Bachelor's thesis

Information and Communication Technology

MICTIS15

2017

Marko Järvenpää

DEVELOPING A SAMPLING SYSTEM FOR QUALITY CONTROL TESTS BASED ON STATISTICAL DATA

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Information and Communications Technology

2017 | 50 pages Instructor: Hazem Al-Bermanei (Ph.Lic)

Marko Järvenpää

DEVELOPING A SAMPLING SYSTEM FOR QUALITY CONTROL TESTS BASED ON STATISTICAL DATA

The main objective of this thesis was to investigate sample plans in quality control and introduce the statistical software Minitab to BonAlive Biomaterials Ltd, Class III medical device manufacturer. Sampling and inspection of sample items is time-consuming and in some cases samples cannot be reused. Hence, sample size is kept to a minimum. On the other hand, the amount of the analysed items should be statistically representative of the entire batch. The goal of this thesis was to find the best possible combination between these two specifications.

The thesis is largely based on the interpretation of ISO standards and the fitting of these to processes of the client company. Minitab 17 has used as a statistical tool.

As a result of the thesis work it is showed that the smallest sample size can be reached when a variable, i.e., numeric format data can be measured. This is especially true when the standard deviation of the process is known. If variable data cannot be used and attribute data methods are used, then sequential sampling gives the smallest average sample size. The risk level that a lot with nonconforming items is accepted does not increase or decrease if the sample plan is changed between single, double, multi, sequential or variable sampling plans. In addition, this thesis presents bulk material sampling and the sample size for that. The results and the company specific processes are classified information and not presented in the thesis.

As a result of the thesis the client company can reduce sample sizes and based on the given training start to use Minitab as a statistical tool for different process analyses.

KEYWORDS:

sample plan, sample system, attribute data, variable data, Minitab

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tieto- ja viestintätekniikka

2017 | 50 sivua

Ohjaaja: FL Hazem Al-Bermanei

Marko Järvenpää

TILASTOLLISEEN DATAAN PERUSTUVAN NÄYTTEENOTON KEHITTÄMINEN LAADUNTARKASTUKSESSA

Tämän opinnäytetyön tavoitteena oli tutkia laaduntarkastuksen näytteenottosuunnitelmia ja esitellä Minitab-tilastotyökalun käyttöä BonAlive Biomaterials Ltd:lle. Näytteiden ottoon ja tarkastukseen kuluu paljon aikaa eikä tuotteita tämän jälkeen voida kaikissa tapauksissa enään hyödyntää. Tästä syystä näytteiden määrä pyritään pitämään mahdollisimman pienenä. Toisaalta näytemäärän tulee tilastollisesti edustaa koko tuotantoerää tai muuta vastaavaa erää. Opinnäytetyössä etsitään näiden vaatimuksien mukainen paras mahdollinen yhdistelmä.

Opinnäytetyö perustuu pitkälti ISO-standardien soveltamiseen asiakasyrityksen prosesseihin. Tilastollisena työkaluna on käytetty Minitab-ohjelmaa.

Tuloksena esitetään, että pienimpään näytekokoon päästään, kun mitattavalla suureella on numeerinen muuttuja ja tuotannon keskihajonta on tunnettu. Mikäli numeerisen muuttujan käyttö ei ole mahdollista ja käytetään kategorista muuttujaa, niin keskimäärin pienimpään näytekokoon päästään sarjallisella (Sequential) näytteenotolla. Riskitaso sille, että asiakas saa viallisen tuotteen on käytännössä sama kaikkien käsiteltyjen näytteenottometodien välillä. Lisäksi työssä esitetään ratkaisu massatavaroiden, kuten nesteiden ja jauheiden, homogeenisuuden toteamiseen tarvittavaan näytemäärään.

Tulosten perusteella asiakasyritys pystyy pienentämään näytekokoja ja hyödyntämään Minitab-ohjelmaa prosessien kyvykkyyksien seurantaan.

ASIASANAT:

näytteenotto, kategorinen muuttuja, numeerinen muuttuja, Minitab

ACKNOWLEDGEMENT

I would like to thank the personnel of BonAlive Biomaterials Ltd for their support and for providing guidance concerning the processes. Special thanks go to B.Sc. Heidi Tuomi for her diligent support. I am gratefully indebted to her for her very valuable guidance on this thesis.

I would also like to thank my thesis advisor Ph.Lic. Hazem Al-Bermanei and M.Sc. Tiina Ferm for her critical and helpful comments. Both are lecturers in the Department of Information Technology at Turku University of Applied Sciences.

27. March 2017 Kaarina

Marko Järvenpää

CONTENTS

FIGURES

EQUATIONS

TABLES

LIST OF ACRONYMS, ABBREVIATIONS AND SYMBOLS

Ac	Acceptance number (ISO 2859-5)
AQL	Acceptance Quality Limit, worst tolerable quality level (ISO 3534-2)
C_p	Process capability index for a process in a state of statistical control. For normal distribution: $C_P = \frac{USL-LSL}{6\sigma}$ (ISO 3534-2)
C_{pkL}	Lower process capability index describing process capability in relation to the lower specification limit, LSL. For normal distribution: $C_{pkL} = \frac{\bar{X}-LSL}{3\sigma}$ (ISO 3534-2)
C_{pkU}	Upper process capability index describing process capability in relation to the upper specification limit, USL. For normal distribution: $C_{pkU} = \frac{(USL-\bar{X})}{3\sigma}$ (ISO 3534-2)
C_{pk}	Minimum process capability index. Smaller of upper process capability index and lower process capability index (ISO 3534-2)
DPMO	Defects Per Million Opportunities. The number of defects in a sample divided by the total number of defect opportunities multiplied by 1 million. (Minitab)
FDA	Food & Drug Administration, United States of America (FDA)
f_s	Factor that correlates the maximum sample standard deviation to the difference between USL and LSL (ISO 3951-1)
f_σ	Factor that correlates the maximum process standard deviation to the difference between USL and LSL (ISO 3951-1)
K_p	Upper portion of the standard normal distribution (ISO 10725)

k	Form k acceptability constant for single specification limit (ISO 3951-1)
LSL	Lower Specification Limit that defines the lower limiting value (ISO 3534-2)
MSSD	Maximum Sample Standard Deviation (ISO 3951-1)
MPSD	Maximum Process Standard Deviation (ISO 3951-1)
N	Lot size (number of items in a lot) (ISO 3951-1)
n	Sample size (number of items in a sample) (ISO 3951-1)
Nonconformity	Non-fulfilment of a requirement (SFS-EN ISO9000)
P_p	Process performance index describing process performance in relation to specified tolerance for a process that may not be in a state of statistical control. For normal distribution: $P_P = \frac{USL-LSL}{6S}$ (ISO 3534-2)
P_{pk}	Minimum process performance index, smaller of upper process performance index and lower process performance index (ISO 3534-2)
P_{pkL}	Lower process performance index describing process performance in relation to specified tolerance. For normal distribution: $P_{pkL} = \frac{\bar{X}-LSL}{3S}$ (ISO 3534-2)
P_{pkU}	Upper process performance index describing process performance in relation to specified tolerance. For normal distribution: $P_{pkU} = \frac{(USL-\bar{X})}{3S}$ (ISO 3534-2)
PPM	Parts Per Million. The number of nonconforming parts out of a million parts. (Minitab)

R^2	R-squared (R-sq) is a statistical measure of how near the data are to the fitted regression line. How well the regression model explains variation compare to total variation. (Minitab)
Re	Rejection Number (ISO 2859-5)
S	Standard Deviation, sample statistic (ISO 3534-2)
S^2	Sample variance, sum of squared deviations of variables in a sample from their mean divided by the number of terms in the sum minus one. $S^2 = \frac{1}{n-1} \sum_{i-1}^{n} (x_i - \bar{x})^2$ (ISO 3534-1)
USL	Upper Specification Limit that defines the upper limiting value (ISO 3534-2)
Variation	Difference between values of a characteristic. Variation is often expressed as a variance or standard deviation (ISO 3534-2)
$\bar{X}$	Process mean (ISO 3534-2)
α	Producer's risk (ISO 3534-2)
β	Customer's risk (ISO 3534-2)
σ	Standard Deviation, population (ISO 3534-2)

1 INTRODUCTION

The commissioner of the thesis is BonAlive Biomaterials Ltd, Turku. The company manufactures implantable medical devices for bone regeneration. The products are level III medical devices and therefore under strict control of the relevant authorities. The research question is, what are the right sample sizes for the commissioner's purposes? Sampling should be economical as the inspections are partly destructive but must be also statistically relevant.

Several books, publication and internet pages dealing with sample sizes. The mere FDA's medical device section web page shows 290 hits for "sample size"-search (FDA). An approach, what is taken to find the right sample sizes for the commissioner's purposes, is to resort to ISO standards. The reasons for selection of standards are that those are partly already in use and it is easy to justify them to the authorities.

