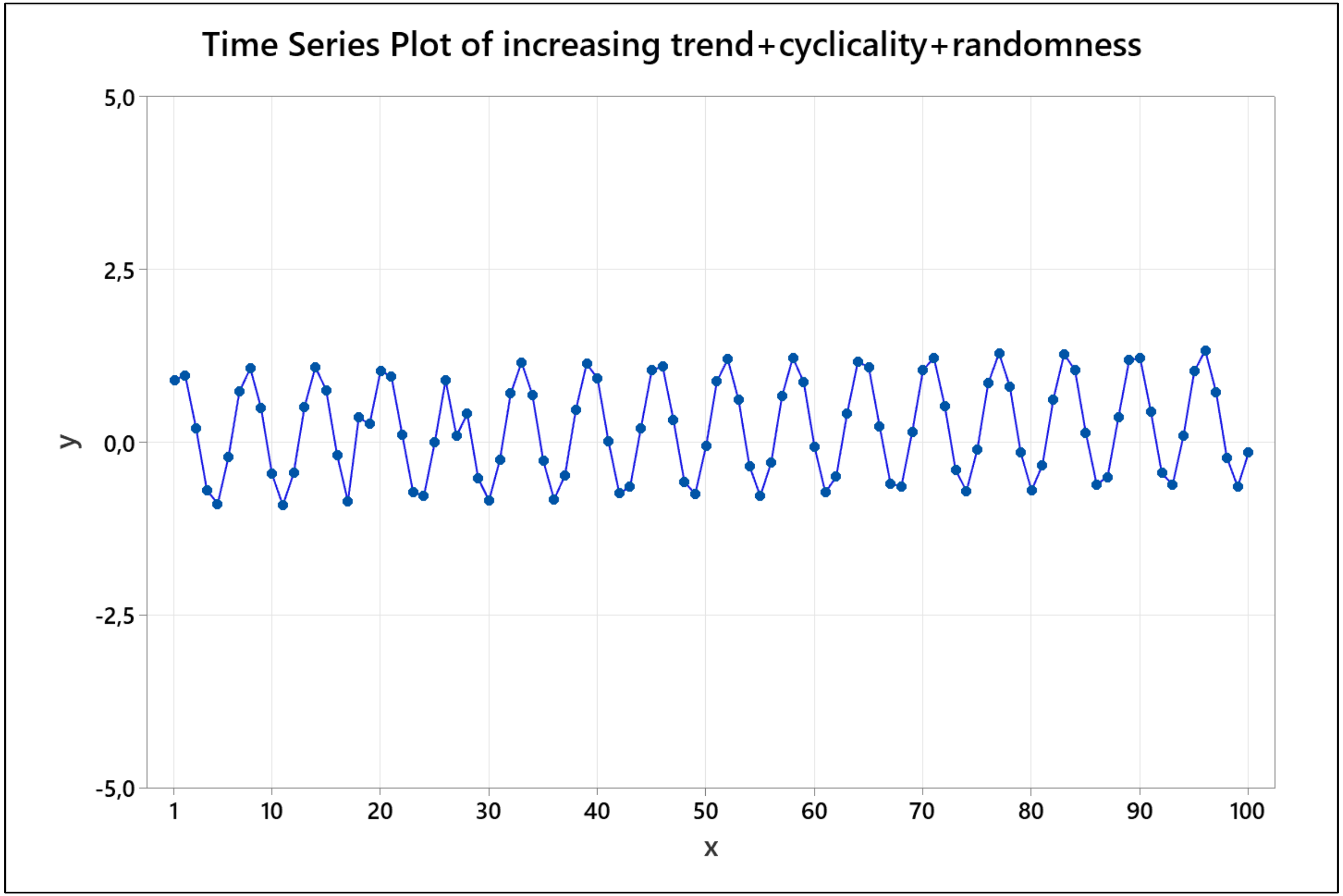
So, what can be done easily?

We can use conventional (e.g., I-MR) and special (e.g., MA, EWMA) control charts!

