QUALITY METRICS AND DATA CONSISTENCY

- Part 1 -

1. The theory of statistics, as it is now understood, can be divided into two parts which are for many purposes better kept distinct. The first function of the theory is purely *descriptive*. It devises numerical and diagrammatic methods by which certain salient characteristics of large groups of phenomena can be briefly described; and it provides formulae by the aid of which we can measure or summarize the variations in some particular character which we have observed over a long series of events or instances. The second function of the theory is *inductive*. It seeks to extend its description of certain characteristics of observed events to the corresponding characteristics of other events which have not been observed. This part of the subject may be called the theory of statistical inference; and it is this which is closely bound up with the theory of probability.
2. The union of these two distinct theories in a single science is natural.

J.M. Keynes, A Treatise on Probability, Macmillan &Co., London (1921) Chapter 27

INDEX

- **INTRODUCTION** : What are Quality Metrics?

  Understanding variability as *« Quality is inversely proportional to variability »*

  Process control strategy requires Quality Metrics

  State of the art

- **PURPOSE** : Create knowledge from available data and therefore:

  ▪ manage possible anomalous or risky situations (OOS, OOT, deviations, *etc.*)

  ▪ communicate awareness in what is done and in the reliability of the processes used

INDEX *(cont.)*

- **CASE STUDY 1** :
  - Use of graphical methods
  - Normal distribution, Normality test and Hypothesis test (α, *P-value*)
  - Anscombe's quartet : Graphics reveal data !
- **CASE STUDY 2**:
  - Control Charts (*I-MR* Chart, *Run* Chart, *Xbar* Chart)
  - Structures in data: clustering, trending, *etc.*
  - ANOVA, *t-test*, *2-Variances test* and comparison between two series of data
  - Example: evaluation of Supplier data
  - Bland-Altman (or Tukey Mean-Difference) plot
  - Central Limit Theorem: *regardless of the shape of parent population, the distribution of means quickly approaches the normal distribution*

BACKGROUND

In August 2002, FDA announced the **Pharmaceutical CGMPs for the 21st Century Initiative**.

In that announcement, the Agency explained its intent to integrate **quality systems** and **risk management** approaches into the existing programs with the goal of encouraging industry to adopt modern and science-based manufacturing technologies.

These two concepts became the foundation of many FDA and ICH Quality Guidelines finalized later such as ICH Q8 (*2006*), Q9 (*2006*) , Q10 (*2008*), Q11 (*2012*) and Q12 (*2020*).

Pharmaceutical cGMP's for the 21st Century: A Risk-Based Approach (https://www.fda.gov/media/77391/download)

BACKGROUND

In 2015, as part of the **Pharmaceutical CGMPs for the 21st Century Initiative**, FDA sought input from industry on the establishment of an **FDA Quality Metrics Program** as another mechanism to promote continual improvement in manufacturing quality.

Metric is a word we will often meet and therefore it is worth spending on some words.

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

METRIC?

Definition

A standard of measurement

History

The metric system was invented in France in the years following the French Revolution, and a version of it is now used in most of the world to measure distance, weight, and volume. Basic metric units include the *kilogram* (the basic unit of weight), the *liter* (the basic unit of volume), and of course the *meter* (the basic unit of length).

Examples of metric in a Sentence (Adjective):

// The *metric* unit of energy is the “joule”

Merriam Webster’s online dictionary

WHAT FDA INTENDS AS QUALITY METRICS ?

« What are Quality Metrics?

Quality metrics are used throughout the drugs and biologics industry to monitor quality control systems and processes. Modern manufacturing includes robust quality metrics programs as a foundation for continual improvement of product and process quality.
Quality metrics are one element of companies' commitment to quality culture. »

Quality Metrics = Quantitative Indicators of Quality

https://www.fda.gov/drugs/pharmaceutical-quality-resources/quality-metrics-drug-manufacturing

WHICH METRICS FDA INTENDS TO CALCULATE ?

- **Lot Acceptance Rate (LAR)** as an indicator of manufacturing process performance.
  LAR = the number of accepted lots in a timeframe divided by the number of lots started by the same covered establishment in the current reporting timeframe.

- **Product Quality Complaint Rate (PQCR)** as an indicator of patient or customer feedback.
  PQCR = the number of product quality complaints received for the product divided by the total number of dosage units distributed in the current reporting timeframe.

- **Invalidated Out-of-Specification (OOS) Rate (IOOSR)** as an indicator of the operation of a laboratory. IOOSR = the number of OOS test results for lot release and long-term stability testing invalidated by the covered establishment due to an aberration of the measurement process divided by the total number of lot release and long-term stability OOS test results in the current reporting timeframe.

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

WHY FDA CONSIDERS THESE METRICS ?

It deals of very general metrics that FDA intends to use to:

- implement a “risk-based” scheduling of drug manufacturing facilities
- improve its ability to evaluate pharmaceutical manufacturing and production operations
- predict, and therefore mitigate, the possibility of future drug shortages

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

WHY QUALITY METRICS ?

Because they are:

- Objective
- Subject to inspection under section 704 of the FD&C Act
- Valuable in assessing the overall state of quality of the product and process

HOWEVER, IN THE SAME DOCUMENT, FDA ALSO STATES THAT…

ARE THESE THE ONLY QUALITY METRICS ?

« FDA understands that establishments involved in the manufacture, preparation, propagation, or processing of human drugs, including oversight to ensure quality, currently use quality metrics as part of the process validation lifecycle and pharmaceutical quality system (PQS) assessment. The metrics described in this guidance could be a part of such oversight.

FDA encourages manufacturers to routinely use additional quality metrics beyond the metrics described in this guidance in performing product and establishment specific evaluations. The selected metrics are not intended to be an all-inclusive set of the quality metrics that manufacturers may find useful to assess a product and manufacturer's state of quality. »

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

TRUE TARGET: UNDERSTANDING VARIATION!

This, in fact, is the real purpose of Quality Metrics !

In the FDA Guidance on Process Validation is stated that manufacturers should:

- Understand the source of variation
- Detect the presence and degree of variation
- Understand the impact of variation on the process and ultimately on product attributes
- Control the variation in a manner commensurate with the risk it represents to the process and the product.

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

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

TRUE TARGET: UNDERSTANDING VARIATION!

These aspects concerning the "***understanding of variation***" often recur in numerous ICH and FDA guidelines (*e.g.*, FDA Guidance on Process Validation, January 2011, page 4) as they are essentials to establish if a ***"process is under control or not".***

The relevance of this is evident if we consider that:

« Quality is inversely proportional to variability »

« Quality improvement is the reduction of variability in processes and products »

D. C. Montgomery, Statistical Quality Control: A Modern Introduction, 7th Edition, Wiley (2013)

TRUE TARGET: UNDERSTANDING VARIATION!

The FDA in its Guidance for Industry on Process Validation (2011) is even more precise:

« …we strongly recommend firms employ, objective measures (*e.g.*, statistical metrics) wherever feasible and meaningful to achieve assurance…»

« …an ongoing program to collect and analyze product and process data that relate to product quality must be established (§ 211.180(e)). »

« …the data should be statistically trended and reviewed by trained personnel…»

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

WHY UNDERSTANDING VARIATION?

Because:

… **Variation** is the *Voice of the Process*…

… **Specifications** are the *Voice of the Customer* …

W.W. Scherkenbach, Deming's Road to Continual Improvement, 1st Ed., SPC Press (1991)

UNDERSTANDING VARIATION

« After establishing and confirming the process, manufacturers **must** maintain the process in a **state of control** over the life of the process, even as materials, equipment, production environment, personnel, and manufacturing procedures change. »

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

UNDERSTANDING VARIATION

« Manufacturers should use ongoing quality programs to collect and analyze product and process information to evaluate the state of control of the process. These programs must be capable of identifying process or product problems and opportunities for manufacturing improvements that can be evaluated and implemented throughout the lifecycle. »

This is the essence of:

- ***Continued Product Verification*** - FDA Guidance on Process Validation (2011) or
- ***Ongoing Process Verification during Lifecycle*** - Annex 15 (EudraLex - Volume 4), ICH Q10, and ICH Q12.

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

FDA & QUALITY METRICS

FDA recognize in fact the need of establishing a ***product control strategy*** as:

- processes, over the time, tend to deteriorate/drift from initial conditions because of many different reasons (*e.g.*, material, personnel, environment, *etc.*)
- improvement in technologies, acquisition of manufacturing experience, *etc*. lay the foundations for a change in the process

and

- the combination of quality metrics along with internal data (*e.g.*, inspection results, recalls, *etc.*) is indicative of the state of quality of the establishment or product.

FDA & QUALITY METRICS

«… FDA encourages manufacturers to routinely use additional quality metrics beyond the metrics described in this guidance in performing these evaluations…»

From this quote it follows a clear tendency of all Regulatory Authorities, FDA first, towards an ever wider use of **quantitative methods** (or **Quality Metrics**) for processes and products monitoring.

Apparently, nothing new has happened from that distant 2002, but it is true that if, to date, **DATA INTEGRITY** is in some way ensured, **DATA CONSISTENCY**, at least equally important, is in fact ignored.

FDA Guidance for Industry (Draft) – Request for Quality Metrics (July 2015)

DATA CONSISTENCY

consistency *n* agreement or harmony of parts, traits, or features < his adversary had to admit the *consistency* of his position >

syn **coherence**, conformity, **congruity**, correspondence; *compare* HARMONY 2

rel **agreement**, concord, consonance, likeness, similarity; apposition, aptness, felicity, fitness, suitability

con incoherence, incongruity; impropriety, inappropriateness, unsuitability

ant inconsistency

Webster's Collegiate Thesaurus, Merriam-Webster, Inc.; 2nd ed. edition (Jan. 2010)

STATE OF THE ART

In the current daily practice, and for sure in most of the cases, **the Quality Metrics in use are:**

- **few** (*e.g.*, arithmetic mean, standard deviations, simple linear graphs obtained using Excel, *etc.*)
- **rough**

and, unfortunately,

- **often incorrectly used !**

In some cases, results shown during inspection or submitted to Regulatory Authorities show inconsistencies with theoretical indications (*e.g.*, microbiological data).

STATE OF THE ART

«…We live in the «Information Age», and much of that information comes to us in the form of numbers. Unfortunately, «information is random and miscellaneous, but knowledge is orderly and cumulative». Before information can be useful it must be analyzed, interpreted, and assimilated. The process of digesting data has been widely neglected at all levels of our educational system.

This deficiency has been characterized as " ***numerical naiveté*** ".

Numerical naiveté is not a failure with arithmetic, but it is instead a failure to know how to use the basic tools of arithmetic [and Statistics, *editor's note*] to understand data. Numerical naiveté is not addressed by the traditional courses in the primary or secondary schools, nor is it addressed by advanced courses in mathematics. This is why even highly educated individuals can be numerically naïve…»

D. J. Wheeler, Understanding Variation - The Key to Managing Chaos, SPC Press, 2nd Ed.(2000)

PURPOSE OF THIS PRESENTATION

Through practical examples, this presentation intends to show how simple quantitative methods (or quality metrics) allow to:

- extract useful information from the available data and therefore
- increase the knowledge of one's own data (and therefore of one's own **real** quality level),
- manage possible anomalous or risky situations (OOS, OOT, deviations, *etc*.),
- communicate awareness in what is done and reliability in the processes used,
- prevent possible unpleasant slips or mistakes during audits 😉

Moreover, all here above can be achieved without using advanced devices : graphs, numerical tables and a pocket calculator will be enough in most cases !

CASE STUDY 1

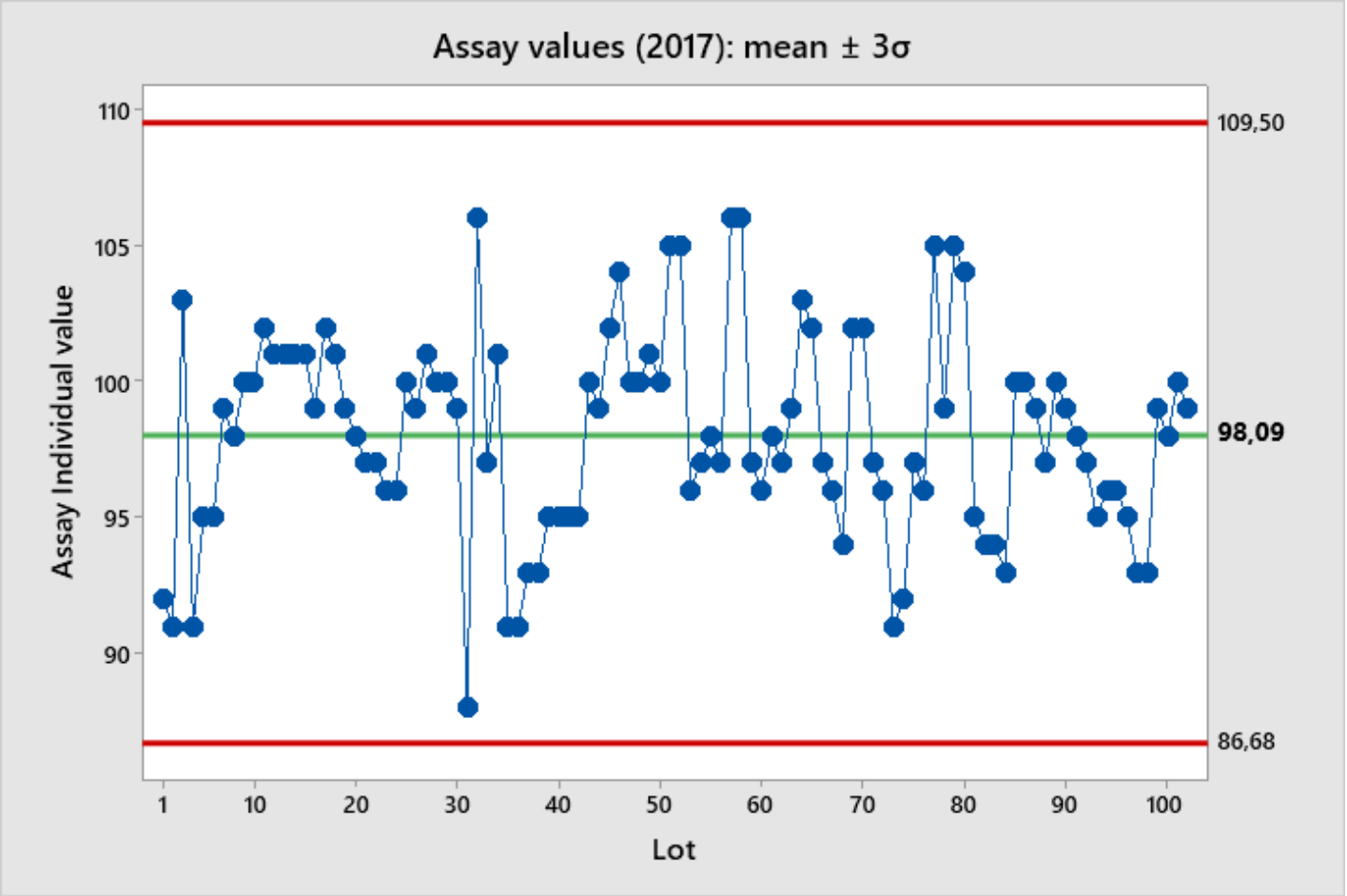
- Use of graphical methods
- Normal distribution, Normality test and Hypothesis test (α, *P-value*)
- Anscombe's quartet : Graphics reveal data !

CASE STUDY 1

Let's consider the HPLC assay values of 102 lots of an API manufactured, for instance, in 2017.

It is common practice to summarize and display this data in a so called « average ± 3σ plot » (*e.g.*, in APQR).

At a first glance, data points do not show anything anomalous and look distributed within the interval « average ± 3σ ».



CASE STUDY 1

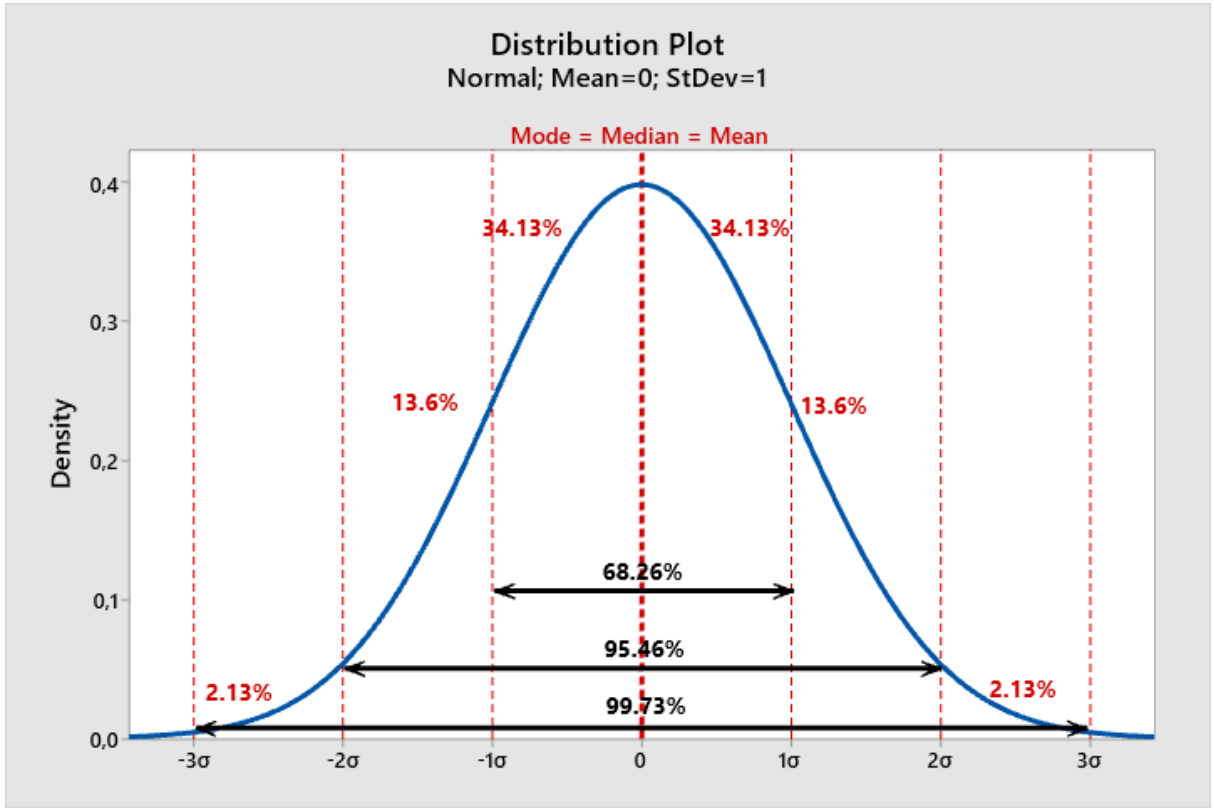
This last finding (*i.e.*: data within the interval « average ± 3σ ») is supported by the descriptive statistics here below reported which display an overall « normal behavior ». In fact:

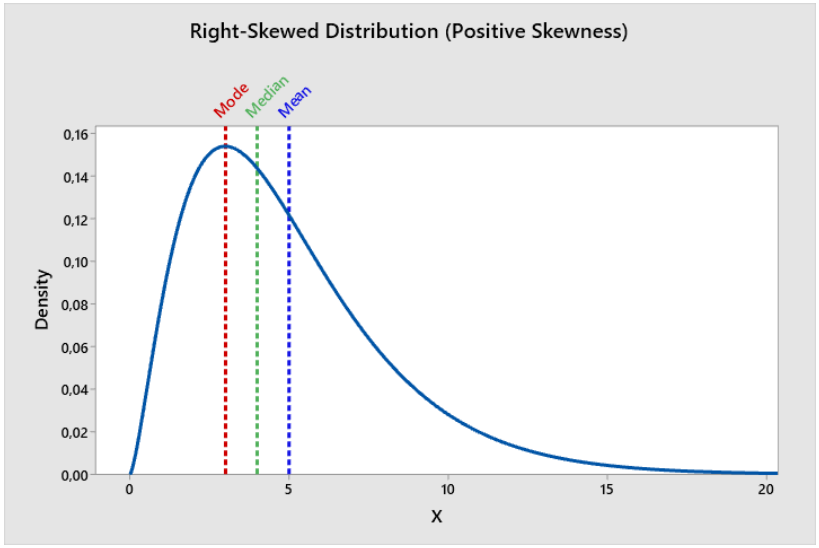
Arithmetic Mean ≅ Median ≅ Mode ; Asymmetry ≅ 0 ; Kurtosis = - 0.20

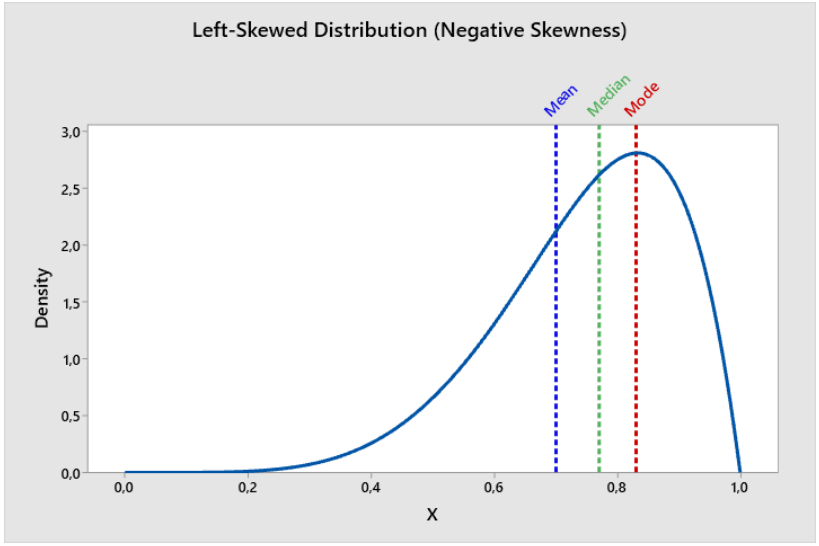
Statistics

Variable	N	Mean	StDev	Minimum	Median	Maximum	Mode	N for Mode	Skewness	Kurtosis
Assay 2017	102	98,088	3,802	88,000	98,000	106,000	100	13	-0,07	-0,20

CASE STUDY 1


Distribution Plot
Normal; Mean=0; StDev=1
Mode = Median = Mean
Density
0,4
0,3
0,2
0,1
0,0
34.13%
34.13%
13.6%
13.6%
68.26%
95.46%
99.73%
2.13%
2.13%
-3σ
-2σ
-1σ
0
1σ
2σ
3σ


Right-Skewed Distribution (Positive Skewness)
Mode
Median
Mean
Density
0,16
0,14
0,12
0,10
0,08
0,06
0,04
0,02
0,00
0
5
10
15
20
X


Left-Skewed Distribution (Negative Skewness)
Mean
Median
Mode
Density
3,0
2,5
2,0
1,5
1,0
0,5
0,0
0,0
0,2
0,4
0,6
0,8
1,0
X

CASE STUDY 1

The **normal curve** is due to the famous French mathematician Abraham De Moivre who mentioned it first in a paper published on November 12, 1733 and distributed only to friends.