Another task is to introduce a statistical software. Currently MS Excel is the used tool but it is not proven to be good enough for statistical purposes. Owing to price, availability and experience of the thesis worker Minitab 17, a statistical software, is chosen. The used tests can be presented also with Microsoft Excel but Minitab 17 is much more visual and versatile for statistical purposes so that is used.

It is crucial to know how accurate and precise the used measurement system is. Also to understand is possible variations due to the measurement system (gage) or due to the measurer, the measurement system analysis is done with Minitab. The thesis covers also other Minitab features, which are suitable for the client company. Minitab video training material is created but not presented in the thesis.

What the company is doing is familiarized by reading work procedures, work instructions and process descriptions. Other way is to go through the whole material flow from incoming raw materials to outgoing ready merchandises and interview the staff about the processes. The questions asked are like what they are doing, why they are doing that, when they are doing that, how they are doing that. Special interest is focused on measurements. What measurement instruments are used if any, is the instrument itself recording the result or is the human operator reading and recording the result? Where are the results saved? Why a measurement is done and for what purposes the result is used? If sampling is used, on what sampling rules are based on? The interview answers are related to the company's processes

and are not presented in the thesis.

Production and quality control activities are within the scope of the thesis. All company specific information, test results, procedures, processes etc. are confidential and thus not presented in the thesis. The pictures and the examples are from artificial test cases.

2 LITERATURE REVIEW OF SAMPLING SYSTEMS

2.1 Standard deviation

Standard deviation is widely used in statistic. It informs how widely data has spread from the mean and it is square root of variance (ISO 3534-1, 2012: 9).

$$s = \sqrt{\frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2}$$

Equation 1. Sample standard deviation

Note, in Equation 1, if the whole population is under investigation, then the symbol σ is used for the population standard deviation and the small letter n is replaced with the capital letter N.

In Equation 1, n is sample size, $\bar{x}$ is sample mean and x_i is value of an individual item. A little bit simplifying: $x_i - \bar{x}$ identifies the distance, how far away an individual item is from the mean. Then distances of all items are summed and finally calculated average value.

When the data is normally distributed, standard deviation shows the following information:

- Around 68% of data are within ±one standard deviation of the mean
- Around 95% of data are within ±two standard deviations of the mean
- Around 99,73% of data are within ±three standard deviations of the mean
- 5,15 standard deviations cover 99% of the data

Thus, if the mean of the data is 10 and the standard deviation is 1, it can be expected that about 68% of values in the range 9…11, and 95% to be in the range 8…12. Anything below 7 or above 13 is very rare. (Tennant, 2001: 13)

.

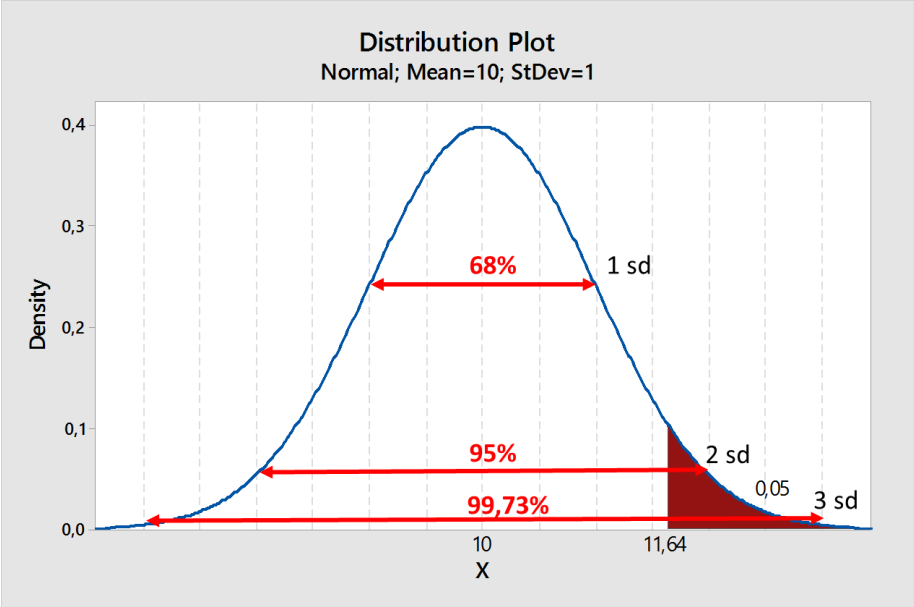


Figure 1. Normally distributed data and standard deviation.

In the Figure 1 is presented how much standard deviations reflects to the normally distributed data, like 99,73% of the data is inside $\pm 3\sigma$. Red tail: If producer's risk α is chosen to be 0,05 → from distribution table z is 1,645 (Apostol, 1969: 536). Meaning that 95% (= 1 – 0,05) of the data has smaller value than 11,645 (= 10 + 1,645). (Minitab 2017)

2.2 Confidence level and interval

Confidence level and interval tells how sure we are that certain data lies between certain limits. The formula for confidence interval is

$$\bar{X} \pm z \times \frac{\sigma}{\sqrt{n}}$$

Equation 2. Confidence interval.

An example could be that we measure length of 100 rods and find that the mean is 10 centimetres. Standard deviation is 1 cm. We like to get 95% confidence interval. From the distribution z-table can be found that 95% confidence level z value is 1,96 (Apostol, 1969: 536) Thus,

$10 \pm 1{,}96 \times \frac{1}{\sqrt{100}} = 9{,}804$ and 10,196

So, it could be said that based on the sample data, we are 95 % confident that the mean length of the rods is between 9,804 and 10,196. (Smithson, 2003: 4-6)

2.3 Population, sample and sampling

Population is totality of items under discussion. It includes every item and could be finite or infinite. Examples are the whole production lot made during the period and incoming raw material batch. Whereas sample is subset of a population. It could consist of one or more sampling units. A person could be sampling unit, one thousand inhabitants form the sample and the population is the inhabitants of Turku. Sampling is an action when a sample is taken or chosen. (ISO 3534-2, 2012: 5-7)

The reason for sampling could be explained with another example. A medical company is receiving 10 000 pieces' shipment of syringes from a supplier. Instead of measuring dimensions of every item, which is population, only a portion of the syringes, which is the sample, could be measured. Depending on risk level the company is willing to take, it could be as low as tens of pieces that need to be measured. So, savings of time and money are remarkable. (ISO 3951-1, 2013: 56-60)

2.4 Sampling systems

A sampling plan is a set of rules by how a batch should be tested. A sampling scheme is collection of sampling plans with switching procedures. A sampling system is a combination of sampling schemes with criteria by how right schemes should be chosen. The measurements can be categorized to attribute or variable measurements. (BS 6001-0, 1996: 35)

2.4.1 Attribute data

Attribute data can be arranged into categories like binary types "conforming" or "nonconforming", "pass" or "fail". It could be also categorical types like day of the week or type of product (BS 6001-0, 1996: 3). Even rating scales and numbers can be treated as attribute data but when the number of categories increase in attribute data, it becomes more variable data. Attribute data can be called also discrete data. An example could be, did

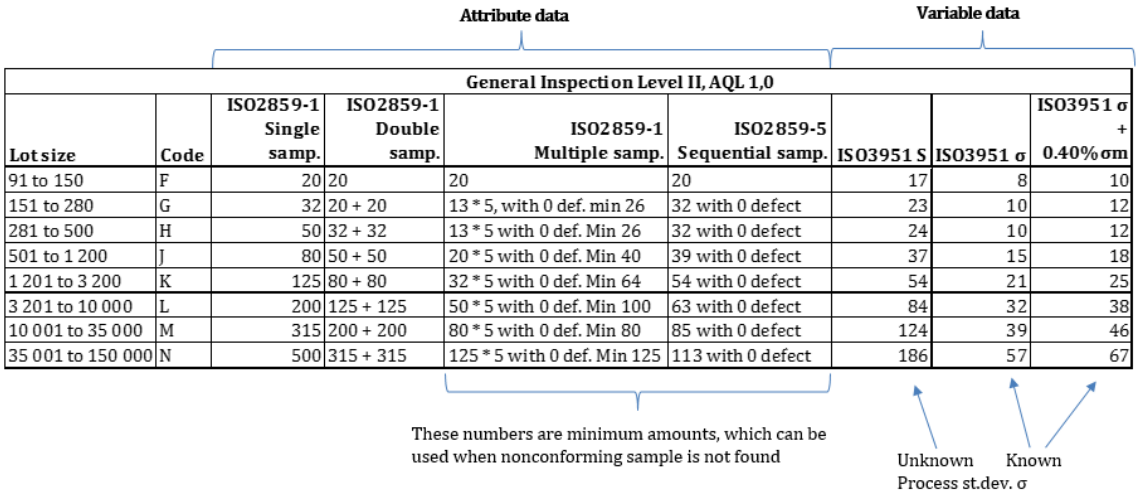
pedestrians stay on the pier or did they drop while walking from one end to another end of the pier?