OUT OF TREND

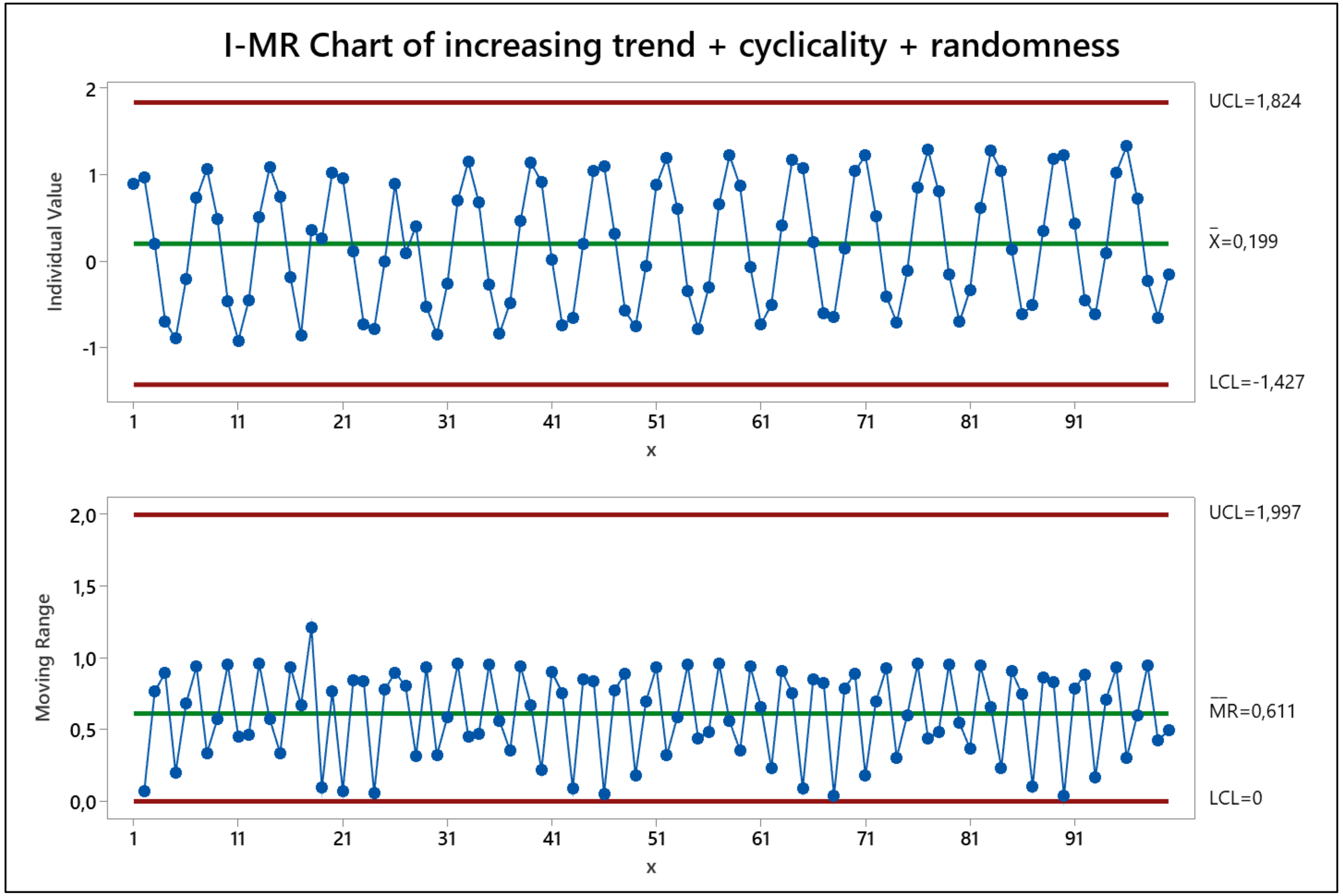
Let's go back to the graph and the data relating to *“increasing trend + cyclicality + randomness”* *seen before.*

The time series plot alongside does not seem to highlight any trend.