The statistical use of the normal distribution began with Laplace and Gauss (distribution of errors) and Quételet made large use of it in Social Statistics (the *average man theory*: the individual person was synonymous with error, while the average person represented the true human being.).

However, this distribution was first called **normal distribution** by Sir Francis Galton in his lecture on *Typical Laws of Heredity* held at the Royal Institution on February 9, 1877.

K. Pearson started using the term only in 1893.

CASE STUDY 1

Beside the graphical shape and the observation that:

Arithmetic Mean = Median = Mode ; Asymmetry = 0 ; Kurtosis = 0

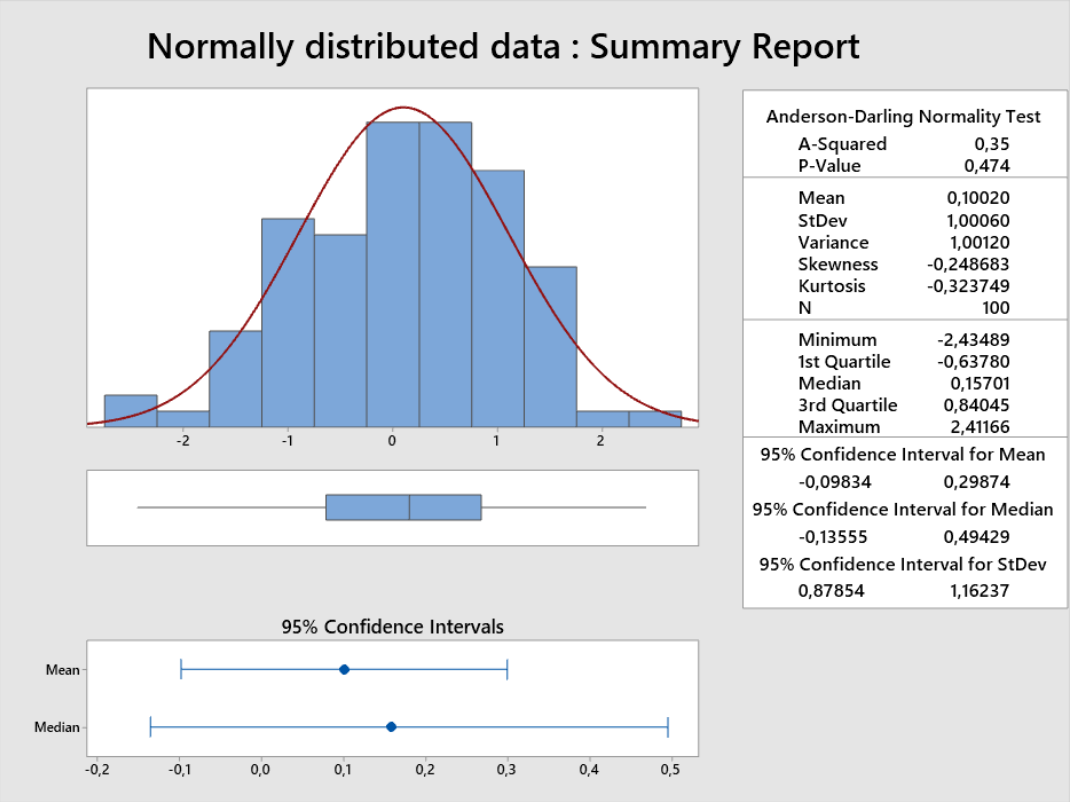
the “normality” of a data distribution can be established using specific statistical tests the so called: ***hypothesis tests*** (see Part 2 - CASE STUDIES 4/5).

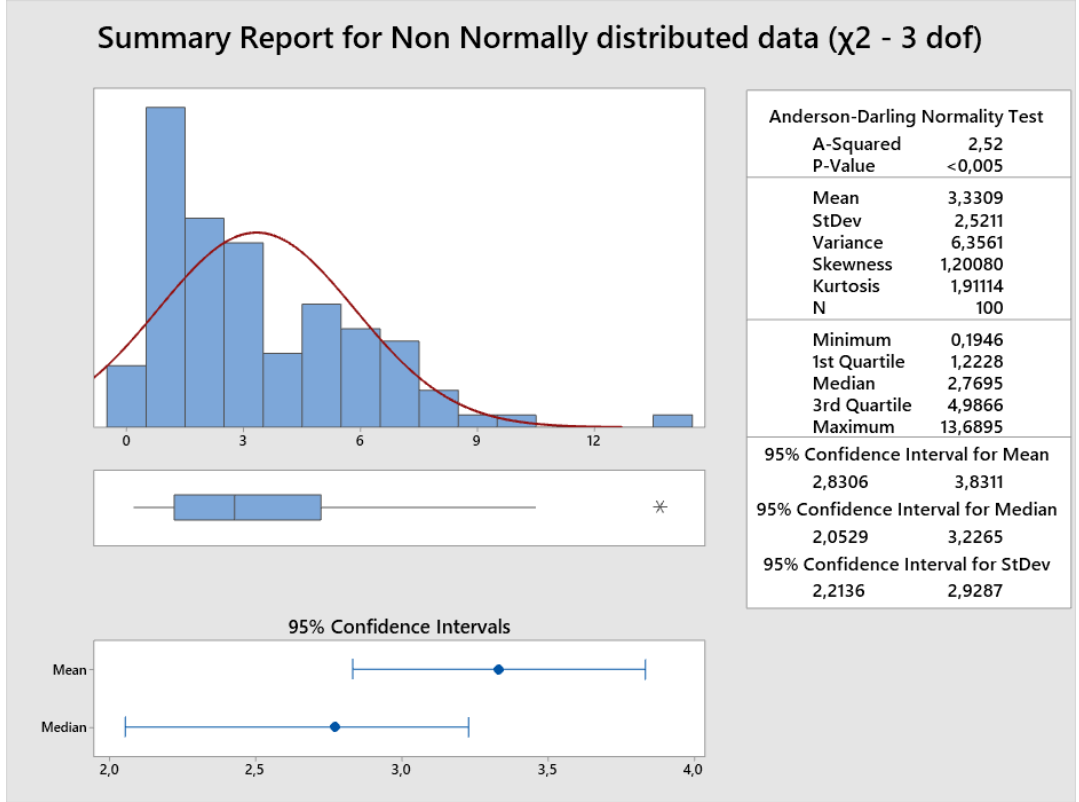
From a very general point of view it can be said that :

- ***P-value***: is a probability value [0, 1] that indicates whether a given distribution is adequate to represent the distribution of data. If *P-value* > 0.05 the distribution is adequate
- **AD** (Anderson Darling) coefficient : is a weighted average of distances from a straight line

CASE STUDY 1

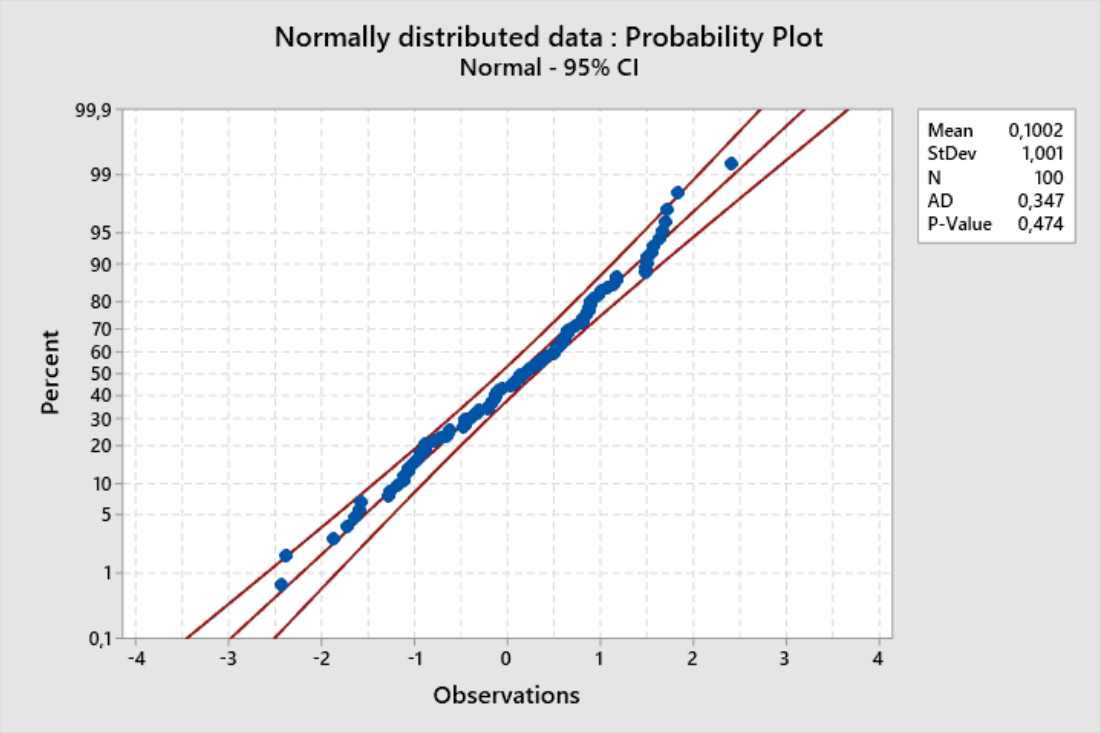
Visual comparison : Normally distributed data *vs.* Non-Normally distributed data

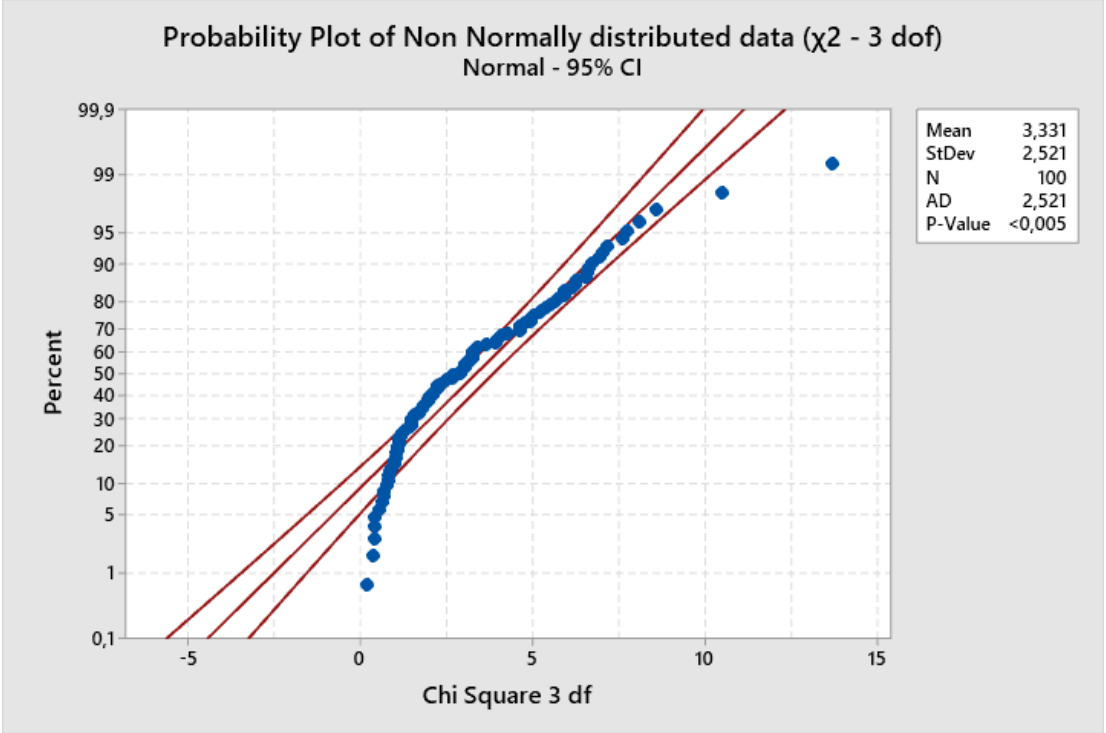




CASE STUDY 1

Visual comparison : Normally distributed data *vs.* Non-Normally distributed data

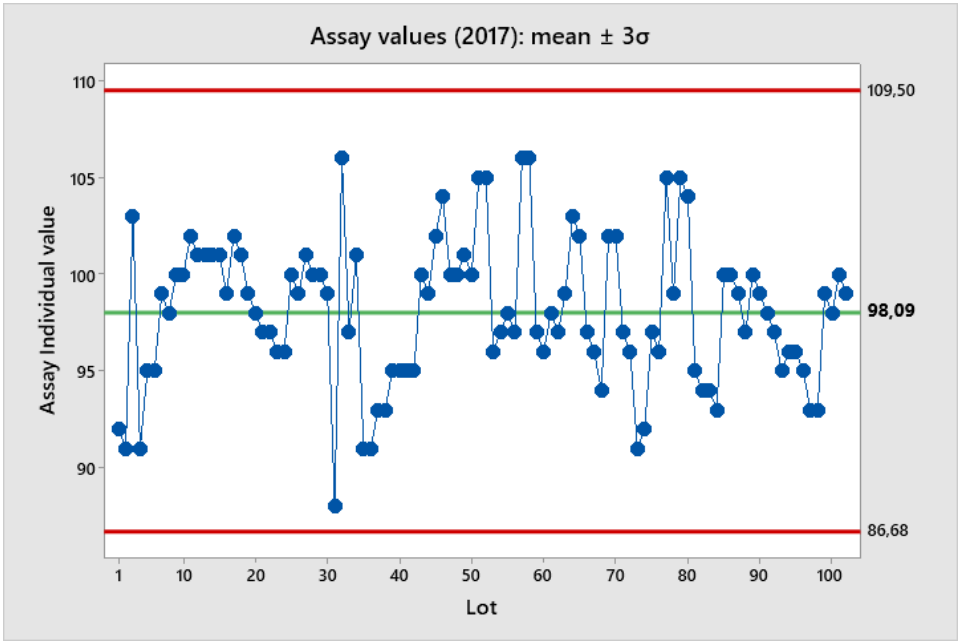


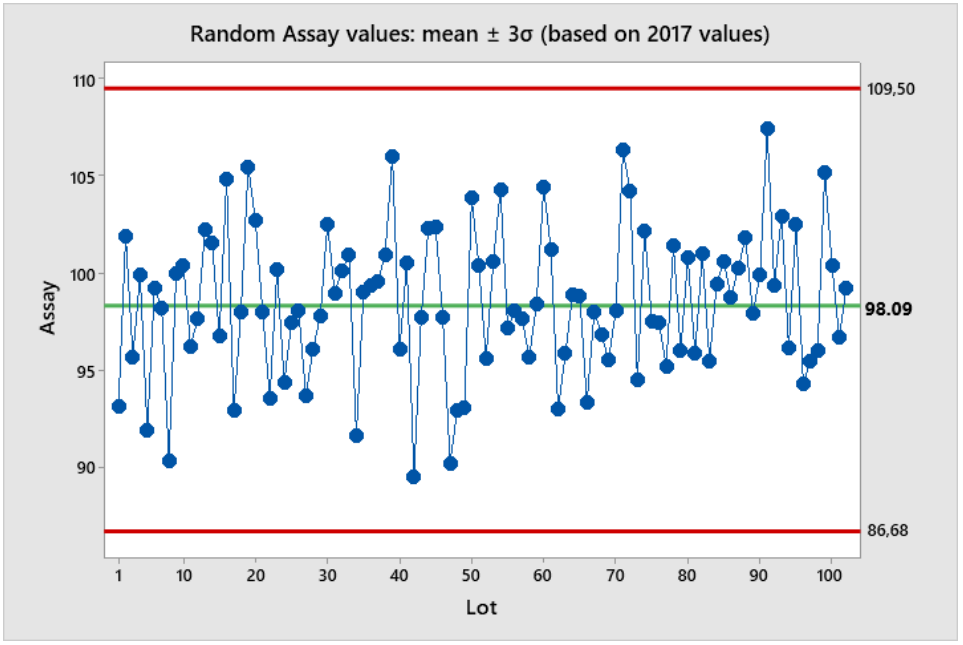


CASE STUDY 1

Let’s go back to the initial diagram:

- Data show VARIABILITY: an unavoidable characteristics !
- Moreover, data do not look «too perfect» such as, for instance, those in the control chart at the bottom that was built up simulating a normal distribution of same mean and standard deviation.





CASE STUDY 1

Please, always keep in mind that:

Round numbers are always false!