2.4.2 Variable data

Variable data can get numeric value for instance length, weight, and temperature (BS 6001-0, 1996: 3). Variable data can be called also quantitative or continues data. It can utilize control charts like X-bar chart for average, R-chart for range between minimum and maximum values and S-chart for standard deviation. With variable data, it cannot only say did pedestrians drop but also where they actually walked. How far from the centre line of the pier and the edges of the pier. If people mainly walk between the left edge and the centre line that could give a hint that maybe the pier is left tilted and an improvement action should be taken before someone drops.

2.4.3 Choosing between attributes and variable methods

There are pros and cons for the both data types. Attribute data is easier to understand and faster to evaluate and needs less skills but on the other hand it usually needs larger sample size. Variable data has better discrimination and generally smaller sample size. It gives better view of the process state and thus it could give earlier warning if the process is slipping (compare to pedestrians on the pier example in earlier chapter). Negative sides are that sampling schemes are often more complicated and distributional shape of the data should be known. When the variable method is used, it might happen that a lot should be rejected even all analysed items are within the limits. It could be sometimes challenging to agree that. (ISO3951-1, 2013: 10)



General Inspection Level II, AQL 1,0								
Lot size	Code	ISO2859-1 Single samp.	ISO2859-1 Double samp.	ISO2859-1 Multiple samp.	ISO2859-5 Sequential samp.	ISO3951 S	ISO3951 σ	ISO3951 σ + 0.40%σm
91 to 150	F	20	20	20	20	17	8	10
151 to 280	G	32	20 + 20	13 * 5, with 0 def. min 26	32 with 0 defect	23	10	12
281 to 500	H	50	32 + 32	13 * 5 with 0 def. Min 26	32 with 0 defect	24	10	12
501 to 1 200	J	80	50 + 50	20 * 5 with 0 def. Min 40	39 with 0 defect	37	15	18
1 201 to 3 200	K	125	80 + 80	32 * 5 with 0 def. Min 64	54 with 0 defect	54	21	25
3 201 to 10 000	L	200	125 + 125	50 * 5 with 0 def. Min 100	63 with 0 defect	84	32	38
10 001 to 35 000	M	315	200 + 200	80 * 5 with 0 def. Min 80	85 with 0 defect	124	39	46
35 001 to 150 000	N	500	315 + 315	125 * 5 with 0 def. Min 125	113 with 0 defect	186	57	67

Figure 2. Sample sizes with AQL 1,0.

In the Figure 2 the Acceptance Quality Limit (AQL) is 1,0 %. Lot size defines Code letter. ISO standard tables, like the sample size tables, are based on the code. Example here is following. The production lot is 3000 pieces. That would mean code letter K. The next four columns are following the ISO 2851 attribute standard. If single sampling is used, then 125 items should be investigated. No matter how many nonconforming items is found. Double sampling has first 80 items to investigate and if number of the nonconforming items is less or equal than accepted, then 80 pieces is only that must be analysed. If more then another 80 pcs need to be analysed. With multiple sampling, at least two times 32 pieces (64 pcs) must be analysed. Whereas with sequential sampling, 63 items are needed to be analyse if nonconforming item is no found. Then there are the two columns telling ISO 3951 based sample sizes for each lot size with variable data. The first one is for so called S-method when the process standard deviation is not known and the second one is for σ-method where the process standard deviation is known. The last column is the example for a process with 40% measurement system variation and the known process standard deviation. (ISO2859-1, 1999; ISO2859-5, 2003; ISO3951-1, 2013)

2.5 Selection of sampling plan for attribute data test cases

ISO 2859 Sampling procedures for inspection by attributes standards cover the sampling plans for attribute data test cases and are found to be suitable for this thesis. Part 1 covers single, double, and multiple sampling plans. Part 5 covers sequential sampling plans.

For the single sampling the standard gives three number; sample size, acceptance number and rejection number. Lot size defines the sample size. Company quality author can design the inspection level. Normally the general inspection level II is chosen but for inspections that are not so critical also some special inspection level can be chosen. The acceptance number tells how many nonconforming units can be in a sample and if number of nonconforming units is equal or more than rejection number then the whole lot is not accepted. (ISO 2859-1, 1999: 14)

In the double sampling plan, there are two sample sets and the first sample size is smaller than sample size in the single sampling. So, if the quality is clearly better or worse than AQL then only the first sample is needed to investigate. In intermittent case, also the other sample is needed to investigate. The both sample sets have own acceptance numbers and rejection numbers. (ISO 2859-1, 1999: 14)

The multiple sampling is extension of the double sampling plan. There are five sample sets and thus sample size could be smaller. (ISO 2859-1, 1999: 14)

In the sequential sampling plan acceptance and rejection decision is given as soon as the evidence is strong enough, either to accept or reject. There are graphical methods with the acceptability charts and the numerical methods to make decision. The standard gives values of the needed parameters. (ISO 2859-5, 2005: 16)

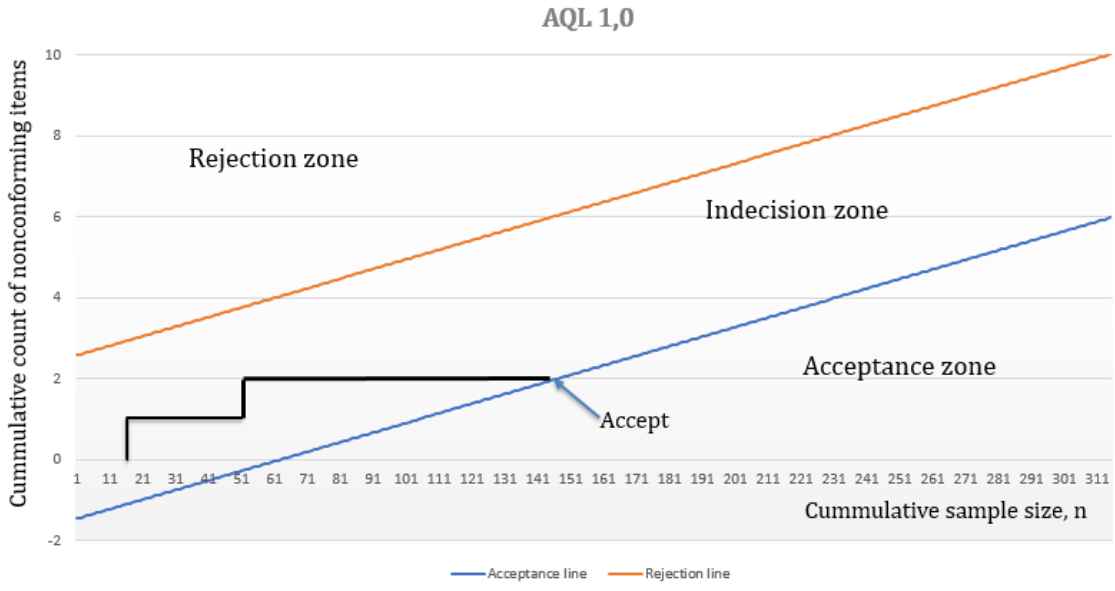


Figure 3. Sequential sampling graphical method. AQL 1,0 and sample size code letter L.

In the Figure 3 case, the first nonconforming item has been found after 17 items have been investigated and the next nonconforming is the 50th, but as the both are still in indecision zone the sampling can continue. As there were no other nonconforming item found the sampling can be ended when 147 items have been investigated. That is the point where the black cumulative number of nonconforming item line crosses the blue acceptance line. In such a case, the single sampling size would have been 200 items, the double would have needed only 125 items and the multiple 150 items. On the other hand, if there would have been no nonconforming item then items to be investigated would have been with sequential 63, with single 200, with double 125 and with multiple 100 items like can be seen in the below example.

Attribute data, AQL: 1,0%, General inspection level II, Lot size 5000 pcs means code L

Single sampling:	n= 200, Ac=5, Re=6
Double sampling:	n1= 125, Cum = 125, Ac=2, Re=5
	n2= 125, Cum = 250, Ac=6, Re=7
Multiple sampling:	n1= 50, Cum = 50, Ac=#, Re=4
	n2= 50, Cum = 100, Ac=1, Re=5
	n3= 50, Cum = 150, Ac=2, Re=6
	n4= 50, Cum = 200, Ac=4, Re=7
	n5= 50, Cum = 250, Ac=6, Re=7

means acceptance is not permitted for this sample size. Ac is acceptance number and Re is rejection number. Cum means cumulative, total number of items e.g. n3: 50 + 50 +50 = 150. In the same n3 case Ac means that if 2 or less nonconforming units have been found, sampling can be stopped and the lot can be accepted. If 6 or more nonconforming units have been found the lot should be rejected. If 3, 4 or 5 nonconforming units have been found sampling should continue at least with the next 50 pieces.