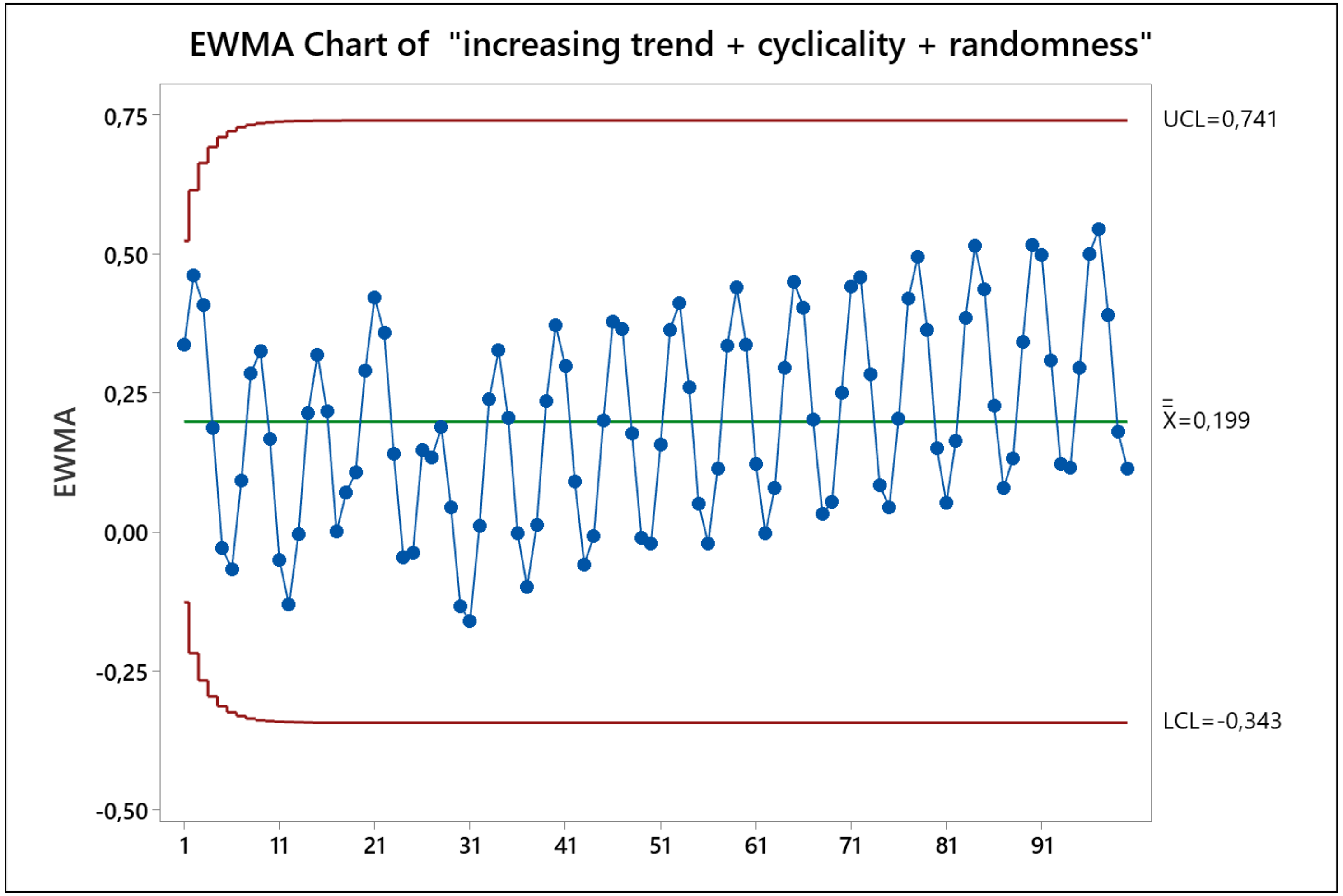
OUT OF TREND

The traditional *I-MR* (or Individual-Moving Range) Shewhart control chart doesn't seem to highlight anything either.