Samuel Johnson (1709-1784)

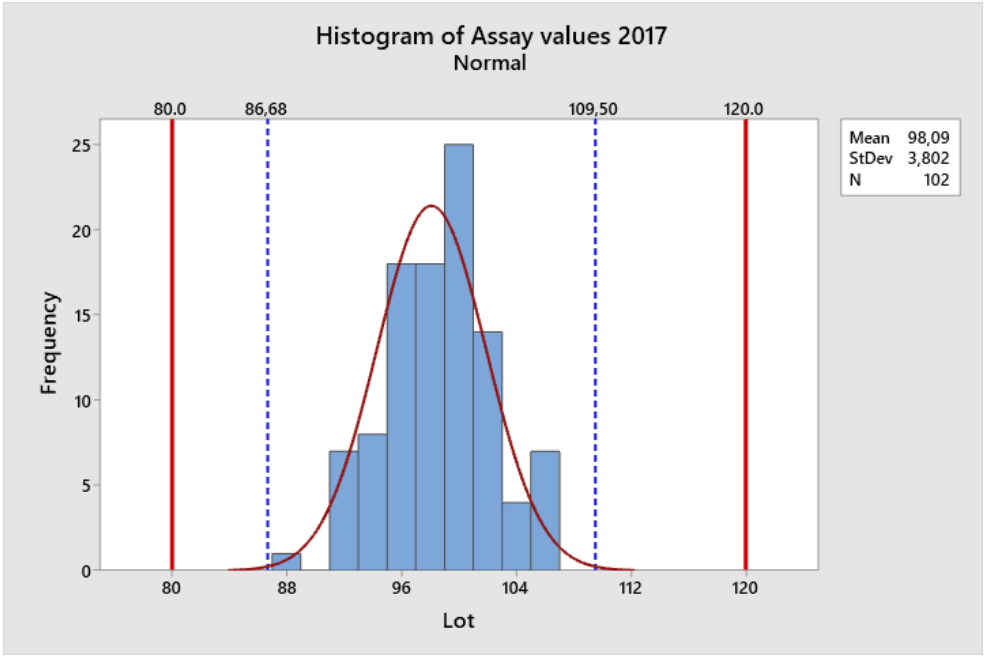
Torture numbers, and they will confess to anything

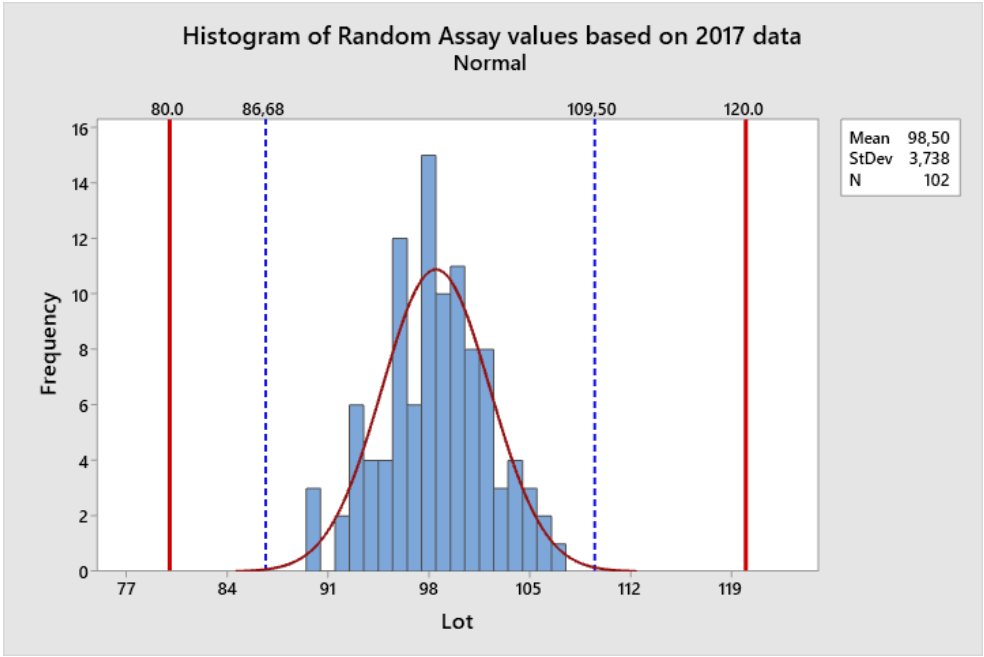
Gregg Easterbrook, New Republic (1999) vol. 221, page 42

CASE STUDY 1

In fact, comparing the histograms of the two data distributions it is immediately evident that the one resulting from PC generated data looks denser and more detailed with respect to the other. Bars are thinner.

Data generated by PC are different among them even if just slightly while those experimentally obtained sometimes recur as on the other hand expected !





CASE STUDY 1

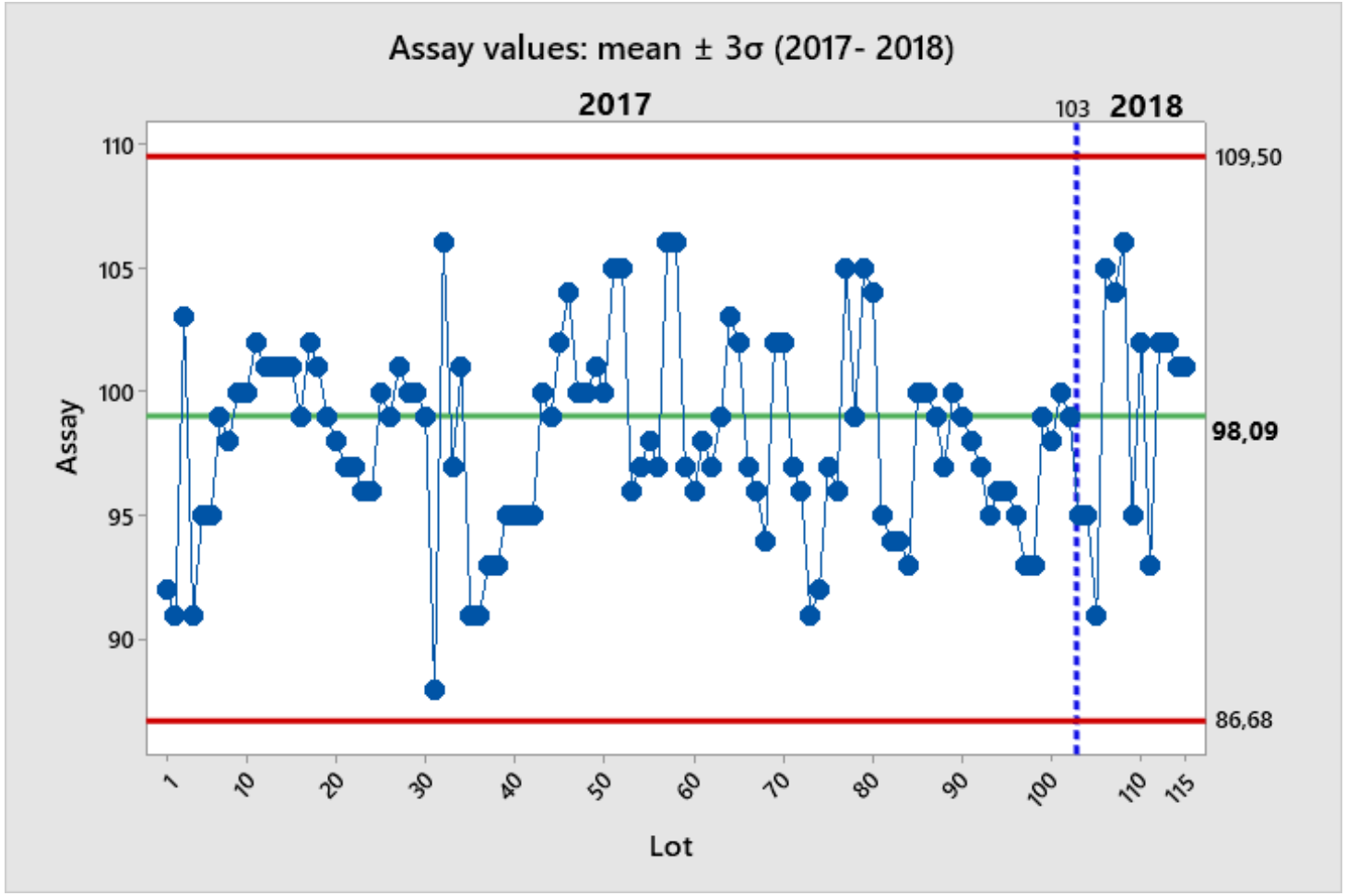
In light of what seen until now everything seems OK and the commonly used approach of the « average ± 3σ » plot (e.g., in APQR) looks correct.

Unfortunately, this way is WRONG as it is unable to detect any drift of the process or lack of control, unless it deals of something huge!

CASE STUDY 1

With the beginning of the new year (2018) some new lots are produced, and the previous control card updated, obtaining the result shown here on the right.

The situation seems to remain unchanged if compared to the previous year. In fact, the control chart does not show any noteworthy difference.

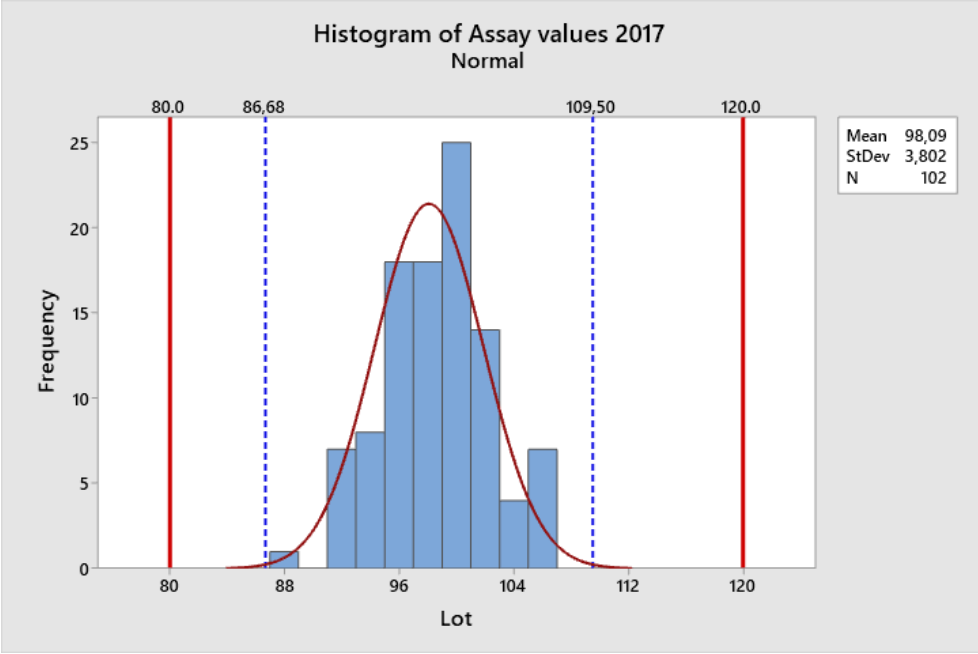


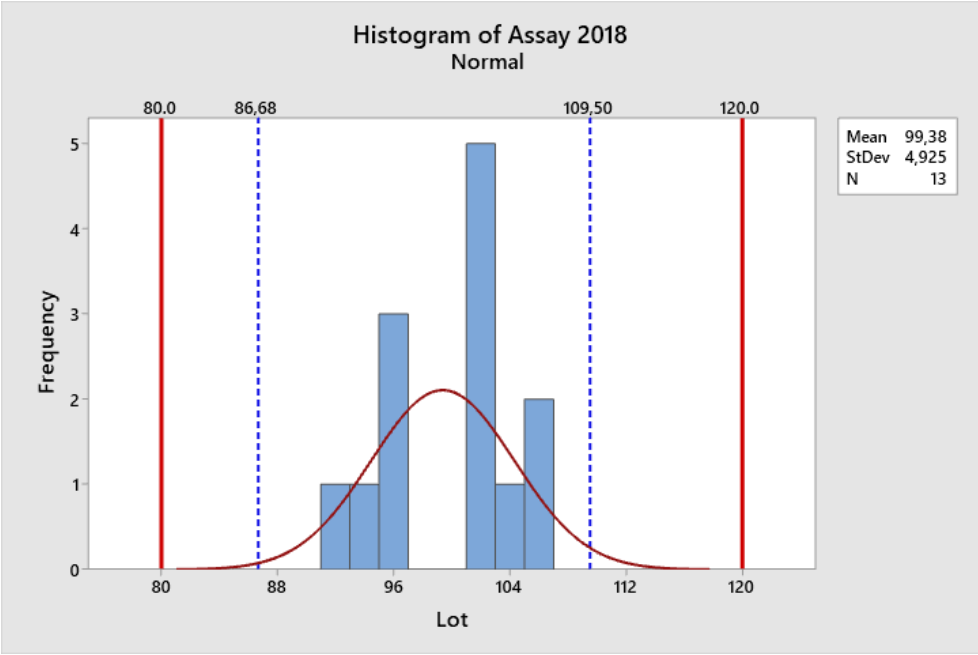
CASE STUDY 1

However, comparing the ***histograms*** of the two series of productions shows that:

lots produced in 2018 belong to two distinct sub-populations!

Something (***special cause***) happened in the process that the investigation system (*i.e.*, the « average ± 3σ » plot) could not reveal.

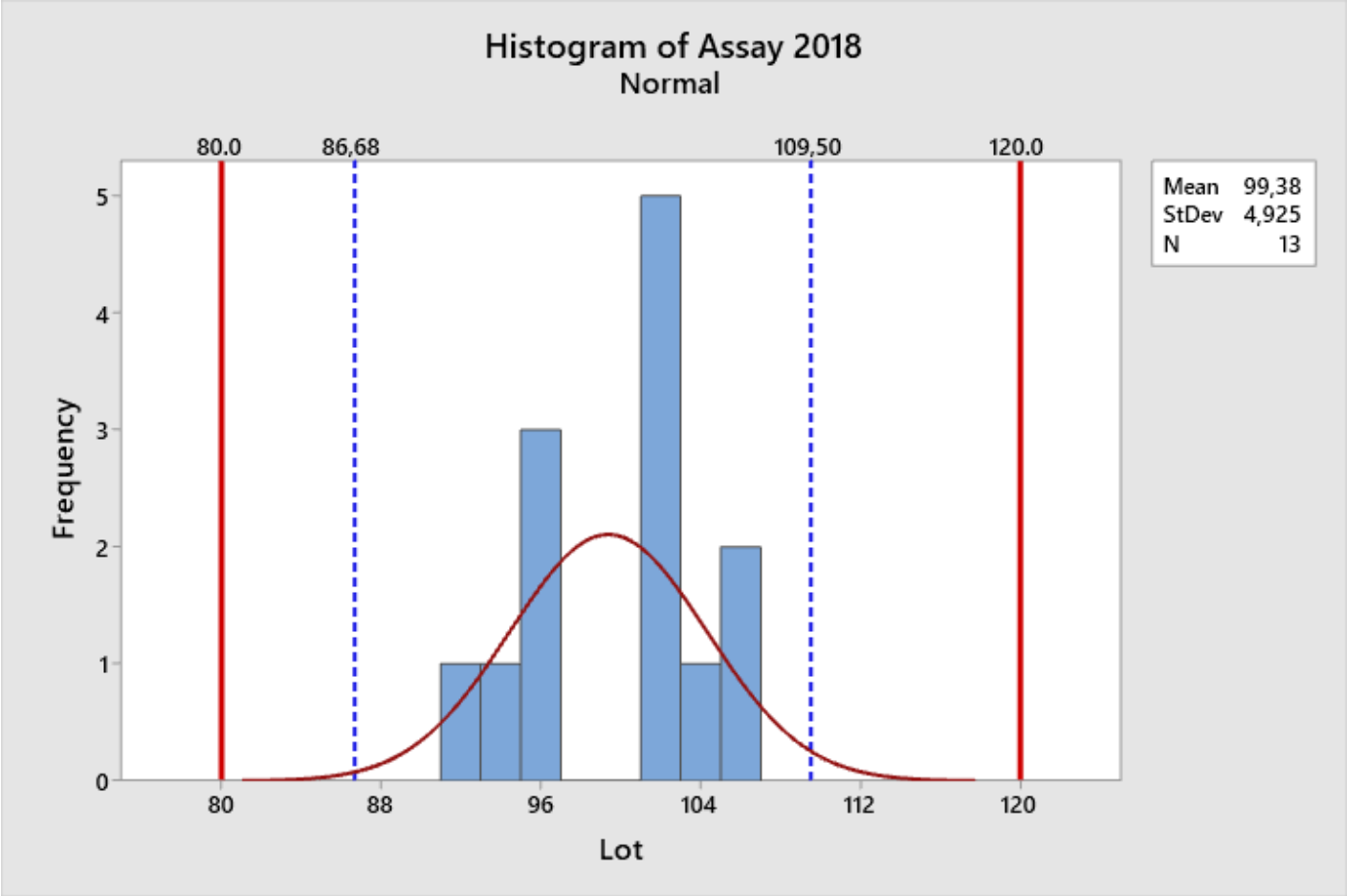




CASE STUDY 1

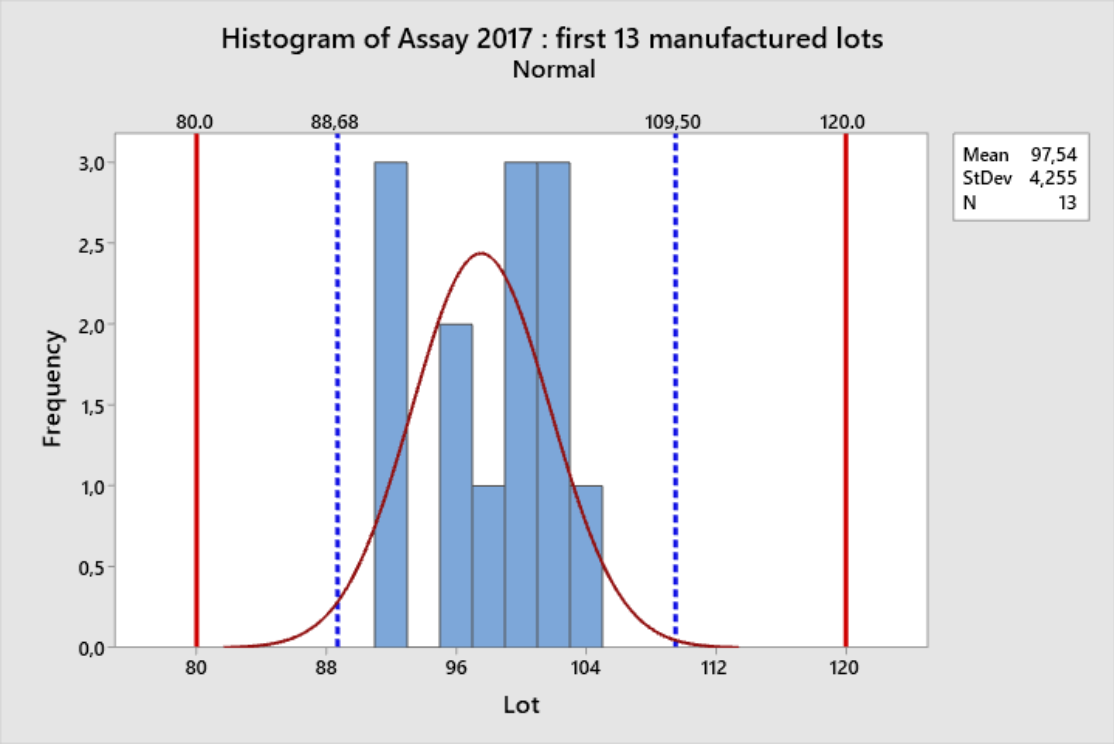
As shown, not only the new lots produced are within specifications and well between the limits « average ± 3σ », but even, on average, they have a higher assay than those produced in the previous year (99.4 *vs.* 98.1).

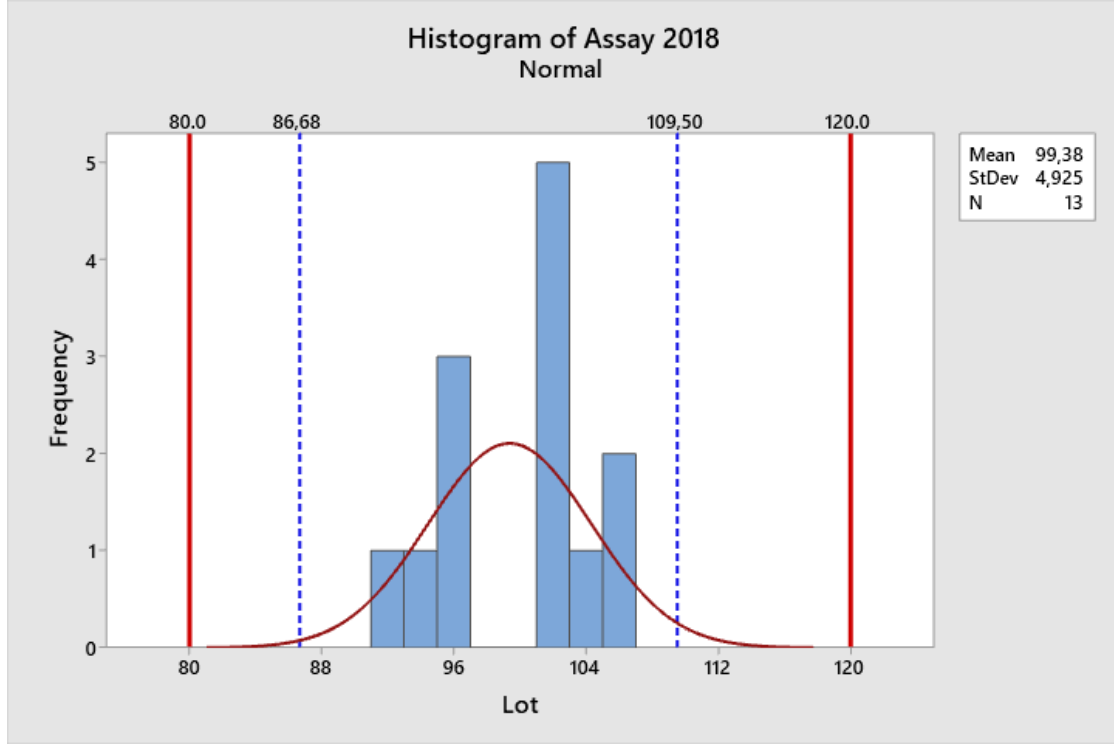
In this case, the use of a simple histogram highlighted a problem not revealed with the usual data analysis systems.