The number of lots that are expected to be accepted at a certain product quality level is about the same for all; single, double, multiple, sequential sampling and variable samplings. In other words, risk level of accepting a lot with nonconforming items is not increasing or decreasing if sample plan is changed between these six. (ISO 2859-1, 1999: 14 & ISO 2859-5, 2005: 21, ISO 3951-1, 2013:8)

2.6 Selection of sampling plan for variable data test cases

ISO 3951 standards could be used when the process is stable and has normally distributed output. Should also have continues scale quality characteristic and lower limit or upper limit or the both. (ISO3951-1, 2013: 1)

2.6.1 Choosing between variable S-methods and variable σ-methods

If the variable data method has been decided to use, then the next question is, do we know the process standard deviation. If the standard deviation is not known, S-method is used but when the standard deviation is considered to be known and stabile then σ-method can be taken in use because the sample size is considerable smaller with σ-method (ISO3951-1, 2013: 11). Estimation of σ should be based on 10 previous lots, unless quality author agrees else (ISO3951-1, 2013: 25). Also with a case of σ-method, the sample standard deviation is calculated to check that the process standard deviation stays stable (ISO3951-1, 2013: 21).

2.6.2 Single or double specification limits with variable data?

According to ISO3951-1 double limits calculations have to used when there are the both, upper and lower limits. Because double limit calculations can first appear daunting, there is also more conservative approach. It is generally the best practice to use one-sized tolerance intervals, if there are both upper and lower specification (so it is bilateral) use the closest specification limit to the sample mean (Durivage 2016).

2.6.3 S-method, acceptability criteria for single specification limit

First is calculated the quality statistic. If mean $\bar{x}$ is outside the specification limits then s, standard deviation, is not even needed to calculate and the lot can be rejected.

$$Q_U = \frac{U - \bar{x}}{s} \quad or \quad Q_L = \frac{\bar{x} - L}{s}$$

U is upper limit and L lower limit. Acceptability constant, k, can be find from ISO 3951-1's table B.1. If the quality statistic is greater or equal to acceptability constant the lot can be accepted.

In a case of upper limit if $Q_U \geq$ k a lot can be accepted and if $Q_U \leq$ k a lot should be rejected. Correspondingly, in a case of lower limit if $Q_L \geq$ k a lot can be accepted and if $Q_L \leq$ k a lot should be rejected (ISO3951-1, 2013: 13).

Example, the same case will be presented later in the Figure 9 (capability analysis):

$Q_L = \frac{\bar{x}-L}{s} = \frac{65{,}98-65{,}80}{0{,}09} = 2{,}12$, then is $Q_L \geq$ k; 2,12 > 1,91 → lot can be accepted.

2.6.4 S-method, acceptability criteria for double specification limits

For double specification limits coefficient f_S is needed from ISO 3951-1's table D.1. After that the maximum allowable MSSD (maximum sample standard deviation) can be calculated. Above example continues.

MSSD $= s_{max} = (U - L)f_s =$ (66,20-65,80) x 0,229 = 0,092

If the process standard deviation s > s_{max} then the lot can be rejected. In the example case s is almost greater than allowed max value. In a case s is smaller then control charts of ISO 3951-1 should be used. Use of control charts is not discussed in this thesis. (ISO3951-1, 2013: 19)

2.6.5 σ-method, acceptability criteria for single specification limit

As the process standard deviation, limit, and acceptability constant, k, are all known the acceptance value $\bar{x}_U = USL - k_\sigma$ *or* $\bar{x}_L = LSL + k_\sigma$ should be calculated before the inspection. Note, k values are from different table (ISO 3951-1's table C.1) than S-method uses.

The lot should be accepted if $\bar{x} \leq \bar{x}_U$ or $\bar{x} \geq \bar{x}_L$ and rejected if $\bar{x} > \bar{x}_U$ or $\bar{x} < \bar{x}_L$

Above example continues. Let assume that the process standard deviation σ is 0,09.

$\bar{x}_L = LSL + k_\sigma =$ 65,80 + 1,900 x 0,09 = 65,97

$\bar{x} \geq \bar{x}_L$: 65,98 > 65,97 → the lot can be accepted. (ISO3951-1, 2013: 21)

2.6.6 σ-method, acceptability criteria for double specification limits

As the process standard deviation, limits, and factor f_σ (ISO 3951-1's table E.1) are all known the maximum acceptance value MPSD (maximum process standard deviation) should be calculated before inspection. Above example continues.

MPSD $= \sigma_{max} = (U - L)f_\sigma =$ (66,20-65,80) x 0,184 = 0,074

Because of the process standard deviation σ (0,09) is bigger than σ_{max} (0,07) the process is unacceptable to produce items between the limits and the sampling is worthless before it is proven that the process variability has been improved to the needed level.

If σ is smaller or equal to σ_{max} then the upper allowance bound $\bar{x}_U$ and the lower allowance bound $\bar{x}_L$ are calculated. For that k value is needed from ISO 3951-1's table C.1.

$\bar{x}_U =$ USL – kσ and $\bar{x}_L =$ LSL + kσ

Then samples are inspected and the sample mean calculated. The acceptance criterion is $\bar{x}_L \leq \bar{x} \leq \bar{x}_U$ and the lot is rejected if $\bar{x} < \bar{x}_L$ or $\bar{x} > \bar{x}_U$

Above example continues with the assumption that the process standard deviation σ is 0,09.

$\bar{x}_U = USL - k_\sigma =$ 66,20 - 1,900 x 0,09 = 66,03

$\bar{x}_L = LSL + k_\sigma =$ 65,80 + 1,900 x 0,09 = 65,97

$\bar{x}_L \leq \bar{x} \leq \bar{x}_U$: 65,97 ≤ 65,98 ≤ 66,03 → according to the lot mean the lot can be accepted although the mean is alarming close to the lower allowable limit. (ISO3951-1, 2013: 22)

2.7 Normal, reduced and tightened inspection

The both ISO standards, 2859 for attribute data and 3951 for variable data, introduce normal, reduced and tightened inspections and switching rules between them. Inspections is started with the normal inspection and if within five lots at least two lots have been rejected the tightened inspection is taken in use. When five tightened inspections have accepted, the normal inspection can be reinstalled. Compare to the normal inspection, the tightened has bigger sample size and/or higher k-value.

Whereas if ten lots on the normal inspection have accepted then the reduced inspection can be taken in use and should came back to the normal inspection if a lot is not accepted.

Compare to the normal inspection, the reduced has smaller sample size and/or lower k-value. (ISO2859-1, 1999: 16, ISO3951-1, 2013: 25)

2.8 Bulk material

ISO 10725 standard covers acceptance sampling plans and procedures for the inspection of bulk materials such as powder or liquid. For the homogeneous material sample size estimation, the standard gives couple of methods. The simplest model is

$$n = \left(\frac{(k_\alpha + k_\beta)\sigma}{USL - LSL} \right)^2$$

Let us assume that we have a process with the upper limit 175 and the lower limit 30. The process standard deviation is 17. Producer's and customer's risks both are 5%, meaning α and β are 0,05. From the normal distribution table (e.g. from Apostol, 1969: 536) it can be seen that k value for 0,05 is 1,645. The distribution table value is also called z value. (ISO 10725, 2000: 54)

Thus

$$n = \left(\frac{(1{,}645 + 1{,}645) \times 17}{175 - 30} \right)^2 = 0{,}15$$

→ only 1 item to be investigate (rounded up to the nearest integer).

Limits for the mean are $\bar{x}_L = LSL + k_\sigma$ and $\bar{x}_U = USL - k_\sigma$

In the example $\bar{x}_L = 30 + 1{,}645\ x\ 17 = 58$ and $\bar{x}_U = 175 - 1{,}645\ x\ 17 = 147$

More complex method is taking account things like cost of a measurement, cost of preparing a test sample and cost of drawing a sampling increment. As a summary, it can be presented that so far than the relative standard deviation σ/(USL-LSL) ≤ 0,25 sample size is 2. The summary is based on the fact that the all sample size tables 3…22 in the standard have sample size 2 when the relative standard deviation ≤ 0,25, no matter what are the costs. Complex method cases where the relative standard deviation > 0,25 is not covered in this thesis. (ISO 10725, 2000: 19, 22…30)

For medical industry in United States of America, Court for the district of New Jersey has declared in a case United States of America versus Barr Laboratories Inc. that no regulation, guideline, or publication requires any specific blend sample size but the proper sample for proportion homogeneity testing in both validation and normal production lots, is three times the active ingredient dose size. (United States District Court, 1993: 12)

3 LITERATURE REVIEW OF MEASUREMENT SYSTEM ANALYSIS AND PROCESS CONTROLS

3.1 Measurement system analysis

The measurement system analysis helps to understand how big measurement system variation is compared to the process variation and to process specific limits.

$$\sigma_{Total} = \sqrt{\sigma^2 + \sigma_m^2}$$

Equation 3. Total standard deviation (ISO3951-1, 2013: 87)

In Equation 3, the total deviation σ_{Total} is caused by the process variation σ^2 and the measurement system variation σ_m^2. If measurement system variation is less than 10% from process variation, then

$\sigma_{Total} = \sqrt{\sigma^2 + (0{,}1\sigma)^2} = \sigma\sqrt{1 + 0{,}01} = 1{,}005\sigma$

Thus, if the measurement system is causing max 10% deviation then the total standard deviation might increase only 0,5%, which can be ignored (ISO3951-1, 2013: 87).