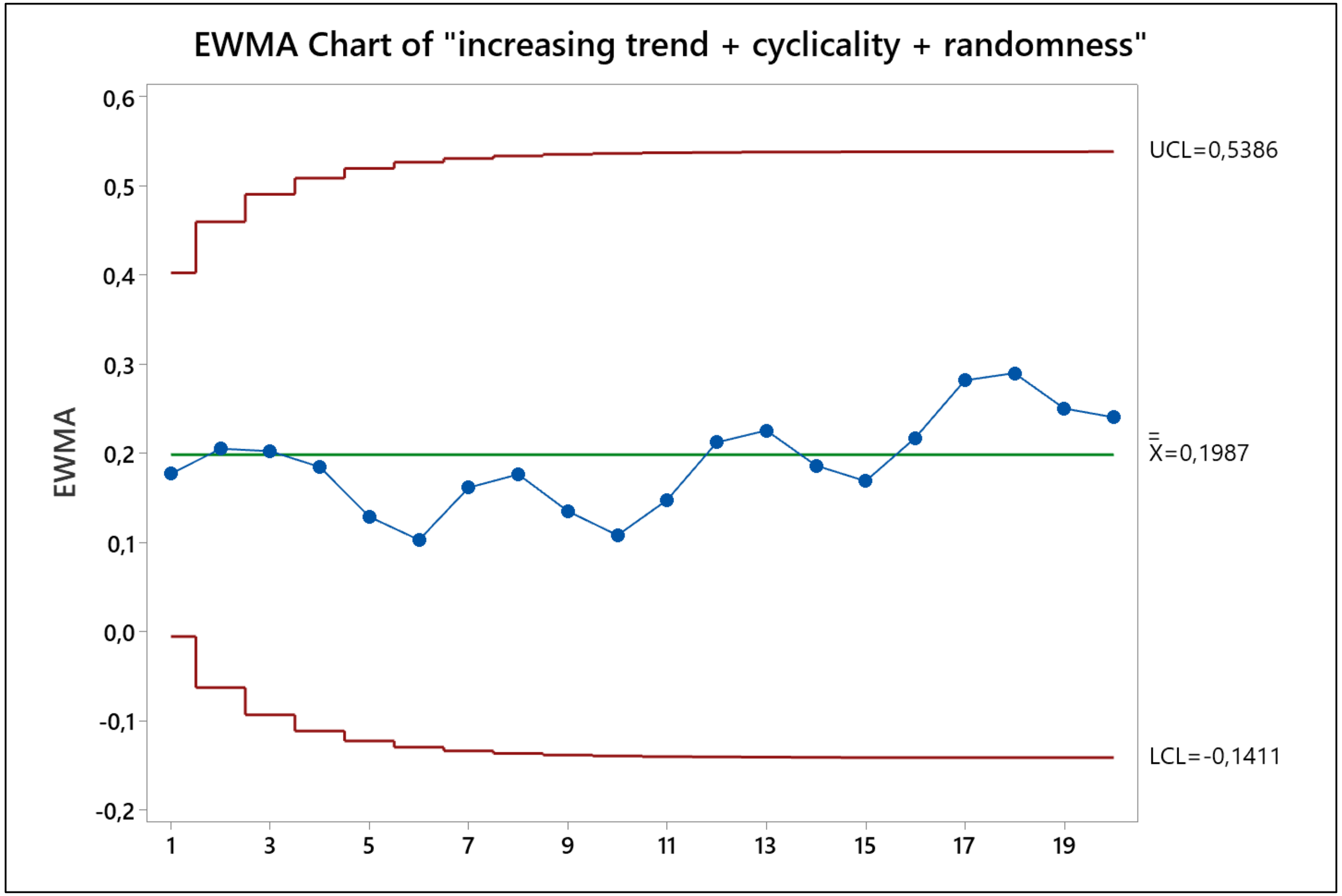
OUT OF TREND

The “special” *EWMA* (or Exponentially Weighted Moving Average) control chart, precisely because of the principle on which it is based, even with standard settings highlights a possible trend of growth in the data.