CASE STUDY 1

The comparison with the trend exhibited by the first 13 lots produced in 2017, shows that what observed in 2018 is anomalous!

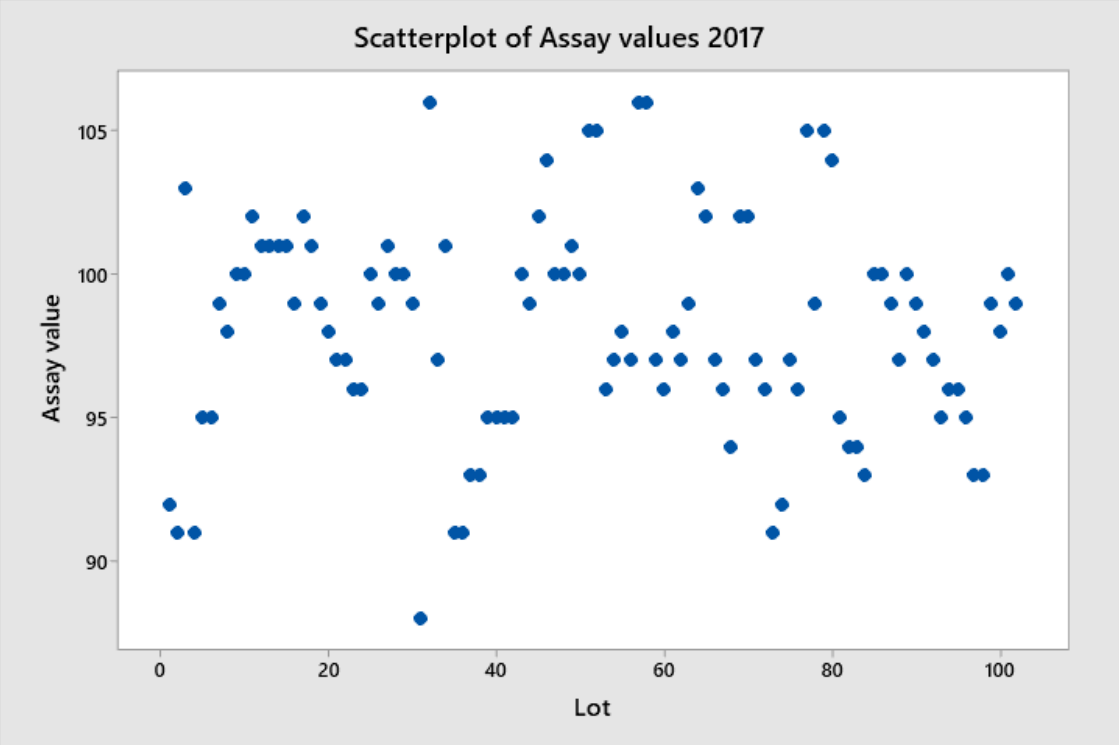


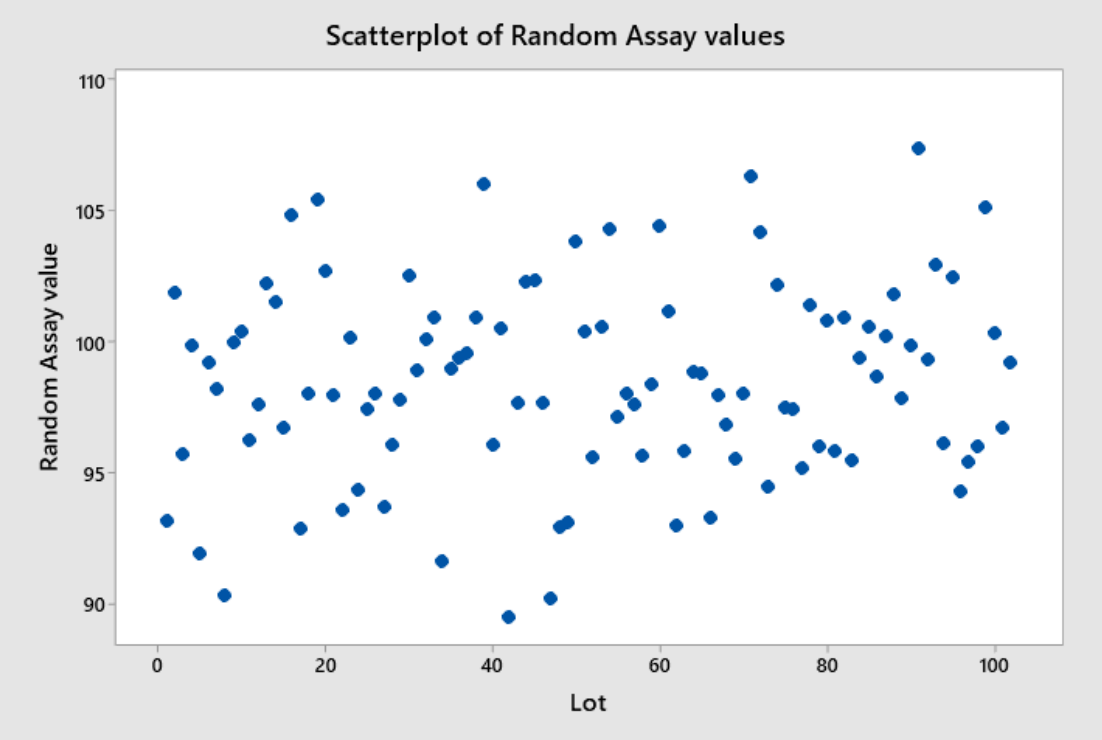


CASE STUDY 1

As an alternative to the histogram it could be considered another type of graphic representation, *e.g.*, the *scatterplot* or *dispersion diagram*.

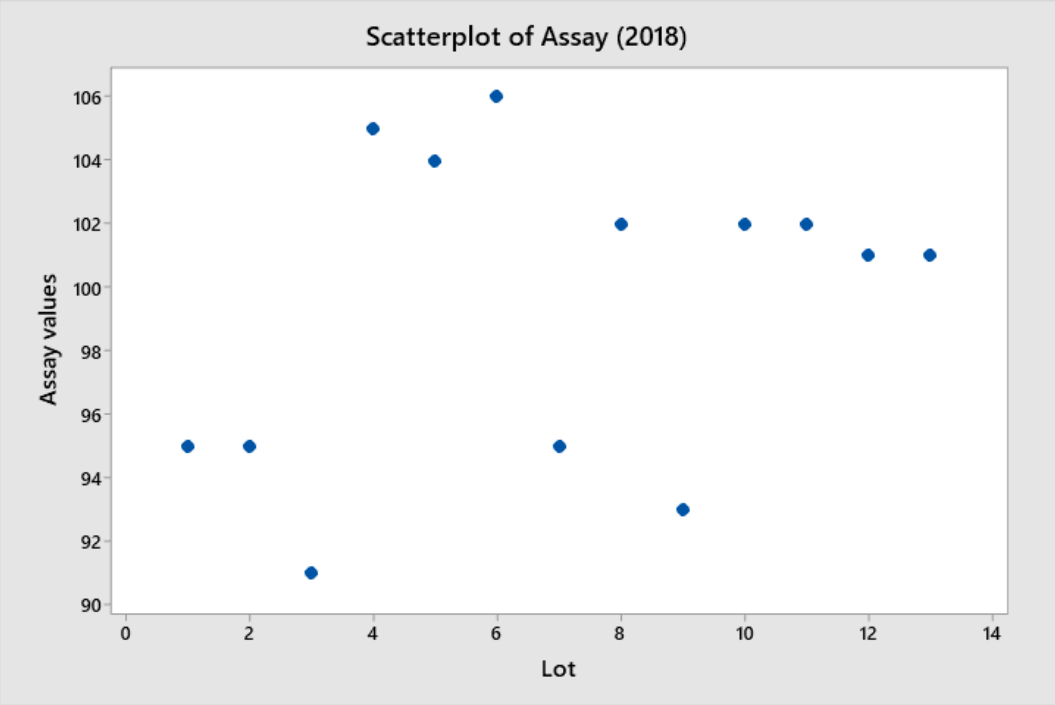
Here below the data 2017 are compared with computer generated random values and between the two plots no difference can be observed.

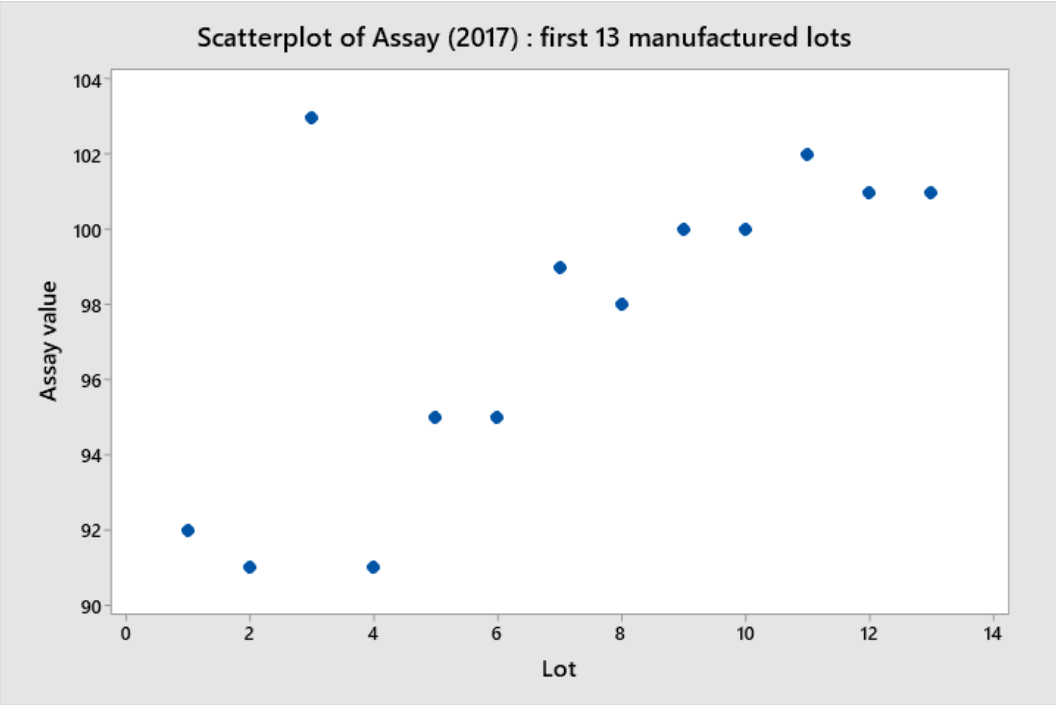




CASE STUDY 1

This indistinguishable situation also persists in the case plotted here below: what the scatterplot returns is, in this case, less immediate than the histogram → the optimal graphic representation must be assessed case by case !





CASE STUDY 1 - CONCLUSION

Graphics reveal data !

E.R. Tufte, The visual display of Quantitative Information, 2nd Ed. (2001)

There is no statistical tool that is as powerful as a well-chosen graph

J.M. Chambers et al., Graphical Methods for Data Analysis, Chapman and Hall (1983)

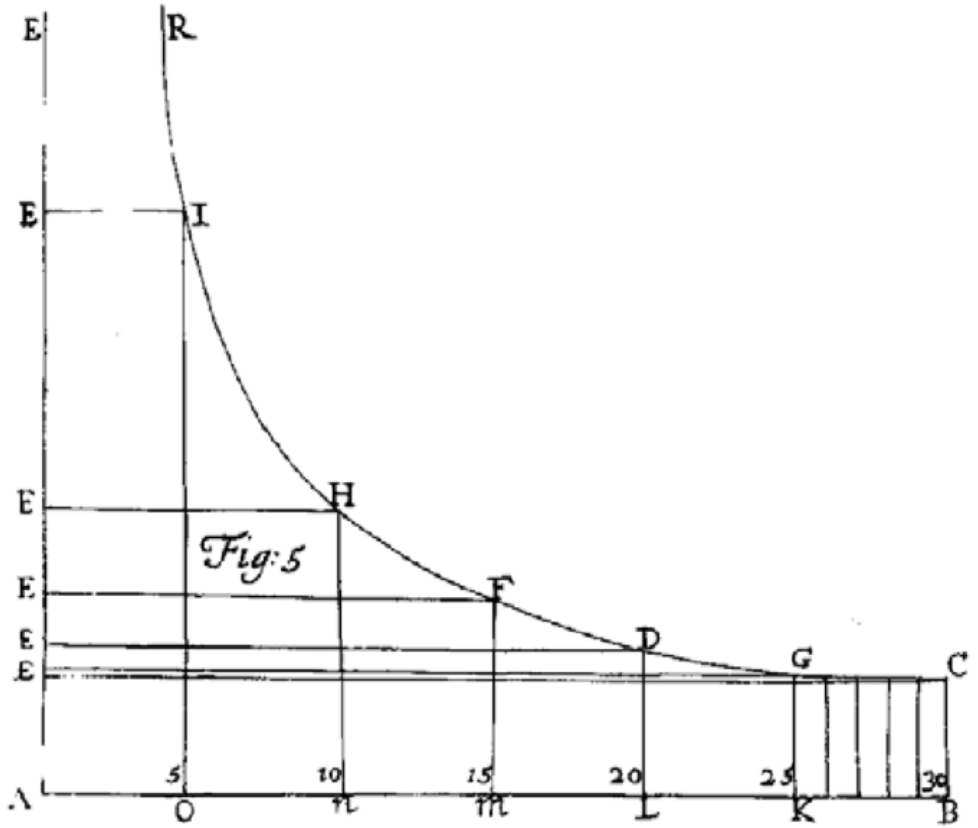
Always graph your data in some simple way, always !

E.R. Ott, Process Quality Control, 1st Ed., McGraw-Hill (1975)

CASE STUDY 1 - ADDENDUM

First known plot derived from observational data (but not showing the data directly) of a theoretical curve relating barometric pressure (y) to altitude (x) obtained in 1686 by Edmund Halley.

Coordinate systems and relations between graphs and functions $y = f(x)$ were introduced by Descartes and Fermat in 1630 !

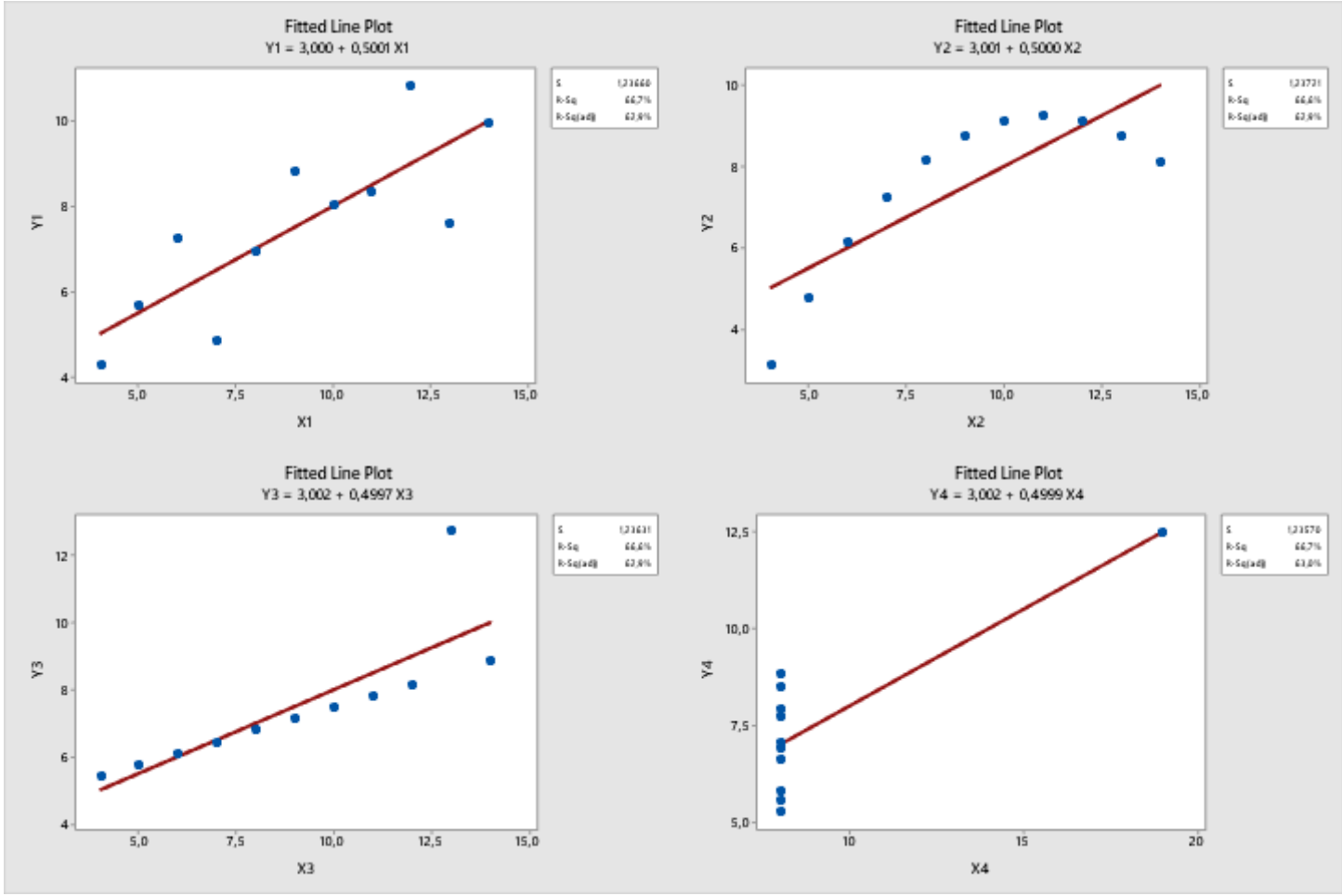


E. Halley, On the height of the mercury in the barometer at different elevations above the surface on the earth, and on the raising and falling of the mercury on the change of weather - Philosophical Transactions (1686) pp. 104-115

ANSCOMBE'S QUARTET

All four of these datasets are described by exactly the by identical descriptive statistics (mean and variance) and same linear model regardless of how they are arranged.

This is a clear example of why Ellis Ott's suggestion should always be kept in mind !



F.J. Anscombe, Graphs in Statistical Analysis – American Statistician, Vol. 27, No. 1. (1973)

CASE STUDY 2

- Control Charts (*I-MR* Chart, *Run* Chart, *Xbar* Chart)
- Structures in data: clustering, trending, *etc.*
- ANOVA, *t-test*, *2-Variances test* and comparison between two series of data
- Example: evaluation of Supplier data
- Bland-Altman (or Tukey Mean-Difference) plot
- Central Limit Theorem: *regardless of the shape of parent population, the distribution of means quickly approaches the normal distribution*

CASE STUDY 2

The statement “Graphics reveal data” is true, but the use of *Statistical Process Control* (SPC) most powerful tools, *i.e.*, S*hewhart’s control charts* (1924), would have highlighted, since the beginning, some « anomalies » in the initial data set that seemed ok if evaluated just using the « average ± 3σ » plot.

In this regard, the next slide is explanatory.