If there is a measurement error, is that owing to repeatability and/or reproducibility? Repeatability means that the same appraiser or measurer is getting the same result when he or she repeats the measurement with the same unit and with the same measuring equipment at the same place under the same conditions. Repetition should be done over a short period. Reproducibility means that other persons get the same result when they repeat the measurement with the same unit and with the same measuring equipment at the same place under the same conditions. (ISO 3534-2, 2006: 3.3.5 & 3.3.10)

Other relevant terms are accuracy, precision, and discrimination. Accuracy is the different between measured average value and the correct value. Precision is total variation of measurements. How close measurements are to each other. Repeatability and reproducibility compose precision. (Minitab 2017)

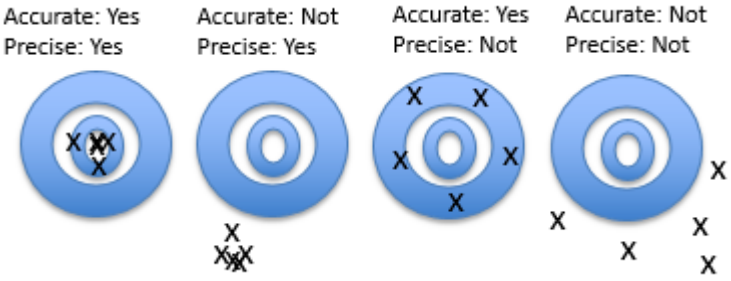


Figure 4. The meaning of accurate and Precise

In Figure 4, the most left shooting target has accurate and precise hits as all are in small area and in the middle of the target. The second target left has bad accuracy as the hits are far away from bull's eye, but good precision as the hits are next to each other. The third left has widely spread hits so precision is bad but average of the hits is middle so accuracy is good. The most right-hand side has nothing good as the hits are widely spread and average is far away from the middle.

Discrimination tells how small detail the measurement system can measure.

Figure 5. Example of discrimination. Two tape measures

Continue with the Figure 5. Let us assume that we have two nails. One 2,8-centimetre-long and another 2,2-centimetre-long. The question is which one is longer. With the left side tape measure, it can be only said that the both are between 2 and 3 centimetres but the right-hand side tape measure can distinguish the nails according to their length.

P/T is precision to tolerance ratio. P/T shows what part measurement error is taking from tolerance. Accepted tolerance is upper specification limit minus lower specification limit. %R&R variation indicates measurement part from total variation.

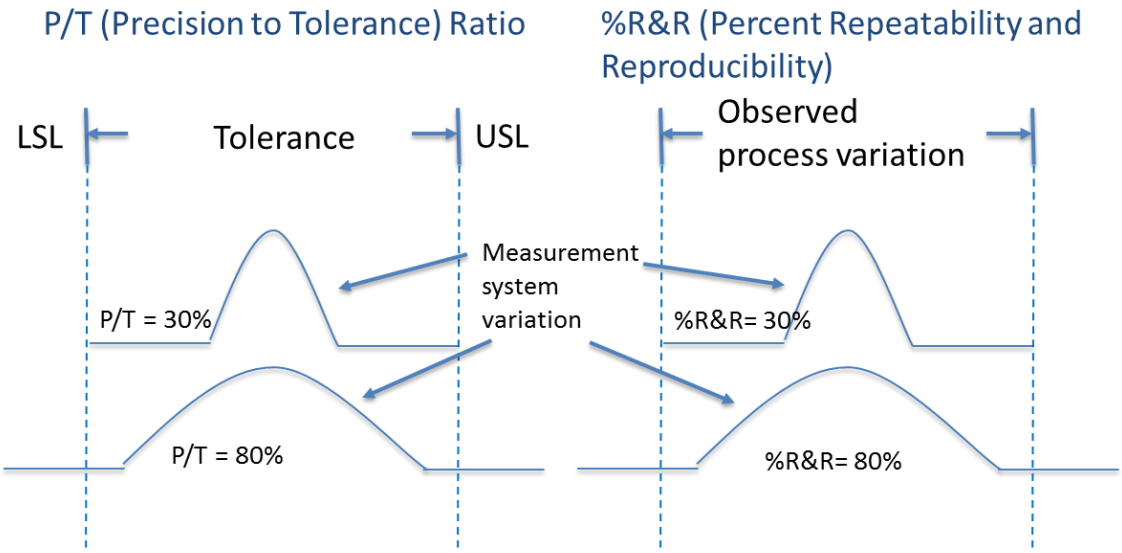


Figure 6. P/T Ratio and %R&R

In the Figure 6 left-side there are two samples. Above measurement system variation covers 30% from tolerance and in left below case it covers 80%. Some items that would be accepted in 30% measurement system variation process is rejected in 80% measurement system variation process. On the right-hand side above measurement system is causing 30 % of total variation and 80 % in right-hand side below case.

3.2 Statistical process control and process stability

In the statistical process control the data is recorded over time and statistical tests are performed to see if a process is stable and therefore predictable. A stable process has unvarying mean and distribution, and has only random variation without any special causes. (AIAG, 2005: 14)

Control charts plots the data over time. The control limits (e.g. $\pm$ 3σ) add visualisation to see if a process has variation due to special causes. Also, occurrences that are not normally part of a process are easier to notice. The special causes could be for instance outliers (data points outside control limits), trend (continued fall or rise), pattern (equal shape), shift in mean or shift in variation. (Minitab 2017)

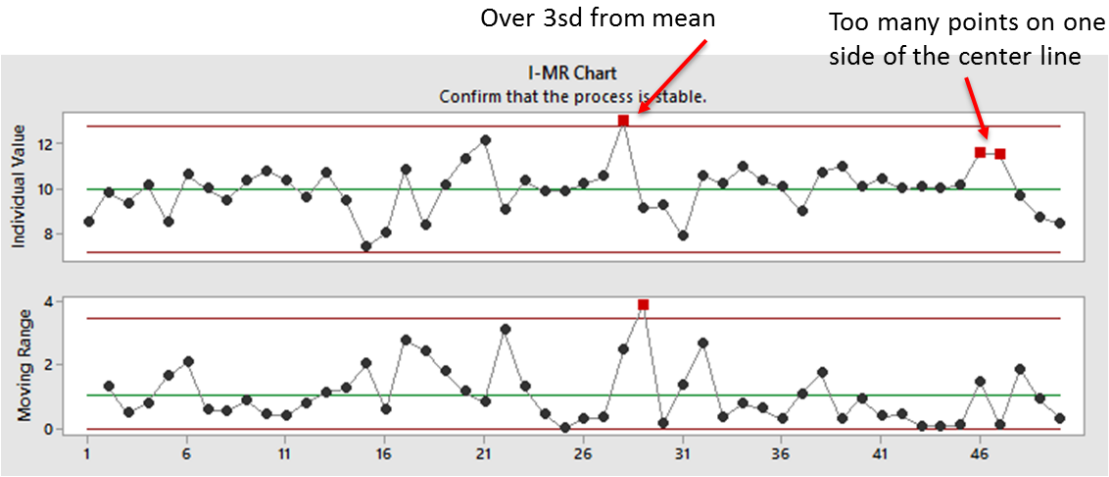


Figure 7. Individual Moving Range Chart.

The red control lines are 3 standard deviations from the mean. Individual data points show stability of process. Even if the process is stabile it could be still incapable. In the Figure 7, there is one measurement with too high value and shift in mean meaning too many measurements in a row over the mean. The Moving Range shows stability of the process variation.

3.3 Process normality

Many statistical tests rely on population normality; thus, a normality test can be an important step in the analysis. The null hypothesis for a normality test states that the population is normal. Population normality can be assessed with a normal probability plot, which plots the ordered data values near against expected values if the sample's population is normally distributed. If the population is normal, the points form an approximately straight line. An informal normality test, called "the pencil test", can be used as a quick visual assessment. Visualise “a pencil" lying on top of the plotted line: if it hides all the data points on the plotted line, the data is likely normal; if data points can be seen beyond the edges of the pencil, the data is likely nonnormal. (Minitab 2017)