OUT OF TREND

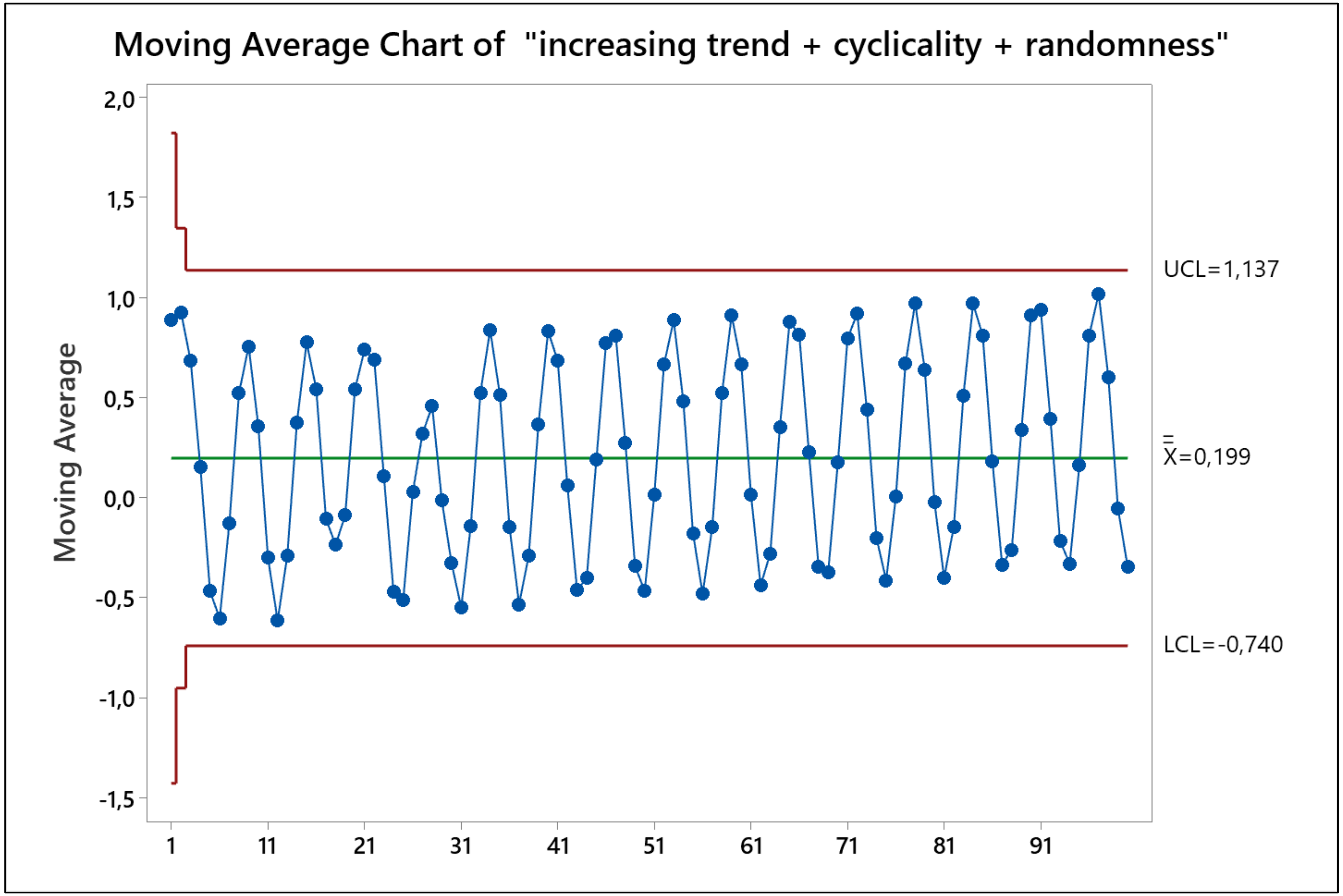
By adjusting the settings (*i.e.*, grouping data in groups of five), the “special” *EWMA* control chart returns the image of a growth trend in the data, albeit a slight one.



OUT OF TREND

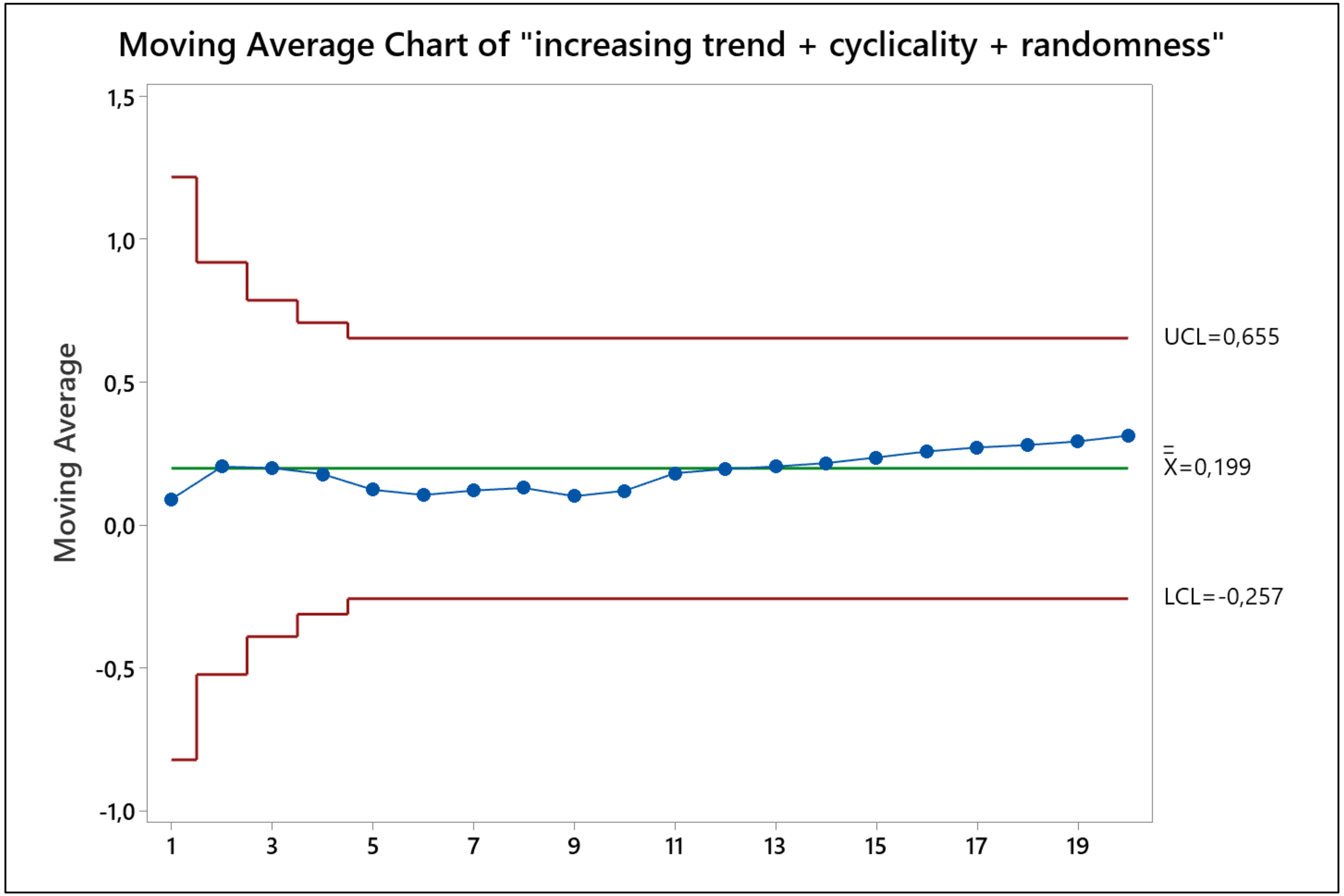
A similar result is obtained using another “special” control chart, *MA* (or Moving Average), with the standard settings.