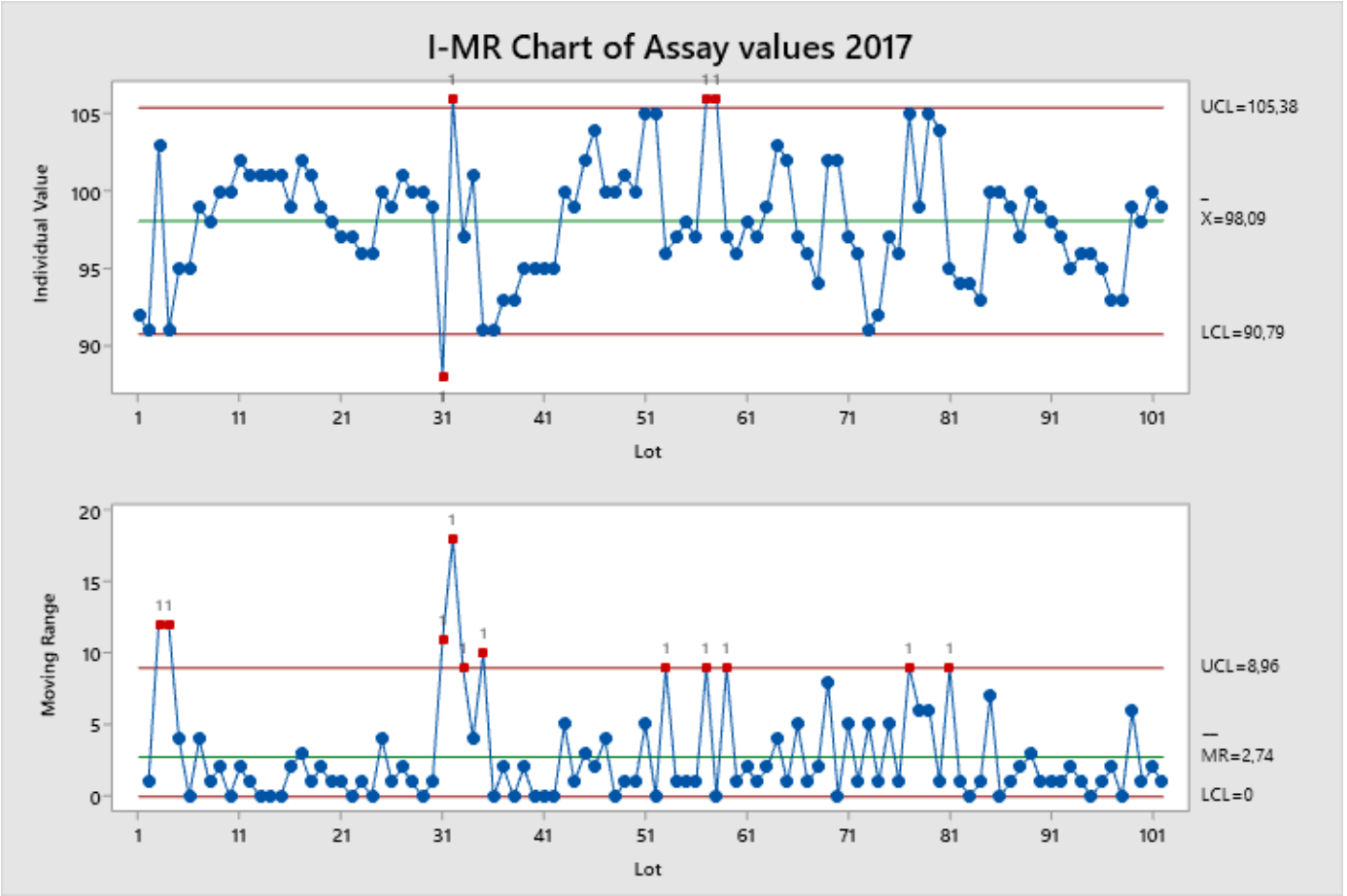
W.A. Shewhart, Economic Control of Quality of Manufactured Product – Van Nostrand (1931)

CASE STUDY 2

The classic «I-MR» control chart (*i.e.*, *Individuals* or Single Observation - *Moving Range* or Mobile Excursion) is shown alongside.

This chart describes the process data both in terms of **position** (process average) and **dispersion** (piece-to-piece variability).

Here, the control limits are based on the average variability measurement (moving average excursion) obtained from the absolute difference between two successive measurements (or moving excursion).

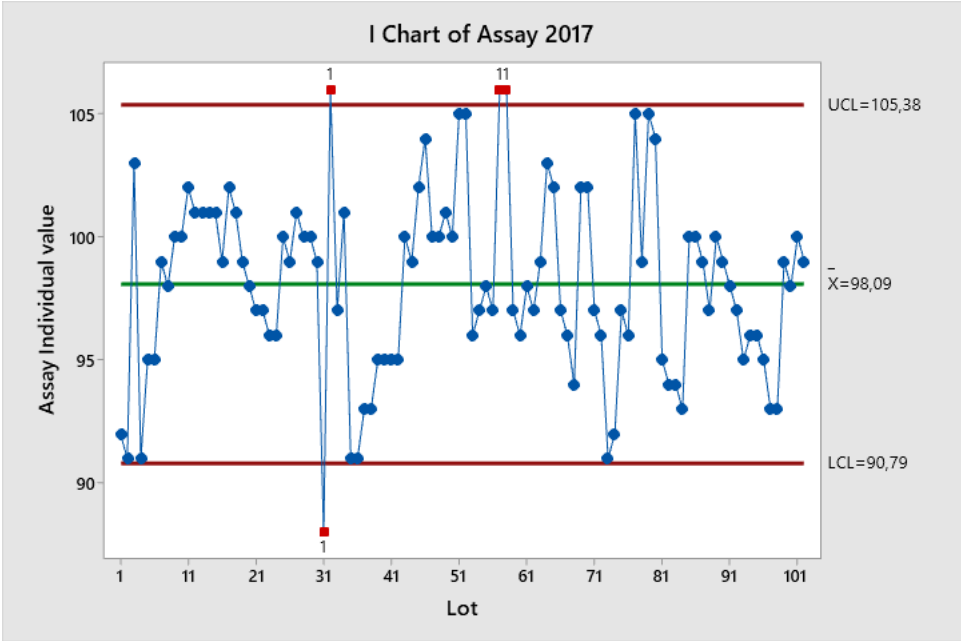


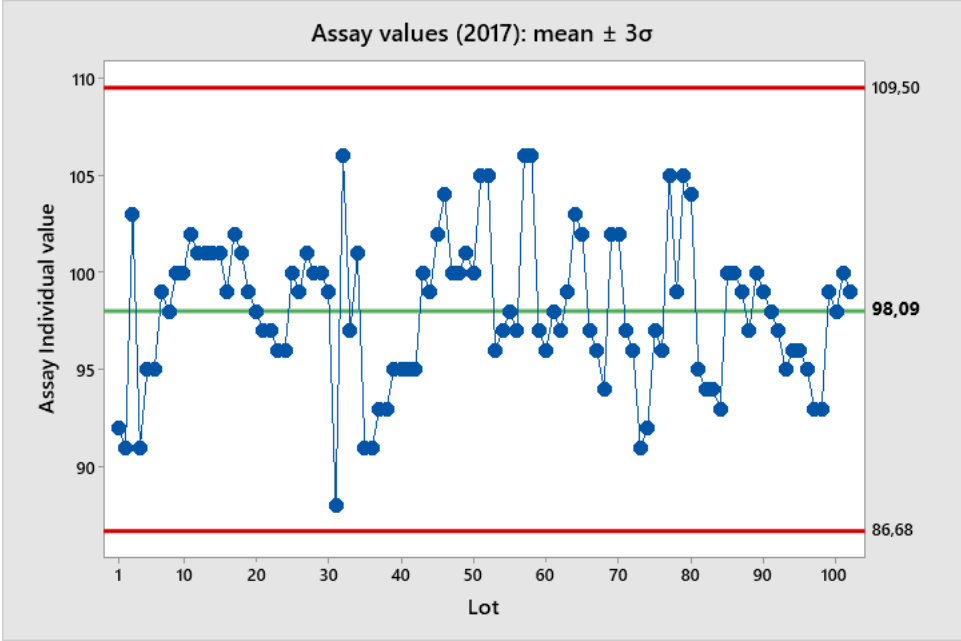
CASE STUDY 2

The control chart I (*Individuals*) shown in the previous slide is now compared to the plot with limits « average ± 3σ » that represents what is normally done in practice.

It is immediately evident that in this last plot all values are within the limits!

The difference lies in the criterion used to define the control limits ☺





CASE STUDY 2

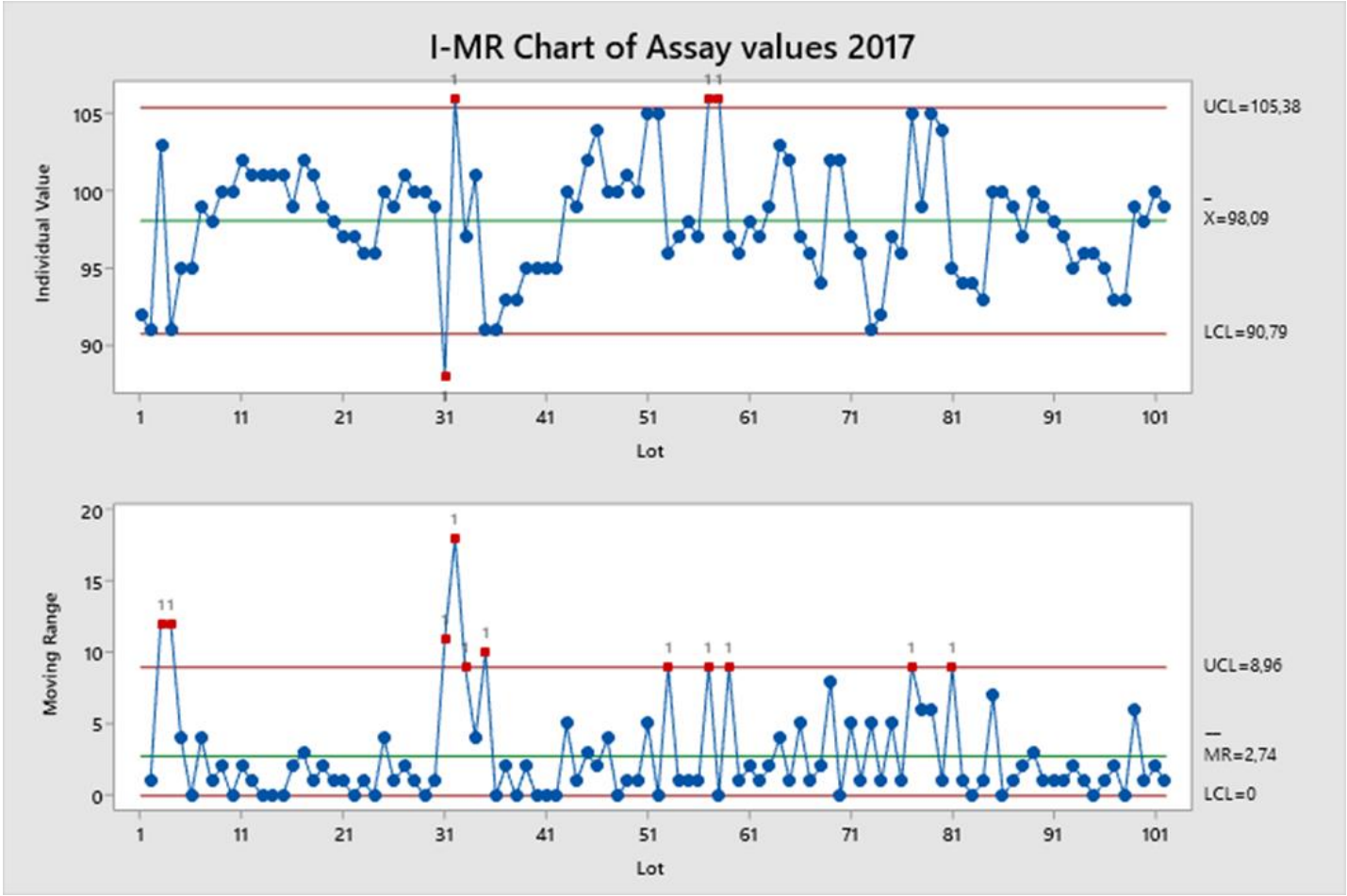
«… The R chart [MR, *editor's note*] reveals any undesirable variation within the subgroups and **is an indicator of the variability of the process under examination.**

It constitutes a measure of the uniformity of the process.

R chart remains in control if the variations within the subgroups remain substantially the same….

$\bar{X}$ charts can also be affected by out-of-control conditions in R charts…

... R chart is analyzed first ... "



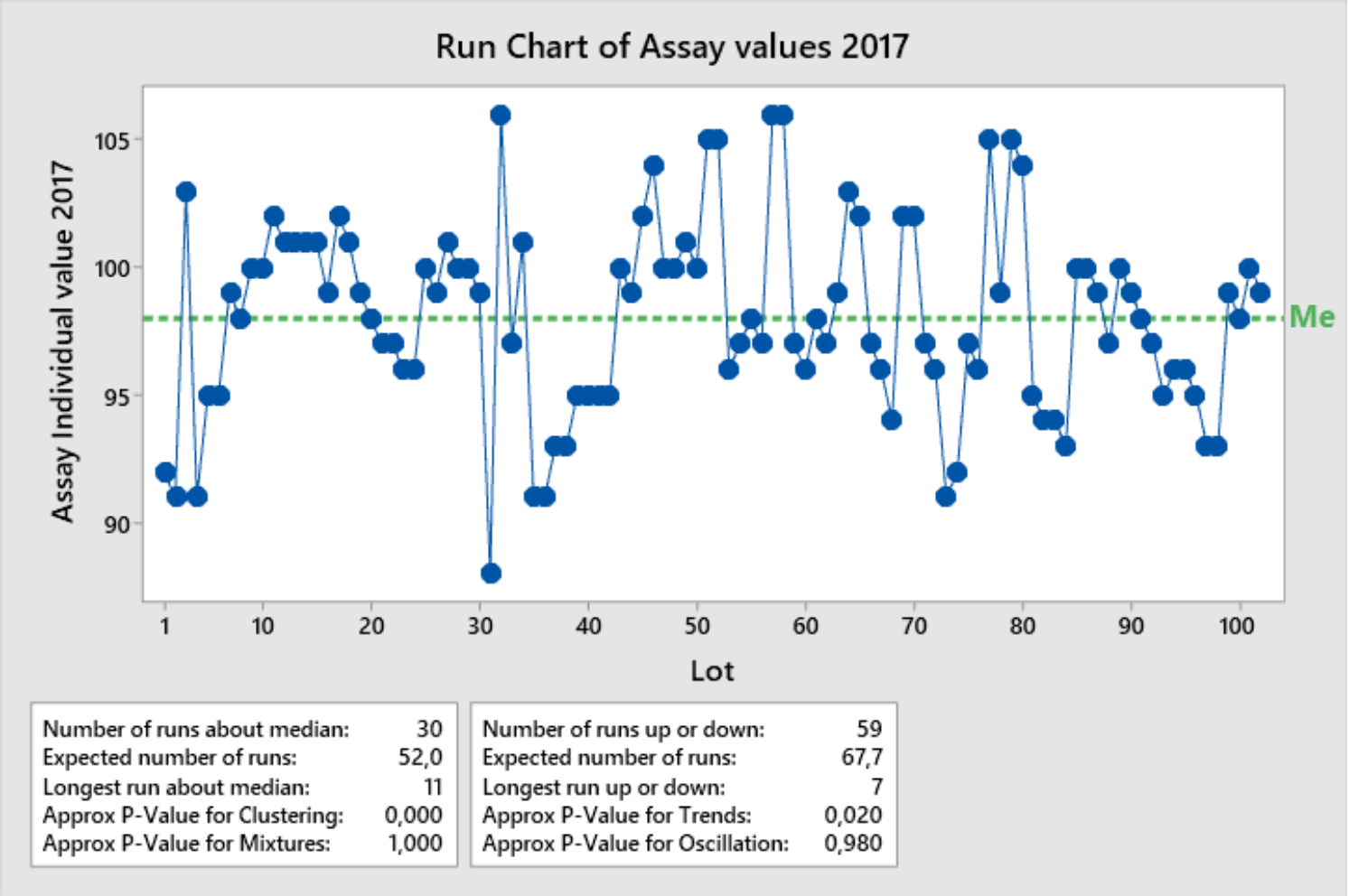
UNI ISO 8528:2004, Shewhart's Control Charts, (2004)

CASE STUDY 2

The use of another chart, the so-called « Run chart », a « median » control chart shown here, would have highlighted the presence of « ordered sub-structures » in the experimental data.

In fact, the values shown at the bottom show that the data show phenomena of:

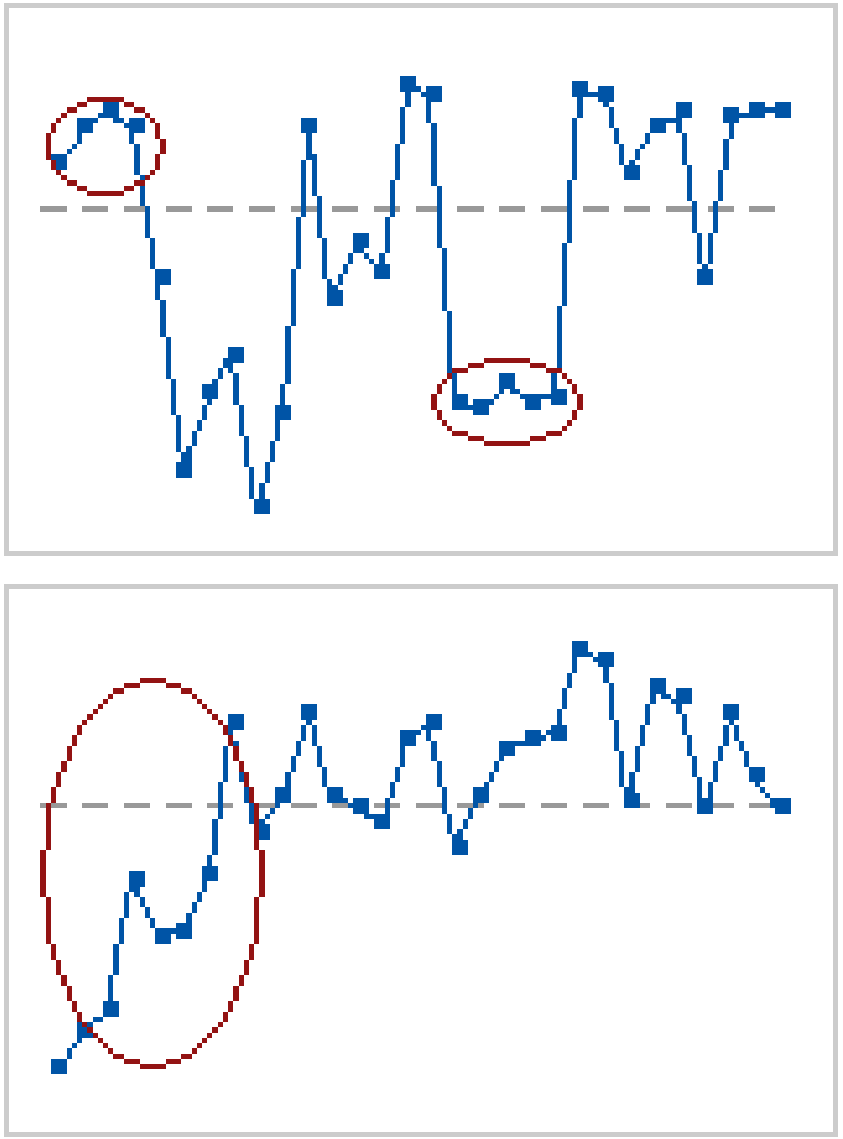
- Clustering
- Trends



CASE STUDY 2

- ***Clustering***: structures characterized by the grouping of data in an area of the graph. They can highlight problems due to measuring system, ***differences between batches***, machine set-up, etc.

- ***Trends***: structure characterized by data aligned upwards or downwards. It may be due to tool wear, ***machinery that does not keep the setting or differences in work shifts***.



E. Belluco, Guide to Statistical Process Control for Minitab, Franco Angeli (2013)

CASE STUDY 2

The finding of « patterns » such as those just shown on a control chart is an indication of the presence of identifiable causes of variability (or ***special causes***) that must be diagnosed and corrected.

The distinction between ***controlled variability*** and ***uncontrolled variability*** due, respectively, to ***common causes*** and ***special causes*** was introduced by W. Shewhart (1931).

W.A. Shewhart, Economic Control of Quality of Manufactured Product – Van Nostrand (1931)

CASE STUDY 2

There are several « criteria » for identifying « special causes » and the most common are summarized in the table opposite.

Each of these « criteria » describes a characteristic « pattern » that can be found in the control chart.