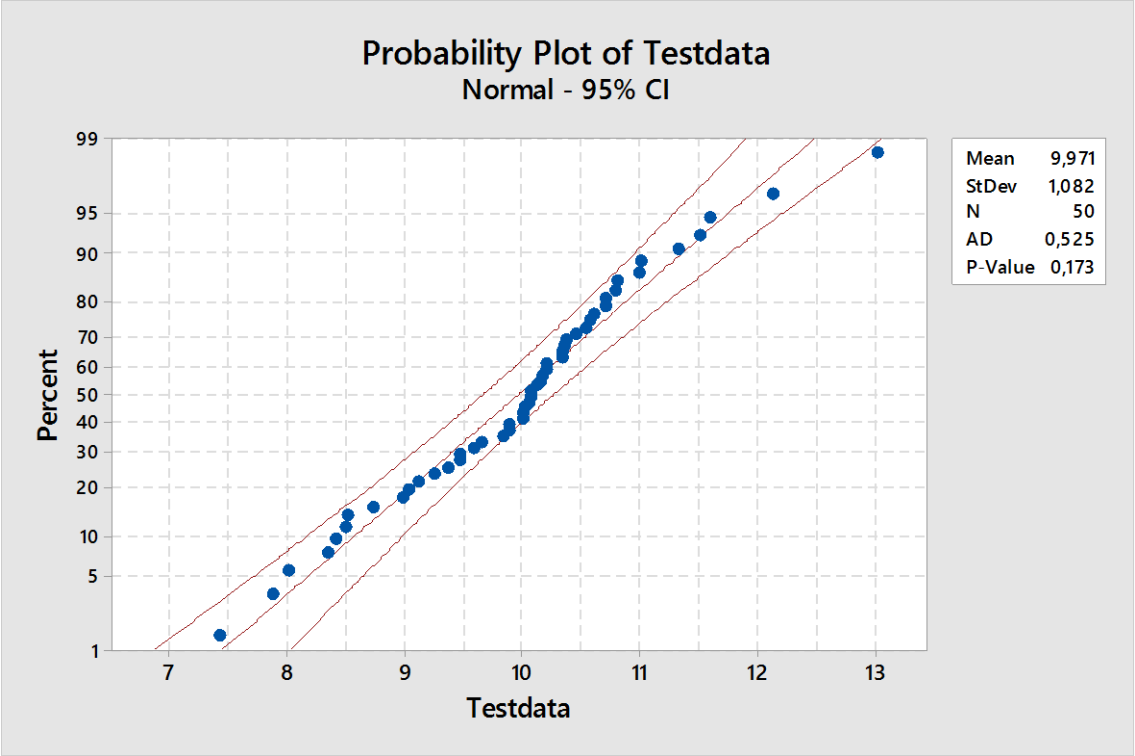


Figure 8. Normality test.

In the Figure 8, P-Value is 0,173, which is more than 0,05 → the data is normally distributed. The outermost lines are 95% confidence interval lines. AD means Anderson-Darling test that tells how good the data adopts a specific distribution. AD is not covered in this thesis. (Minitab 2017)

3.4 Capability Analysis

Process Capability means that the customer requirements can be met. The capability analysis shows how well the process meets the requirements and could provide information about how to improve the process. Before assessing the process capability, the process should be stable. Only a stable process can be used when predicting the future performance. (Minitab 2017)

The capability of a process should be constantly recorded and analysed. The analysis can tell e.g. is the process good enough to meet the customer requirements, how well the process should perform in the future, is there any improvement needed and have the possible improvement actions improved the process.

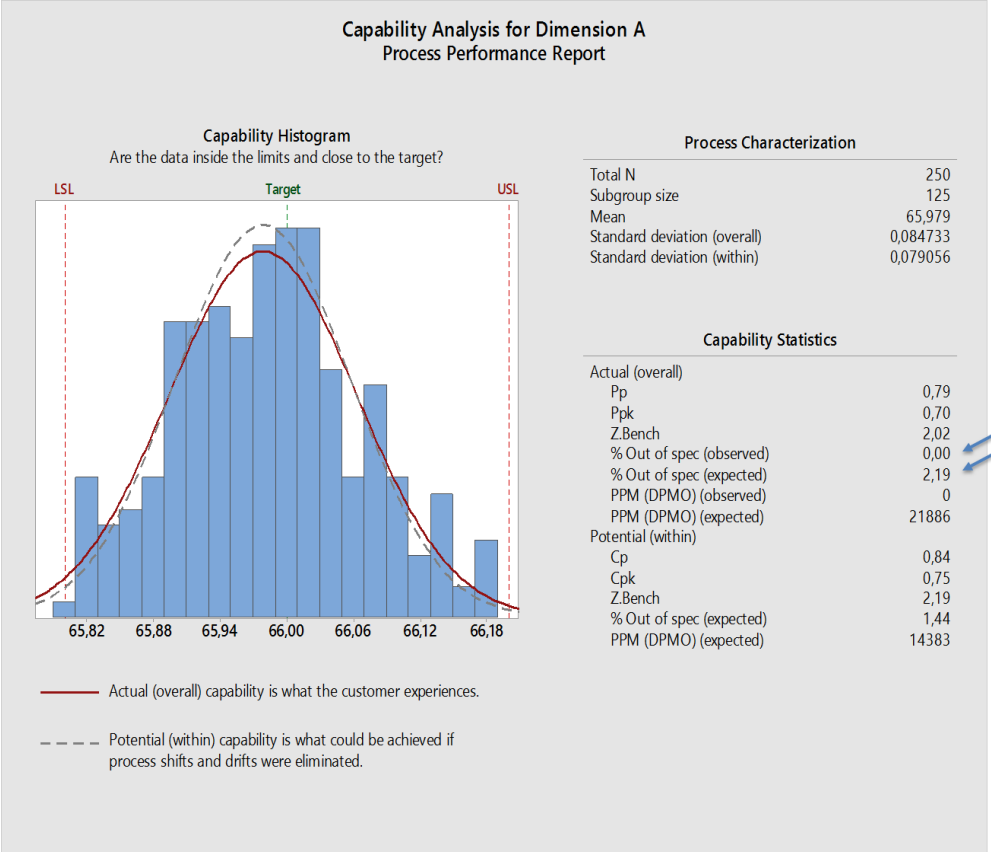


Figure 9. Example of capability analysis.

In the Figure 9, there are two 125 pieces samples combined to one 250 pieces sample. The lower and the upper specific limits are known. Minitab informs the capability that is observed and estimates capability for the whole population.

The difference between Pp (process performance index) and Cp (process capability index) is that Cp is calculated first for each subgroup then combined so it accounts only the variation within the subgroups whereas Pp is calculated directly from the all samples i.e. from total n. If the process is stable with minor variation Cp and Pp should be about the same. c_{pk} (and P_{pk}) describes how capable the process is. Minimum value for c_{pk} depends on application. Some processes have more critical quality targets and thus are needing higher c_{pk} than some not so critical process. Pp is normally used when only sample standard deviation is known but not process standard deviation. Cp is used when also process standard deviation is known. (Minitab)

Table 1, c_{pk} and the corresponding process yield.

C_{pk}	Sigma level (σ)	Process yield	Process fallout / PPM
0,33	1	68.27%	317311
0,67	2	95.45%	45500
1	3	99.73%	2700
1,33	4	99.99%	63
1,67	5	99.9999%	1
2	6	99.9999998%	0,002

In the Table 1, Sigma level is the same than standard deviation. Note, in Six Sigma terminology (statistical process control and total quality management ideology, originally presented by Motorola in the 1980s) 1,5σ drift in the process is expected to happen eventually and thus 6σ will be only 4,5σ with 3,4 DPMO (defects per million opportunities). It should be also noted that in the case of shift or drift the population has moved closer to the one specification limit and further away from the other specification limit. Thus, the shifted 1,5-sigma case is not equivalent to the failure rate of a 4,5-sigma process with the mean centered on zero. (Tennant, 2001: ix, 25)

3.5 Hypothesis test

Hypothesis tests are tests to help decision making. Test cases can be for instance is the data normally distributed or do data sets statistically differ from each other. There are five essential terms. Null hypothesis H_0, which is the tested theory and the alternative hypothesis H_a, which would be true if the null hypothesis H_0 is not true. The p-value indicates whether an observed relationship is statistically significant. Alpha risk (α) and Beta risk (β) are described in the Figure 10.

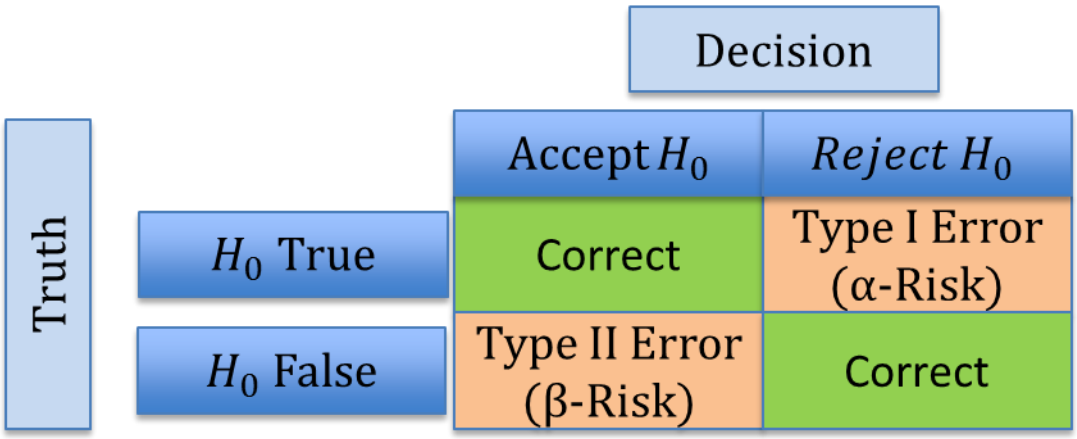


Figure 10. Decisions and errors

An example null hypothesis H_0 could be that the product is conforming, in other words, it works as specified. Type I error means that the products are scrapped even if they could have been sent customer. Alpha risk (α) describes likelihood of making a type I error. Type II error means that the products have been sent customer even when they should have been scrapped or repaired. Beta risk (β) describes likelihood of making a type II error. (Wickens & Hollands, 2000: 18)