In this case, unlike the *EWMA* chart at "standard settings", nothing is observed.



OUT OF TREND

By refining the control chart settings (*i.e.*, grouping data), the *MA* chart also returns the image of a weak growth trend in the data.



CONCLUSIONS

CONCLUSIONS

Summing up what has been said so far, **for Outliers / OOE (Out of Expectations) and OOT (Out of Trend) data there is no one-size-fits-all approach.**

- **Outliers / OOE (Out of Expectations) :**

  For them there are ***graphical methods*** and ***statistical tests***. The former are the simplest to use, above all because they do not require any hypotheses on the distribution of the data, while the latter must be used with great care to prevent big mistakes.

CONCLUSIONS

- **OOT (Out of Trend) :**
  - In this case it is necessary to establish the evaluation time horizon and the criteria so that a trend can be considered significant or not. A preliminary evaluation of the data using control charts or other statistical criteria (*e.g.*, percentiles) can help define meaningful significance criteria.
  - Subsequently, the use of particularly sensitive control charts (*e.g.*, ***EWMA*** or ***MA***) combined with adequate data grouping are certainly useful for highlighting any trends.

REFERENCES

- V. Barnett, T. Lewis, *Outliers in Statistical Data*, 2nd Ed., J. Wiley and Sons (1984)
- D. Hawkins, *Identification of Outliers*, Chapman and Hall (1980)
- R.J. Beckman, R.D. Cook, *Outlier……….s*, Technometrics, 25 (1983) 119-149
- B. Iglewicz, D.C. Hoaglin, *How to Detect and Handle Outliers*, American Society for Quality Control, Vol. 16, (1993)
- R.F. Green, *Outlier-Prone and Outlier-Resistant Distributions*, J. Am. Stat. Assoc., 71 (1976) 502-505
- J.W. Tukey, *Exploratory Data Analysis*, Addison Wesley, 1977

REFERENCES

- W.J. Dixon, F.J. Massey, *Introduction to Statistical Analysis*, 3rd Ed., McGraw-Hill, 1969
- R.E. Shiffler, *Maximum Z Scores and Outliers*, American Statistician, 42 (1988) 79-80
- S. Bolton, C. Bon, *Pharmaceutical Statistics – Practical and Clinical Applica*tions, CRC Press (2010)
- R.J. Hyndman, G. Athanasopoulos, *Forecasting – Principles and Practice*, OTexts (2018)
- D.G. Montgomery, *Statistical Quality Control – A Modern Introduction*, Wiley (2013)
- L.S. Nelson, *The Shewhart Control Chart—Tests for Special Causes*, J. Quality Technology, Vol. 16, 4 (1984) 237-239