	Summary of typical « Special Cause » criteria
1	1 point more than 3 standard deviations from centerline
2	7 points in a row on same side of centerline
3	6 points in a row, all increasing or all decreasing
4	14 points in a row, alternating up and down
5	2 out of 3 points > 2 standard deviations form centerline (same side)
6	4 out of 5 points > 1 standard deviations form centerline (same side)
7	15 points in a row within 1 standard deviation of centerline (either side)
8	8 points in a row > 1 standard deviation from centerline (either side)

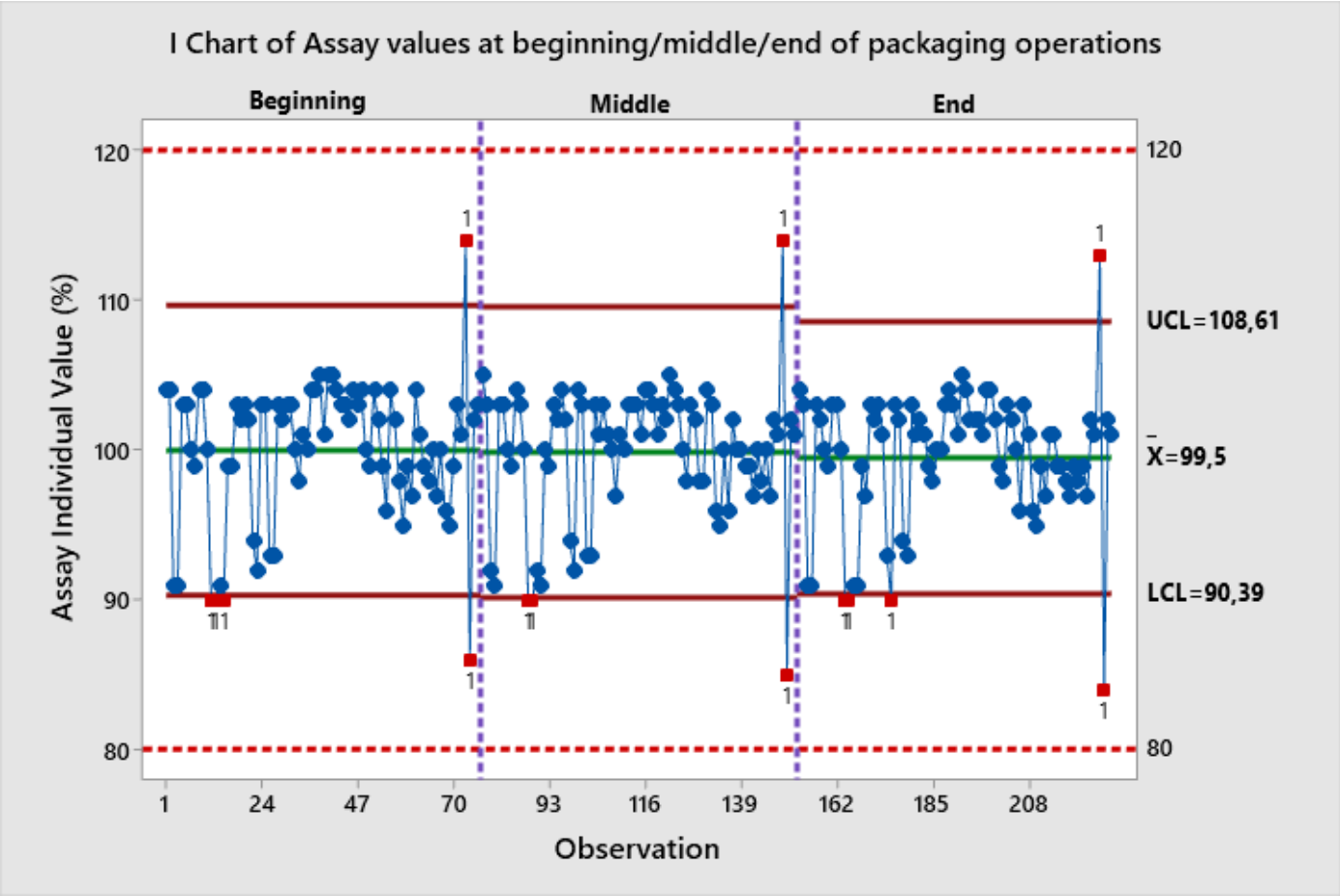
DaimlerChrysler, Ford, and General Motors, Statistical Process Control (SPC) Reference Manual, 2nd Ed., AIAG (2005)

AT&T Technologies, Inc., Statistical Quality Control Handbook, AT&T Technologies, Inc.(1984)

CASE STUDY 2

Let's now consider one-year production (*i.e.*, 76 lots) of a finished API and the assay values measured on samples taken at the beginning, middle and end of the packaging for each lot (*i.e.*, 228 data).

The whole bunch of data should provide information on the year production quality. For the sake of comparison these data are reported in an *I chart* like that shown here.



CASE STUDY 2

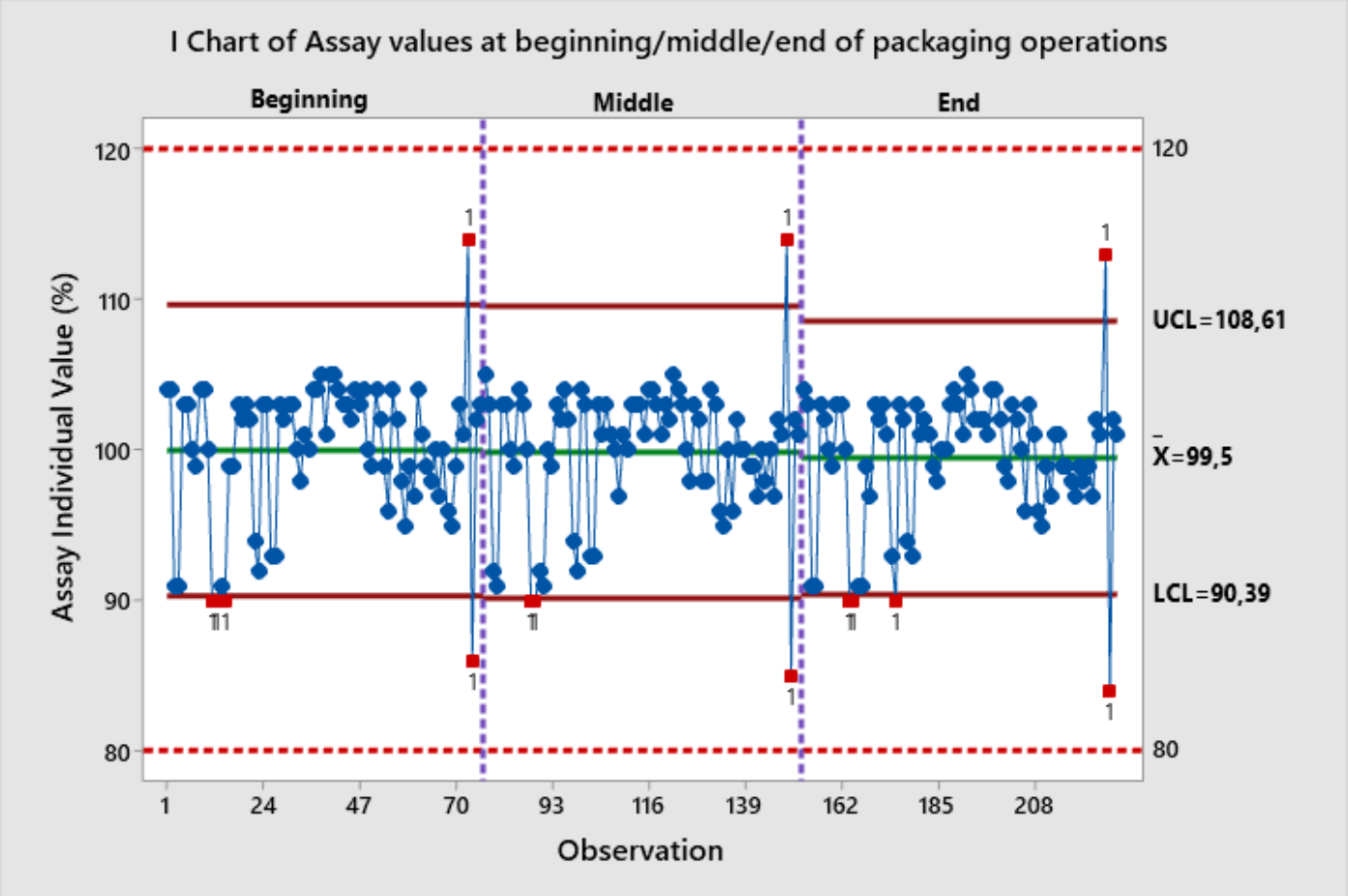
This control chart is informative, but it is too detailed.

IT LACKS SYNTHESIS !

Instead, it is typical of Statistics to synthesize information, as well as to describe it.

Descriptive Statistics uses graphs and several indices:

- Position (mean, median,…)
- Variability (variance,…)
- Shape (asymmetry,…)



CASE STUDY 2

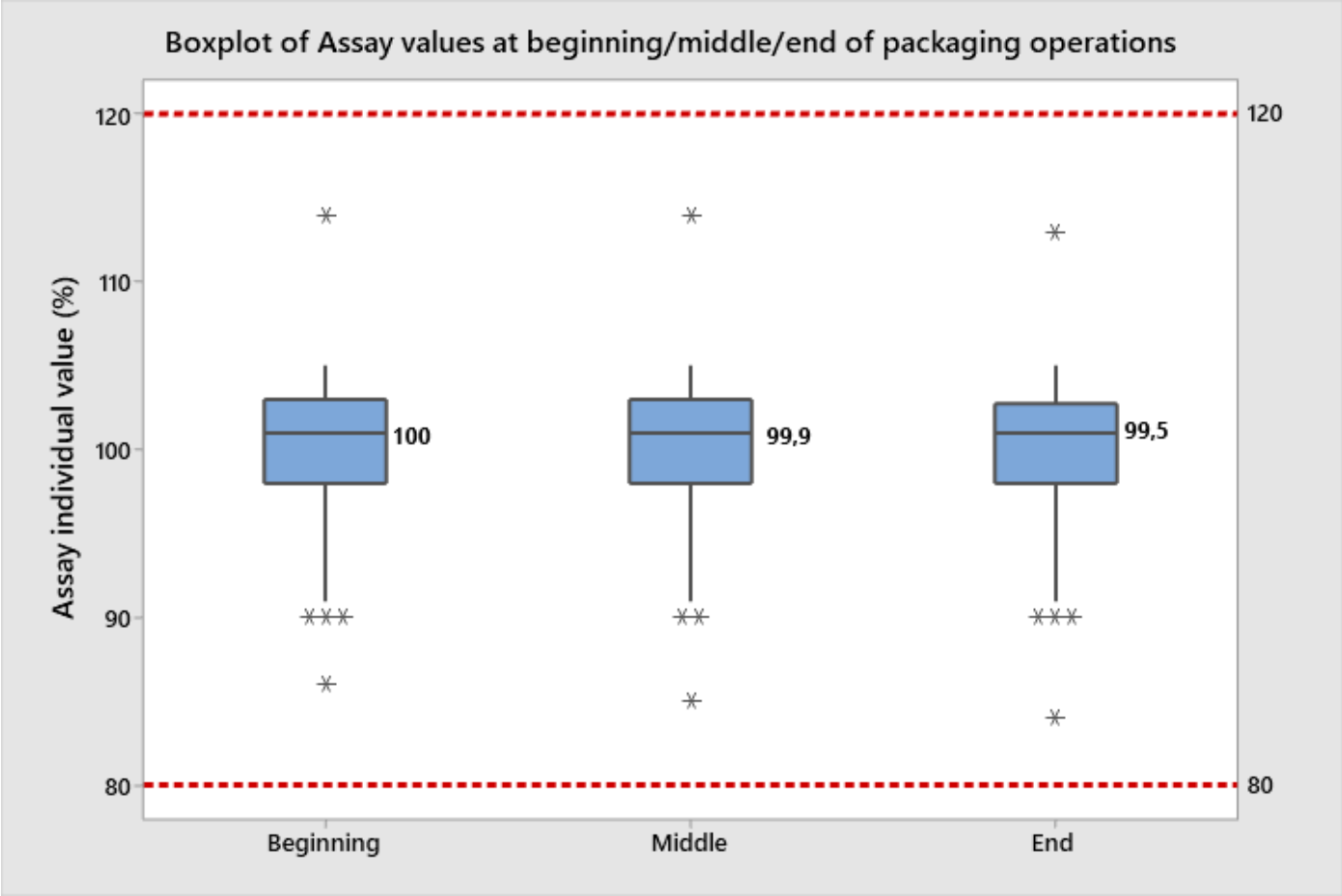
In this case a boxplot is much better as shown here on the side.

Boxplots are widely used as they have great visual impact and easy to understand

From this is evident a high level of constancy in production quality.

Average values are comparable !

The number of outliers and their arrangement indicate that it deals of real data ☺



CASE STUDY 2

The boxplot is a useful graphic technique that helps to visually compare the result obtained in the three cases but does not give any quantitative estimate of the differences existing among them.

More properly, the boxplot does not allow to establish whether, or not, there is a statistically significant difference between the three data groups. For that

ANOVA (Analysis of Variance)

CASE STUDY 2

With One-way ANOVA we want to test the hypothesis that all means are equal, namely that:

H_0: all means are equal *vs.*

H_1: not all means are equal

at a significance level of 5% ($\alpha = 0.05$)

Analysis of Variance

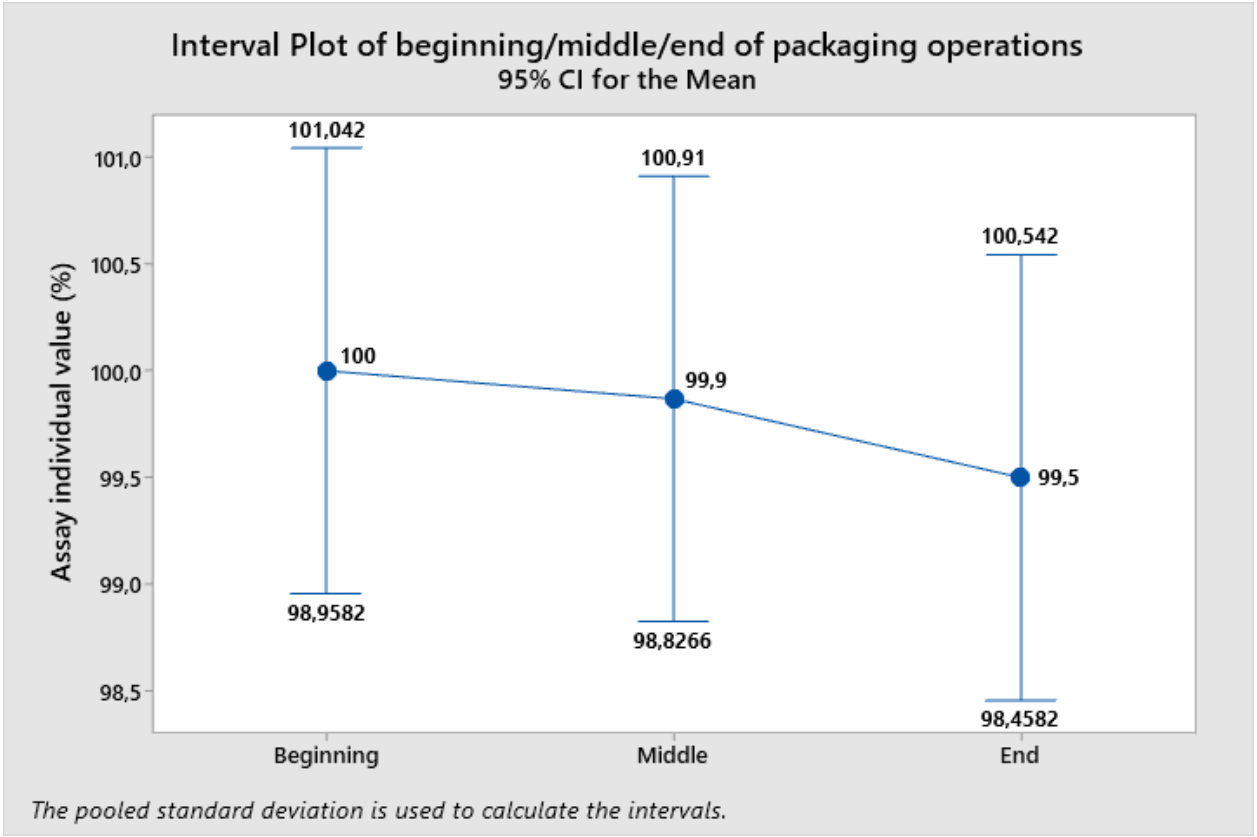
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	10,21	5,105	0,24	**0,787**
Error	225	4779,68	21,243		
Total	227	4789,89			

CASE STUDY 2

Means

Factor	N	Mean	StDev	95% CI
Begin	76	100,000	4,730	(98,958; 101,042)
Middle	76	99,868	4,550	(98,827; 100,910)
End	76	99,500	4,545	(98,458; 100,542)

Pooled StDev = 4,60902



Considering all findings (this and previous slide) there is no evidence to reject the null hypothesis, H_0

⇒ Means can therefore be considered not statistically different one from the other !

CASE STUDY 2

The t-test limits the comparison to just two groups of data, namely that:

H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ *vs.*
H_1: $\mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$
at a significance level of 5% ($\alpha = 0.05$)

In this respect let's consider two series of pH values, one determined in-house on real samples and the other reported on the corresponding certificates of analysis (CoAs) provided by the supplier (*i.e.*, Supplier) together with the samples.

	Sodium Acetate pH values	
	In-house	**Supplier's CoA**
Sample 1	8.1	8.1
Sample 2	8.3	8.1
Sample 3	8.2	8
Sample 4	8.5	8.4
Sample 5	8.5	8.4
Mean value	**8.32**	**8.2**

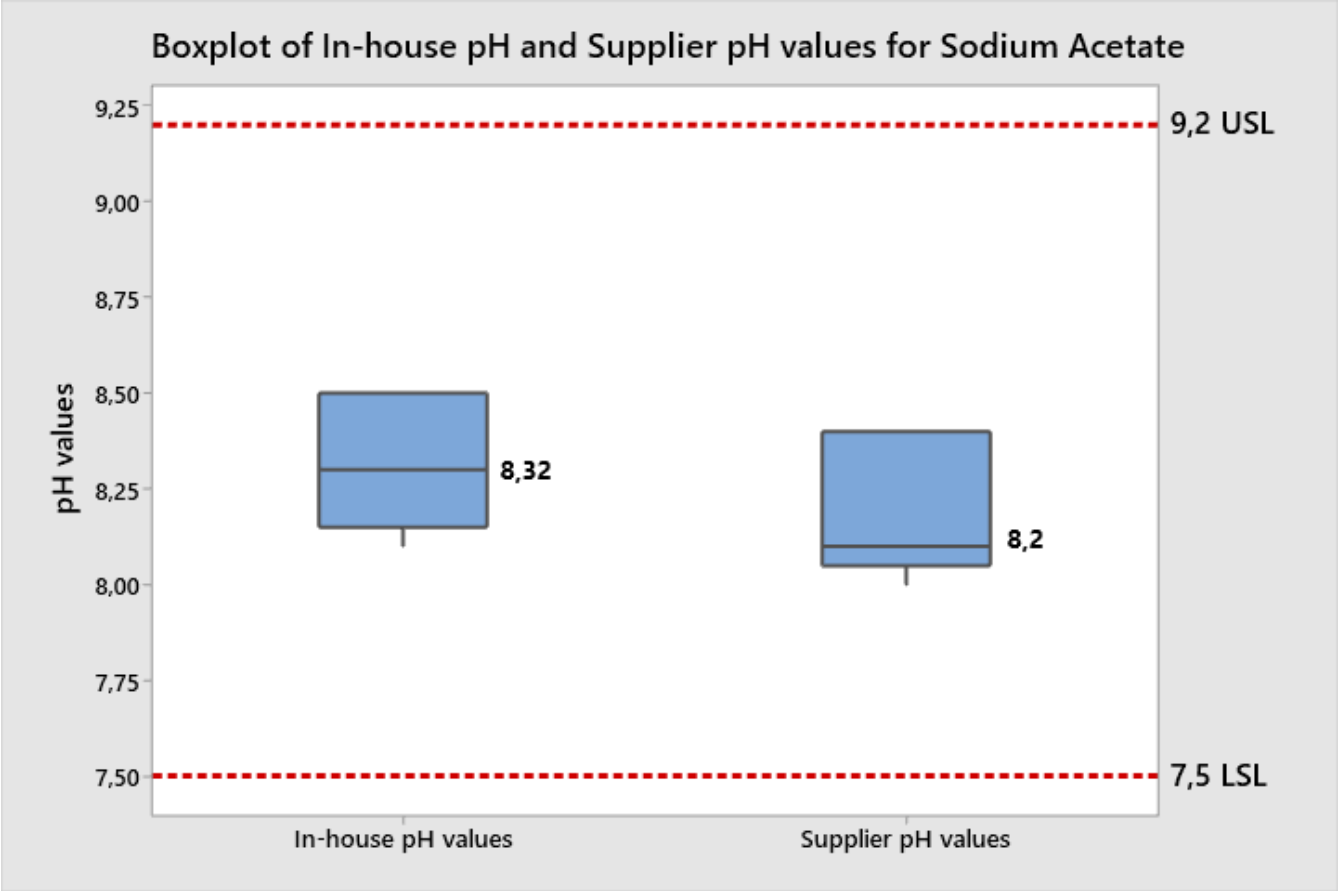
Are, on the average, the two series of values here above statistically different or not?