P-value helps to decide should the null hypothesis H_0 be accepted or not. It shows the probability the data could occur by random chance if the null hypothesis H_0 is true. P-value is linked with Alpha risk (α). Often value 0,05 has been used as an Alpha risk. That would mean 5% risk to make wrong decision. Thus, if P-value is less than 0,05 there is strong evidence against the null hypothesis H_0 and the alternative hypothesis H_a can be accepted. If P-value is 0,05 or more, there is only weak evidence against the null hypothesis H_0 and the null hypothesis H_0 cannot be rejected. (Minitab 2017)

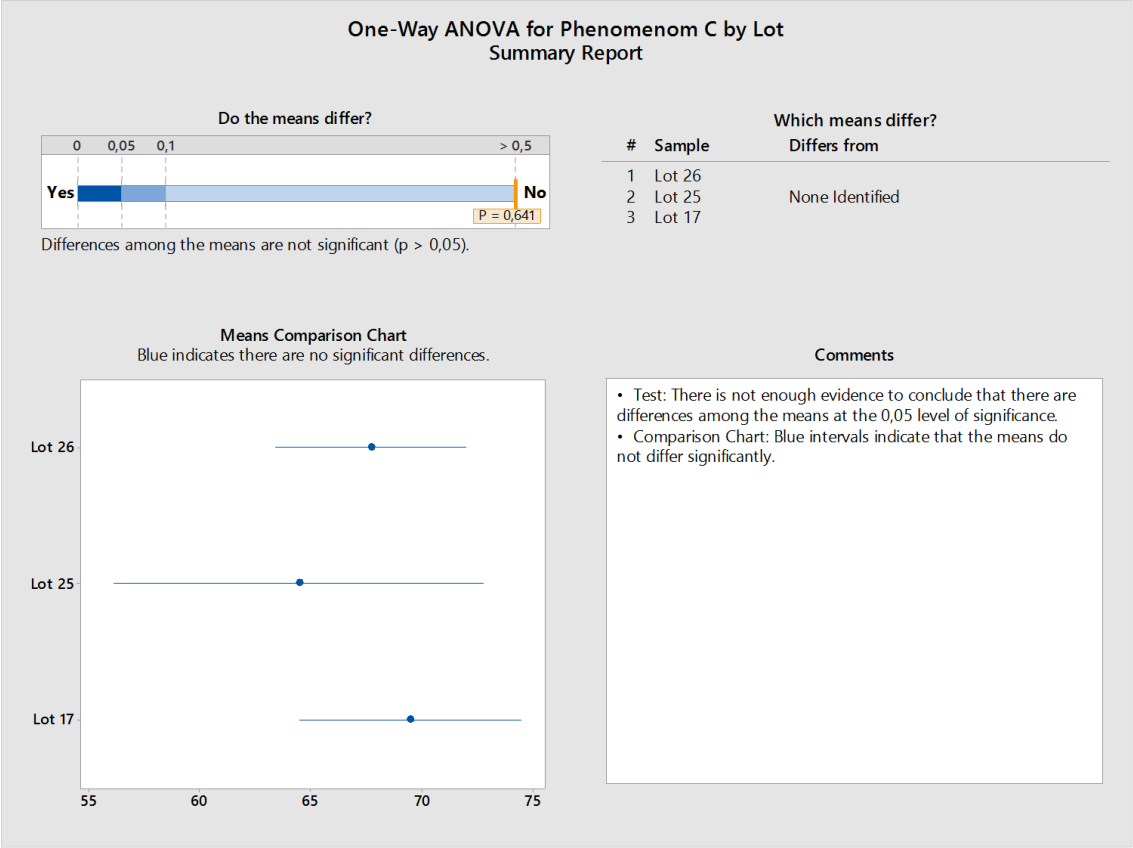


Figure 11. Analysis of variance. Three samples, each has 10 items.

In the Figure 11 the null hypothesis is, are differences among the means of each lot different. α-level is 0,05 meaning that 5% risk is accepted of concluding there are differences when there are none. With relatively high P-value (0,64) it can be said that there is not enough evidence to conclude that there are significant differences between the means. (Minitab 2017)

Power of the test is essential in the hypothesis tests. It tells the probability that difference can be found when it is there. The power also tells the likelihood that the null hypothesis is correctly rejected when it is false. In other words, it is one minus β. Several factors affect power:

- Sample size: More samples tells more about the population and thus increases the power.

- α-level: Large α-value grows the power as the null hypothesis is more likely rejected with higher α-value. Have to note also that higher α-value also increases the probability of the Type I error.
- σ (variability in the population): With small σ it is easier to find the difference so power increases when σ decreases.
- The population: The more different the populations are, the easier it is to detect a difference. (ISO 3534-1, 2006: 11)

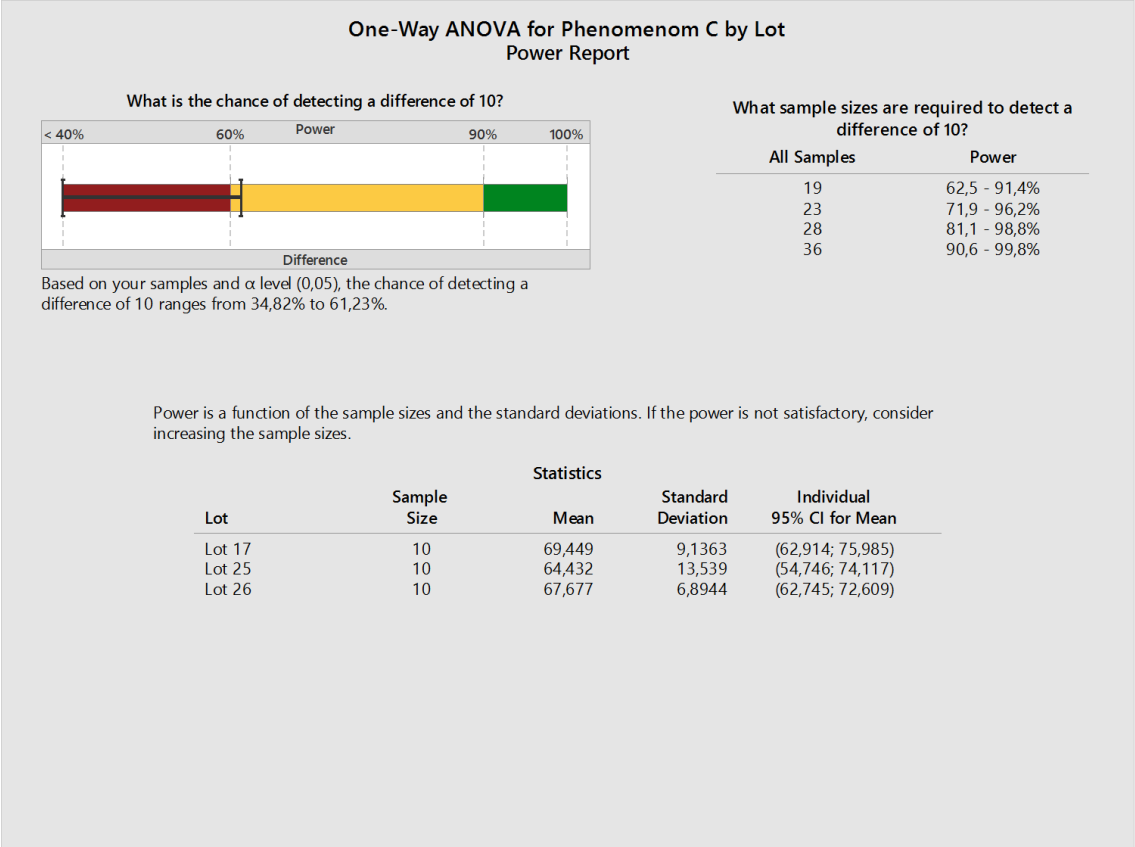


Figure 12. Power report for the same samples than in Figure 11.