CASE STUDY 2

Let’s first look at data visualization using boxplots.

Box widths look rather similar, but, apart from this, we cannot say much more.

The ***t-test*** can tell us if the two mean values are statistically different or not.



CASE STUDY 2

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
In-house pH values	5	8,320	0,179	0,080
Supplier pH values	5	8,200	0,187	0,084

Estimation for Difference

Difference	95% CI for Difference
0,120	(-0,154; 0,394)

Test

Null hypothesis $H_o: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	**P-Value**
1,04	7	**0,334**

As *P-value* > 0.05, we fail to reject H_0 → **No means difference !**

CASE STUDY 2

The fact that there is no statistically significant difference between the average values of the two data groups suggests that, reasonably, there is no difference between the two methods of determining Sodium Acetate pH.

Instead, consider the data in the table here on the side. In this case, Sodium Acetate is provided by a different supplier (*i.e.*, Supplier 1).

	Sodium Acetate pH values	
	In-house	**Supplier's 1 CoA**
Sample 1	8.1	8.6
Sample 2	8.3	8.6
Sample 3	8.2	8.5
Sample 4	8.5	8.9
Sample 5	8.5	8.9
Mean value	**8.32**	**8.7**

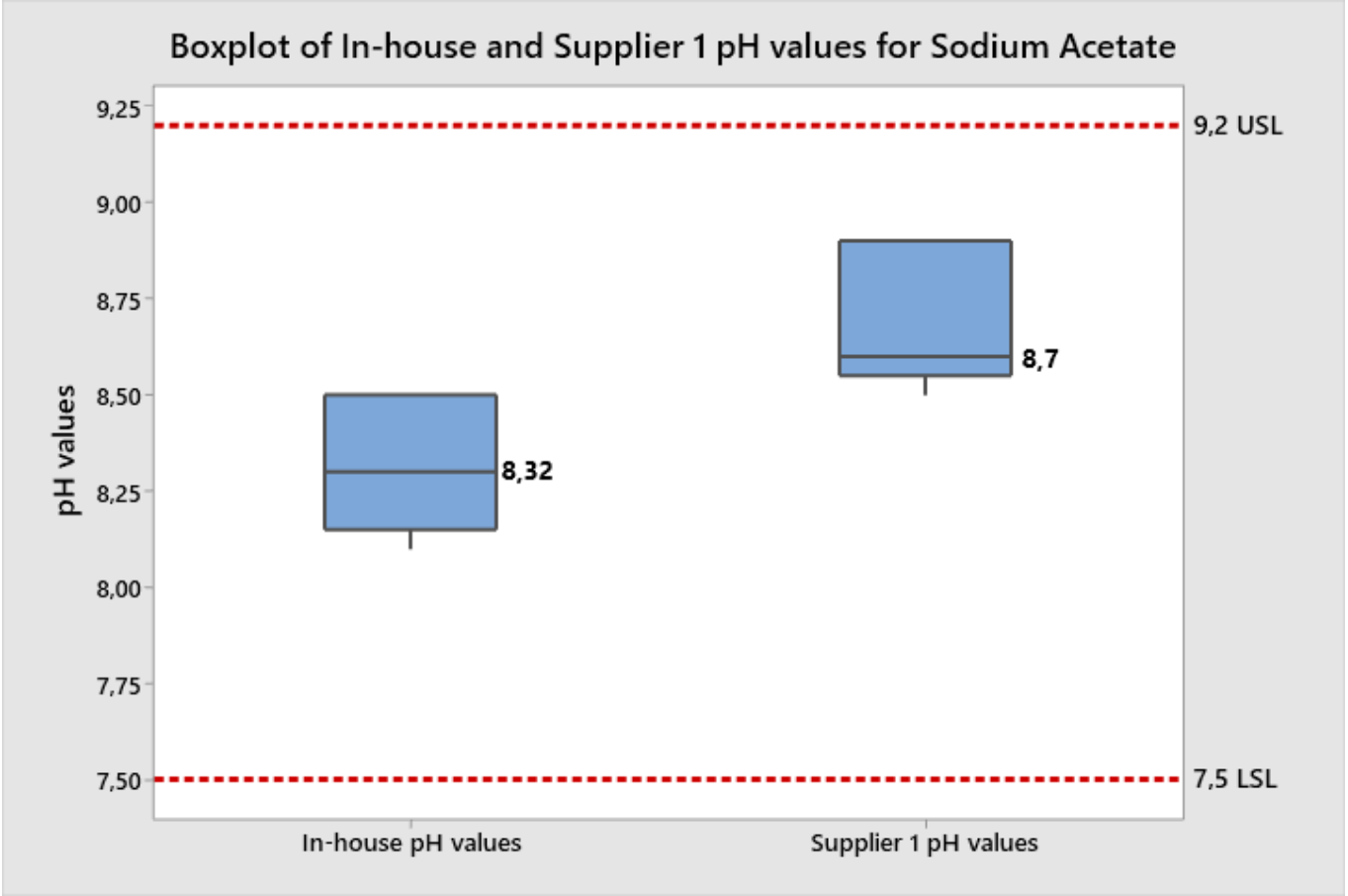
Are the two mean values here above reported, statistically different or not?

CASE STUDY 2

In this case it is evident that the two pH data distributions are shifted from each other.

However, box widths are comparable
⇨ data spreads are similar.

The ***t-test*** can tell us if the two mean values are statistically different or not while the F-test can tell us if data spreads are comparable or not.



CASE STUDY 2

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
in-house pH values	5	8,320	0,179	0,080
supplier 1 pH values	5	8,700	0,187	0,084

Estimation for Difference

Difference	95% CI for Difference
-0,380	(-0,654; -0,106)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-3,28	7	**0,013**

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

There is difference a statistically significant difference between the two mean values !

CASE STUDY 2

The fact that there is a statistically significant difference between the average values of the two data groups suggests that, reasonably, there is difference between the two methods of determining pH.

This finding is not so unusual if comparing data from different laboratories !

In such a case, even if the analytical techniques are different from each other, they should be of comparable precision and accuracy and therefore

test for 2 Variances : Determine whether the variances or standard deviations of two groups differ. You can use this test to compare the process variance before and after you implement a quality improvement program.

CASE STUDY 2

Descriptive Statistics

Variable	N	StDev	Variance	95% CI for σ
In-house pH values	5	0,179	0,032	(0,102; 0,518)
Supplier 1 pH values	5	0,187	0,035	(0,113; 0,510)

Test

Null hypothesis $H_0: \sigma_1 / \sigma_2 = 1$

Alternative hypothesis $H_1: \sigma_1 / \sigma_2 \neq 1$

Significance level $\alpha = 0,05$

Method	Test Statistic	DF1	DF2	P-Value
Bonett	0,02	1		**0,896**
Levene	0,00	1	8	**1,000**

As P-values > 0.05 there is no evidence to reject H_0 !

Data spreads can be considered comparable !

CASE STUDY 2 - BLAND-ALTMAN PLOT

- The Tukey Mean-Difference (TMD) plot in medical research is usually known as the Bland-Altman plot.
- It is used to compare two measurements of the same variable or two measurement techniques
- The Bland-Altman plot is formed by plotting the differences *X1* - *X2* on the vertical axis versus the averages (*X1*+*X2*)/2 on the horizontal axis.

A horizontal line representing the *bias* is drawn at $\bar{d}$

Additional horizontal lines, known as *limits of agreement*, are added to the plot at

$\bar{d} - 1.96\ Sd$ and $\bar{d} + 1.96\ Sd$

The *d*'s are the differences formed as *d* = *X1* - *X2*.

*D.G. Altman, J.M. Bland, Measurement in medicine: the analysis of method comparison studies – The Statistician, 32 (*1983) 307-317

CASE STUDY 2 - BLAND-ALTMAN PLOT

Let's now consider the two series of pH values, one determined in-house on real samples and the other reported on the corresponding CoAs provided by the Supplier discussed earlier.

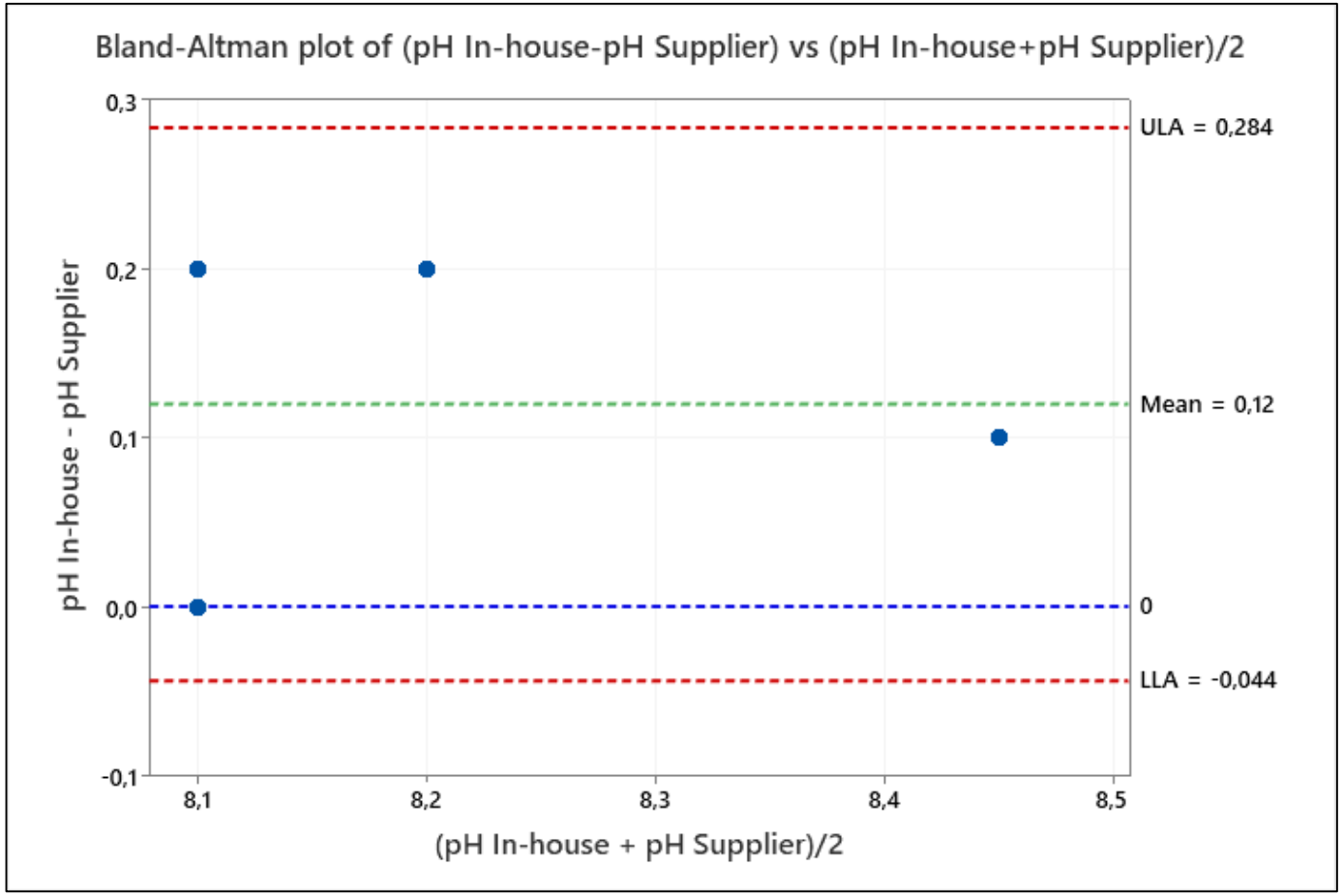
The corresponding Bland-Altman plot is shown in the next slide.

	In-house	Supplier's CoA	In-house - Supplier's CoA	(In-house + Supplier's CoA)/2
Sample 1	8,1	8,1	0,0	8,1
Sample 2	8,3	8,1	0,2	8,2
Sample 3	8,2	8	0,2	8,1
Sample 4	8,5	8,4	0,1	8,45
Sample 5	8,5	8,4	0,1	8,45
Mean value			**0,12**	
Standard Deviation (SD)			**0,0837**	
Mean - 1.96SD			-0,0440	
Mean + 1.96Sd			0,2840	

CASE STUDY 2 - BLAND-ALTMAN PLOT

This plot shows that:

- the four difference values (two are coincident) look randomly dispersed around the mean value (0.12)
- All data points are within the limits of agreement (LLA, ULA)
- The zero value is ***within*** the limits of agreement
- There is practically ***no bias*** as the mean value 0.12 ~ 0.0



CASE STUDY 2 - BLAND-ALTMAN PLOT

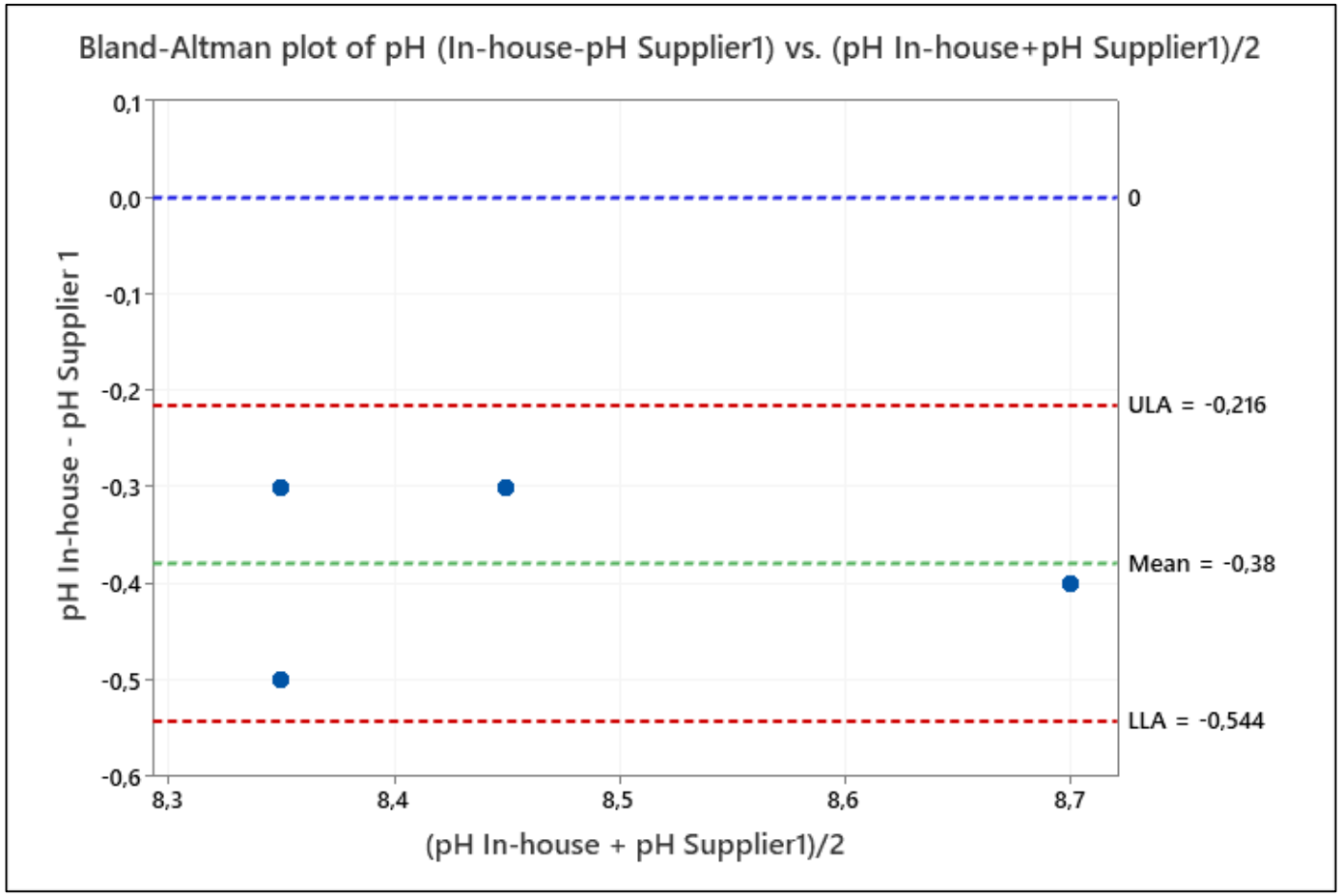
Let's now consider the two series of pH values pertinent to Supplier 1 and repeat the same procedure. The corresponding Bland-Altman plot is shown in the next slide.

	In-house	Supplier's CoA	In-house - Supplier's CoA	(In-house + Supplier's CoA)/2
Sample 1	8,1	8,6	-0,5	8,35
Sample 2	8,3	8,6	-0,3	8,45
Sample 3	8,2	8,5	-0,3	8,35
Sample 4	8,5	8,9	-0,4	8,7
Sample 5	8,5	8,9	-0,4	8,7
Mean value			**-0,38**	
Standard Deviation (SD)			**0,0837**	
Mean - 1.96SD			-0,5440	
Mean + 1.96Sd			-0,2160	

CASE STUDY 2 - BLAND-ALTMAN PLOT

This plot shows that:

- the four difference values (two are coincident) look randomly dispersed around the mean value (-0.38)
- All data points are within the limits of agreement (LLA, ULA)
- The zero value is ***out*** of the limits of agreement
- The mean value is the ***bias*** in fact: -0.38 ≠ 0.0



CASE STUDY 2 - BLAND-ALTMAN PLOT

To summarize we can say that:

- The Bland-Altman plot is used to compare two measurements of the same variable
- There are three study designs that can be analyzed by this procedure:
    1. exactly one data pair *per* observation (no repeatability parameter is computed)
    2. multiple replicates for each method, no pairing
    3. multiple replicates for each method obtained as pairs

  The examples shown here belong to type 1.

- Since:
  - it not necessary that measurements themselves follow a normal distribution
  - a non-normal distribution of differences may not be as serious here as in other statistical contexts

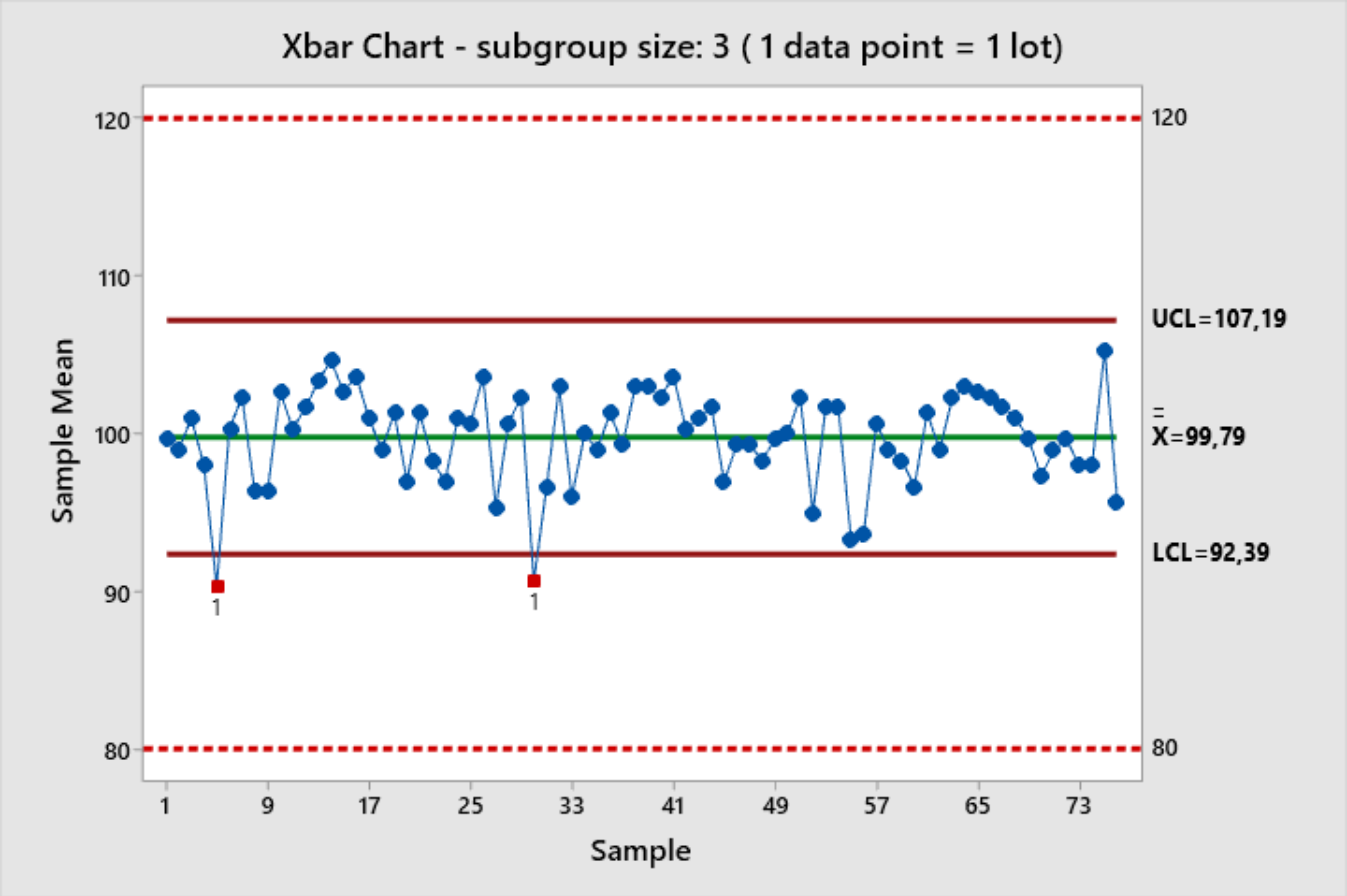
  the Bland-Altman plot is a useful tool to compare data that often meet these requirements (*e.g.*, **related substances contents**).

CASE STUDY 2

Let's come back to the control chart and let's group data in subgroups of each consisting of three values. We now have: 1 data point (*i.e.*, average of three values) *per* lot.

Xbar Chart

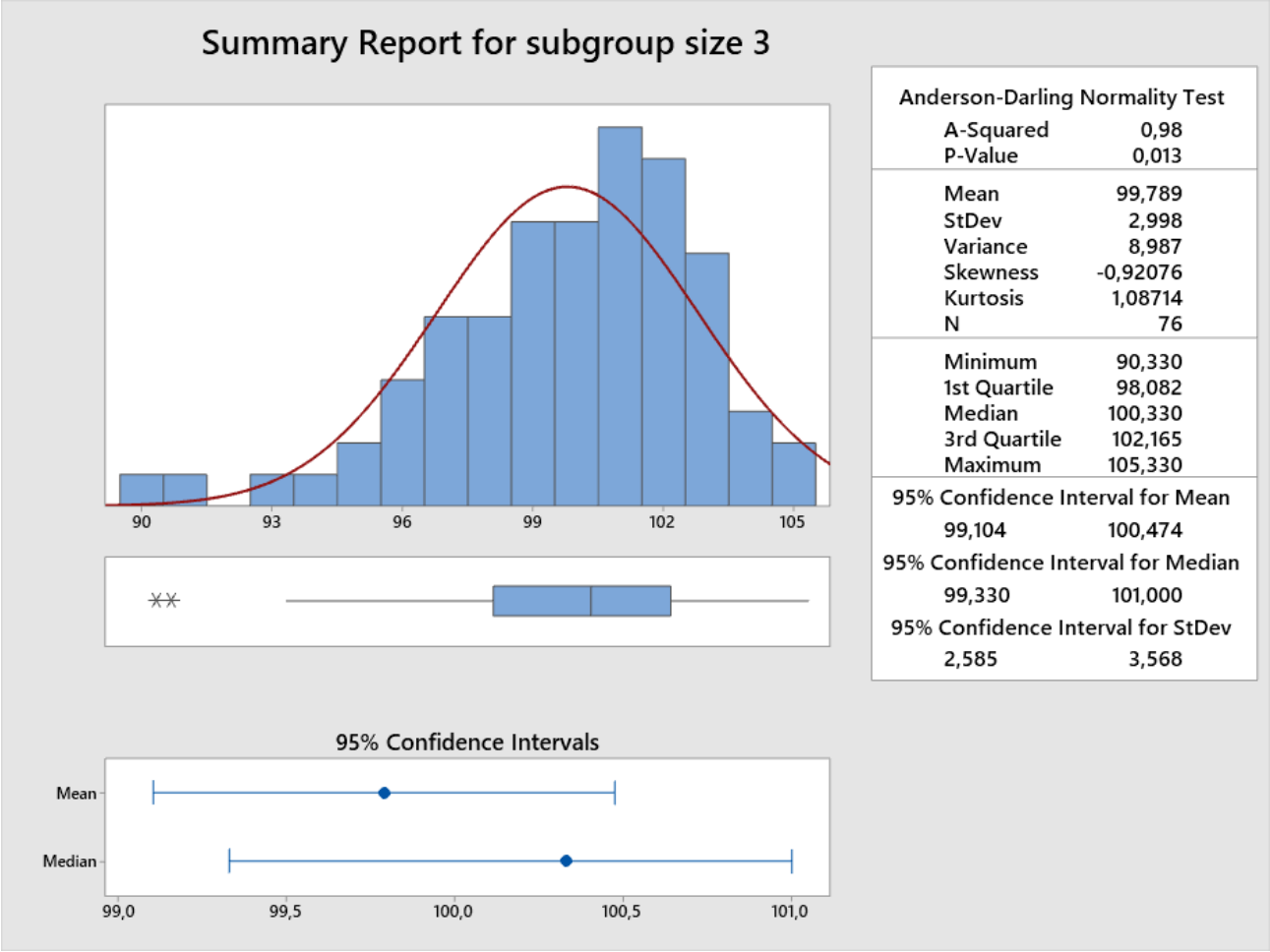
This chart is simpler and more informative than the previous one.



CASE STUDY 2

Interestingly, data points plotted in the *Xbar* chart look nearly normally distributed even though *P-value* < *0.05*.

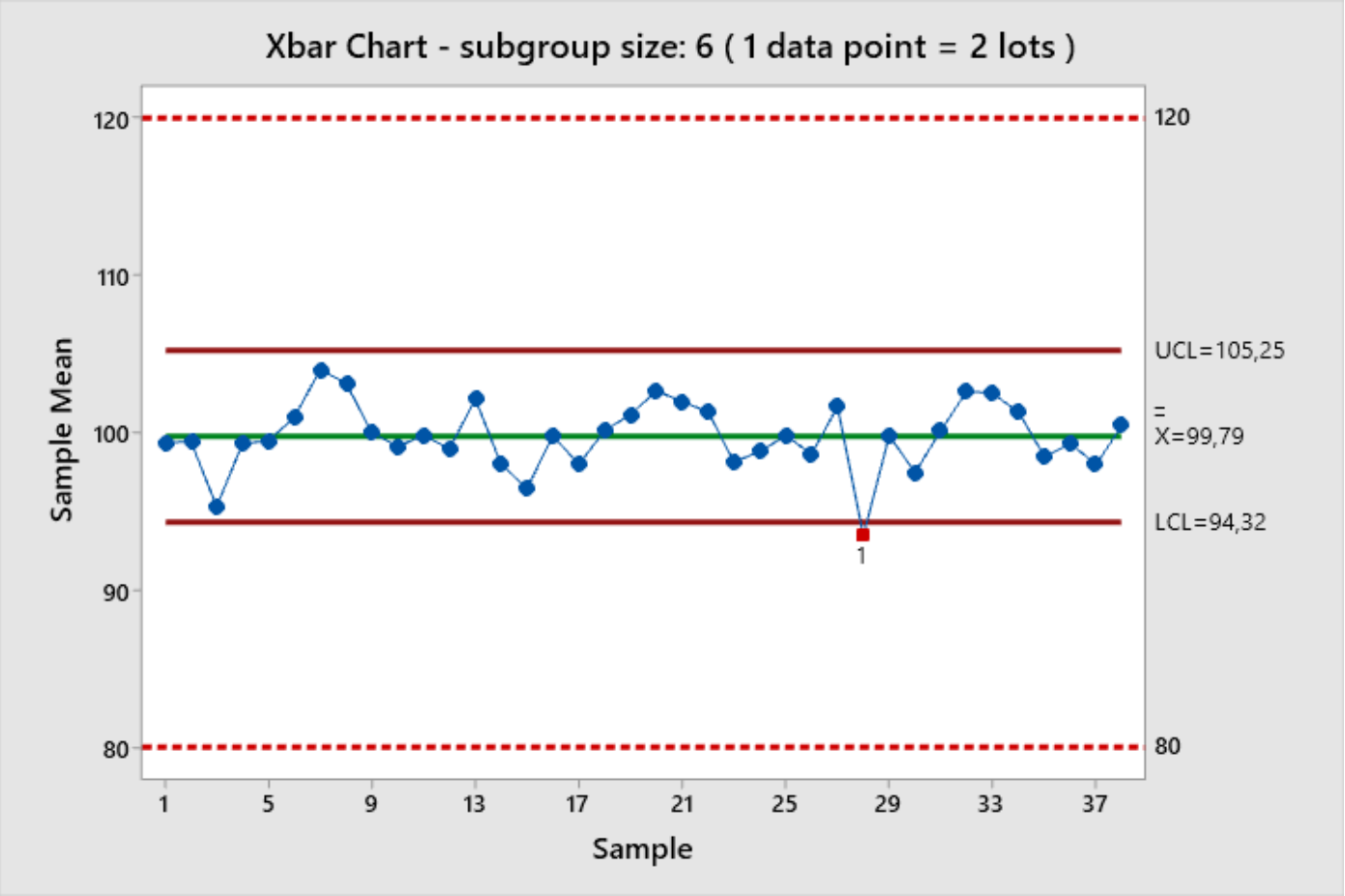
Let's see what happens grouping data in groups of six, *i.e.*, 1 data point *per* 2 manufactured lots.



Anderson-Darling Normality Test	
A-Squared	0,98
P-Value	0,013
Mean	99,789
StDev	2,998
Variance	8,987
Skewness	-0,92076
Kurtosis	1,08714
N	76
Minimum	90,330
1st Quartile	98,082
Median	100,330
3rd Quartile	102,165
Maximum	105,330
95% Confidence Interval for Mean	
99,104	100,474
95% Confidence Interval for Median	
99,330	101,000
95% Confidence Interval for StDev	
2,585	3,568

CASE STUDY 2

As expected, the *Xbar* shown here is much simpler than the previous one and the data trend look very smooth and within the upper and lower control limits.



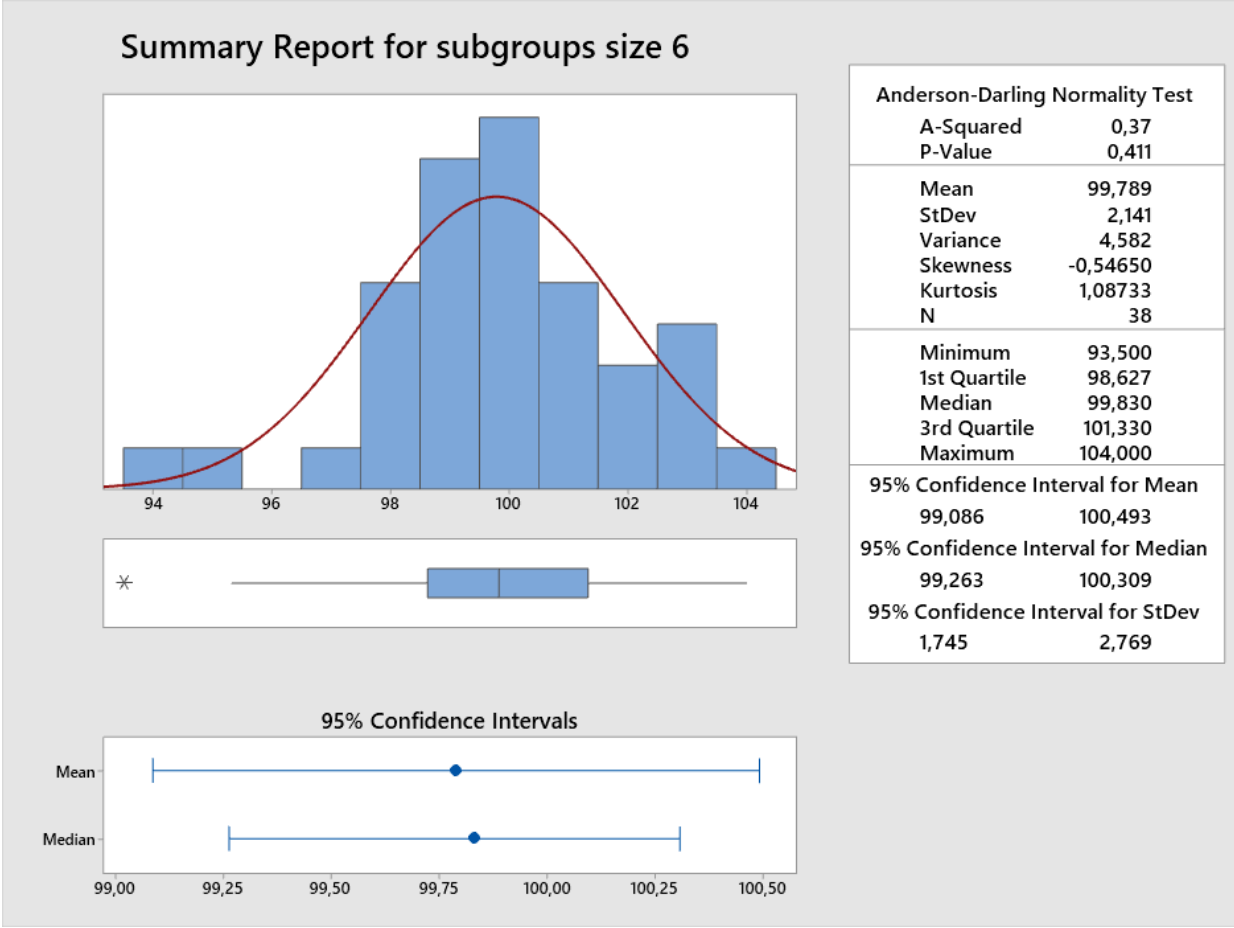
CASE STUDY 2

Data points display now a nicely shaped normal trend as shown here on the right.

This behavior is not unexpected. In fact, it reflects the meaning of the

Central Limit Theorem

one of the most important results in Probability Theory !



Anderson-Darling Normality Test	
A-Squared	0,37
P-Value	0,411
Mean	99,789
StDev	2,141
Variance	4,582
Skewness	-0,54650
Kurtosis	1,08733
N	38
Minimum	93,500
1st Quartile	98,627
Median	99,830
3rd Quartile	101,330
Maximum	104,000
95% Confidence Interval for Mean	
99,086	100,493
95% Confidence Interval for Median	
99,263	100,309
95% Confidence Interval for StDev	
1,745	2,769

CASE STUDY 2

In simple terms the **Central Limit Theorem** states that, under certain conditions, the sum of a large number of random variables is approximately normal.

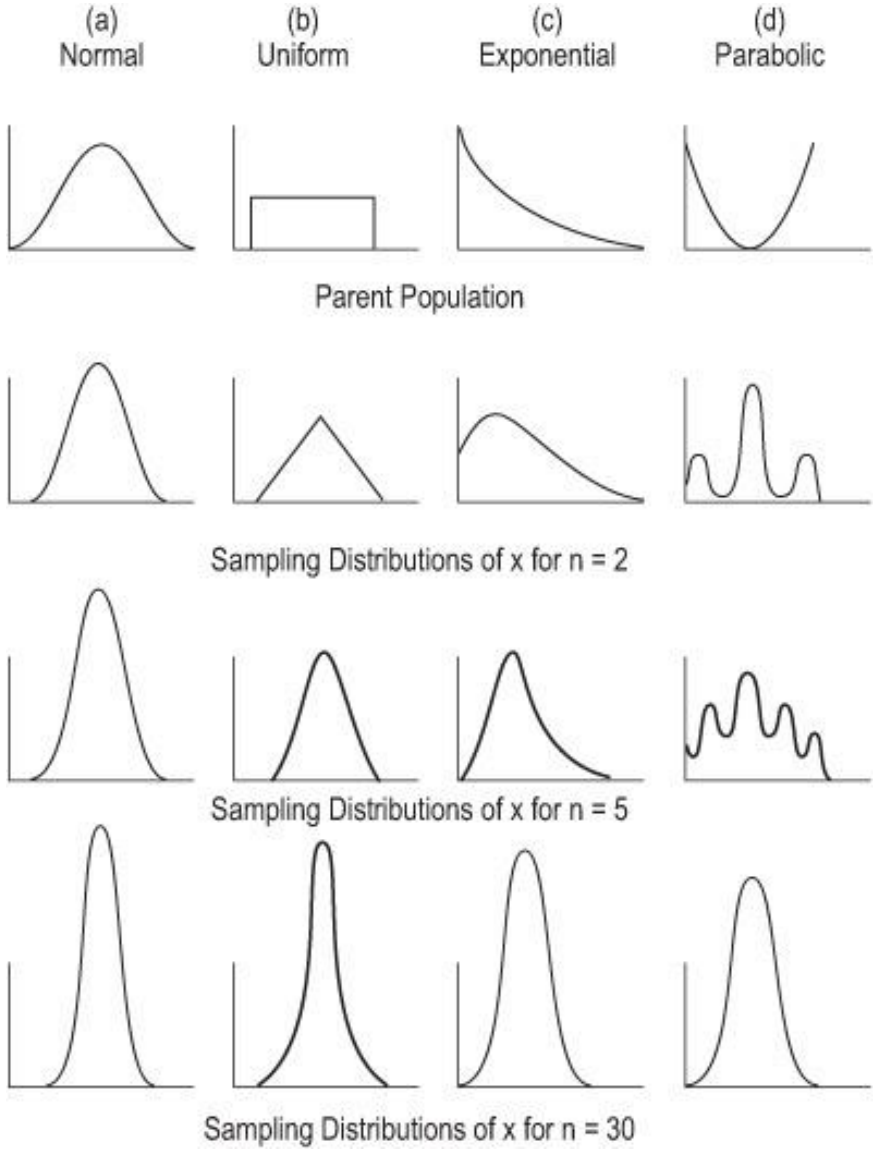
As normality is a requirement for many statistical tests and since many data sets are not normally distributed, it would be convenient converting non-normal data into something that does have a normal distribution.

- The distribution of the ***averages*** (***Xbars***) approaches normality if big enough samples are considered
- This finding reflects the **Central Limit Theorem**
- Calculating averages on subsets of data is therefore a common practice when the underlying data distribution is not normal

CASE STUDY 2

A practical and very important consequence of the **Central Limit Theorem** is that *regardless of the shape of parent population, the distribution of means quickly approaches the normal distribution* as shown here.

Obviously, data can be grouped, but a rational is needed. In this case we had 3 data points (beginning/middle/end) for each lot of API.



M.L. George et al., The Lean Six Sigma Pocket Toolbox, McGraw Hill (2005)

CASE STUDY 2

Until now we have seen three main types of control charts:

- ***I chart***
- ***Xbar chart***
- ***R chart***

These are just a few types of control charts as many other are available, however, for the sake of simplicity it is worth to focus just on these and understand how they can help.

CASE STUDY 2

- *Xbar chart*
  - shows the changes in the average values of the process
  - it displays long-term variability
  - shows if the variability between the subgroup means is greater than can be expected by observing the variability within the subgroups

- *R chart*
  - shows short-term variability
  - shows whether the variability within the subgroups is consistent, stable, or very different between the subgroups

CASE STUDY 2 - CONCLUSIONS

- what in the previous slide provides the rationale for why two control charts (*i.e.*, *Xbar* and *R*) are usually needed !
- for a process to be “under control”, data points must fall within the control limits on both control charts !
- for a good functioning of the *Xbar-R chart* is essential to choose the subgroups appropriately.

  In general, groups consisting of 4 – 5 observations each are considered adequate.

  In the case of an overall number of observations equal to 100, this would lead to 20 – 25 subgroups and in that case is reasonable to expect a normal distribution for average values.