In the Figure 12 case it is chosen that 10 unit (e.g. 10 centimetre) of the measurement difference between the means has practical value. With α-value 0,05, sample size 10 and standard deviations between 6,9…13,5 the power is 35…61 and it is not good. The easiest way to reach 90% power is to increase sample size to 36 pieces. (Minitab 2017)

3.6 Regression test

With regression test it can be seen is there correlation between predictor(s) X(s) and response variable Y. The regression model tells the direction (sign of coefficient), size, and significance of the relationship between response and predictor(s). If there is correlation, then the model can be used to predict how the process work with different predictor values. The null hypothesis is that the coefficient has no effect to the response. Thus, low p-value tells that the predictor is important to the model. R-sq (R^2) is defining the variation in the output explained by the regression model. R-sq values are always between 0 and 100%. (Minitab 2017)

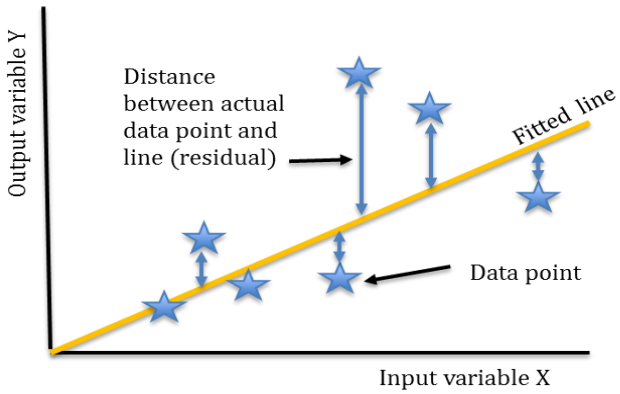


Figure 13. Regression model.

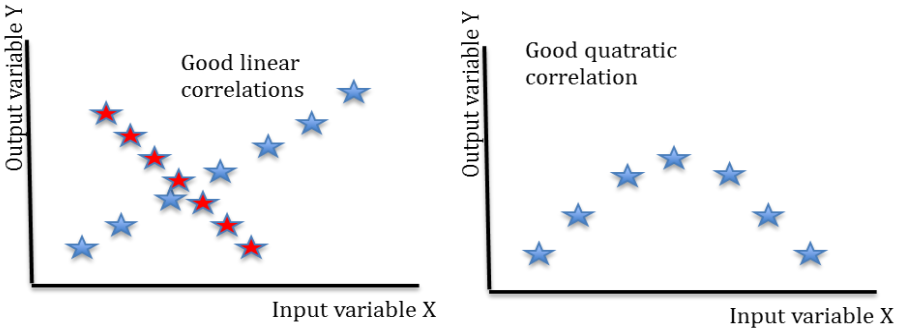


Figure 14. Correlation coefficient.

In the Figure 13 the closer data points are to the fitted line, the bigger R-sq is. In the Figure 14 left side the blue stars have r = 1 and the red stars have r = -1. The right-hand side stars make the parabola curve and therefore it has a quadratic function with a polynomial term x^2.

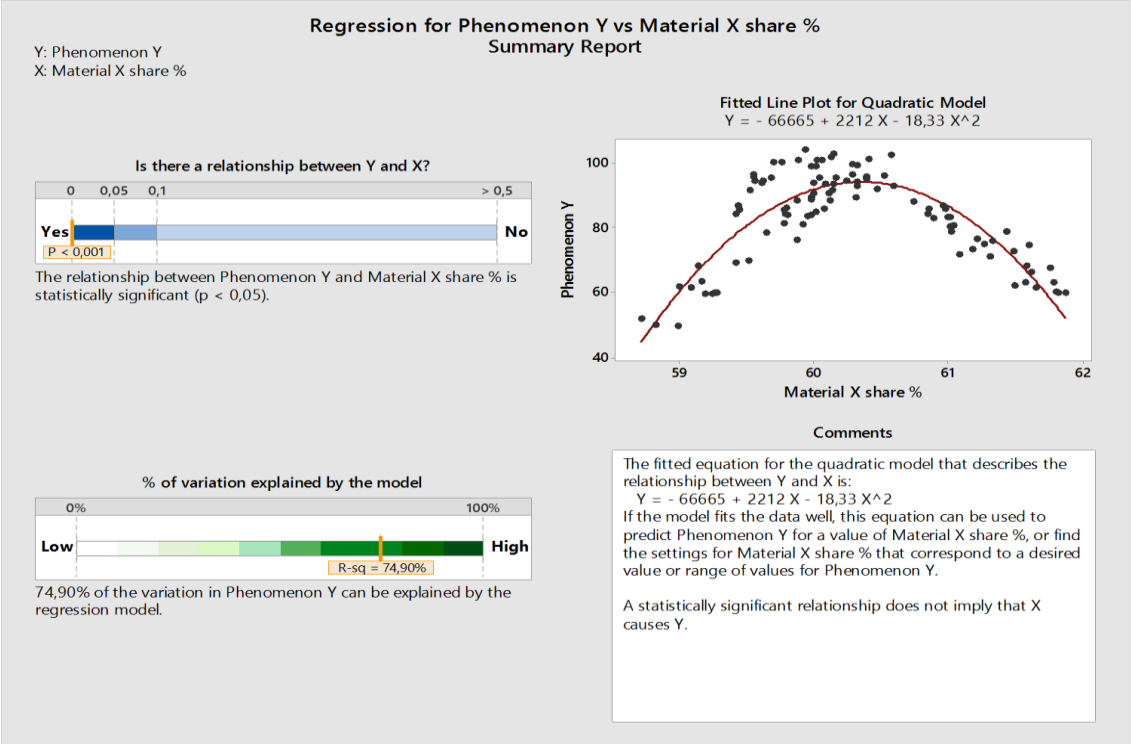


Figure 15. Regression test.

In the Figure 15 there is clear correlation between share of the material X and the phenomenon Y. The left upper picture shows P-value, which is less than 0,001 and clearly less than 0,05 so the relationship between X and Y is statistically significant. The left lower shows that R^2 value is 74,9% and thus 74,9% of the variation of Y can be interpreted by the regression model. The right-hand side shows the plot diagram and the equation. As the red line is curve the model is nonlinear but the quadratic. The sample size is enough as typically at least 40 items is needed to get a good estimate about the relationship. Note disclaimer "A statistically significant relationship does not imply that X causes Y". (Minitab 2017)

4 DATA COLLECTION PLAN AND MEASUREMENT SYSTEM ANALYSIS FOR THE COMPANY

4.1 Data collection plan for the company

Data collection plan is a structured way to gather rational data in order to perform different statistical analysis and to see all measurements in the one sheet. Based on need column titles, in other words, recorded details and thus the content of the plan are varying. There need to be a clear goal what the plan should include and what not. (Eckes, 2000: 72-74)

The data collection plan is made by listing all measurements which are done in production and quality control. The content is based on work procedures, work instructions, process descriptions and interviews. The goal is to collect all measurements which are done in the company.

The content of the data collection plan is:

1. What to measure?
    - Name of parameter: What is actually measured?
    - Type of measure: Is that attribute or variable data, process input X or output Y?
    - Operational definition: Clear explanation on the method of measurement so that the measurement is repeatable if another person does that again.

2. How to measure?
    - What method is used? Is e.g. some equipment used or is it visual inspection?
    - What data tags are defined? A tag could be such as time, lot, date, tester and line.
    - What data collection methods are used? A method could be manual, computer, automated tester etc.

3. Who oversees the measurement?
    - Who is measuring?
    - Who knows the best how to measure?

4. Sample plan

- What physical property is collected? A property could be such as weight, length, colour or torque.
- Where the measurement is done?
- When the measurement is done? How often the measurement is done?
- How many measurements are done? The number of data points or the number of items per lot. (Eckes, 2000: 72-74)

Tables 1 and 2 show example of the data collection plan.

Table 2. Data collection plan, first part

Data Collection Plan					
Define What to Measure			Define How to Measure		
Measure	Type of Measure	Operational Definition	Measurement or Test Method	Data Tags Needed to Stratify the Data	Data Collection Method
Name of parameter or condition to be measured	X or Y input or output attribute or variable data, product or process data	Clear definition of the measurement defined in such a way as to achieve repeatable results from multiple observers	Visual inspection or automated test? Test instruments are defined. Procedures for data collection are defined.	Data tags are defined for the measure. Such as: time, date, location, tester, line, customer, buyer, operator, etc.	Manual? Spreadsheet? Computer based? etc.
Incoming inspection					
Dimensions of applic. Parts	X variable	According WP22, Lower and upper limits	Caliper	Lot, date, tester	Manual

Table 3. Data collection plan, second part

Who will Do it?	Sample Plan			
Person(s) Assigned	What?	Where?	When?	How Many?
State who has the responsibility?	What measure is being collected	Location for data collection	How often the data is collected	The number of data points collected per sample or the number of samples per lot
N.N.	Lenght/mm	Incoming inspection	Every lot	ISO2859 AQL 0,4%, Level II normal

The last row in the tables 2 and 3 is an example row showing what kind of data could be recorded. The example case is from the incoming inspection. The first column shows that

measured parameter is dimensions of applicator parts. The second column tells that type of measurement is variable and that it is input (X) of the process. Y would have mean that it is output of the process, something that customer gets. The third column informs that measurement is done per the work procedure 22 (separate instruction) and there are lower and upper specification limits for the dimensions. The fourth column points that the measurement instrument is caliber. The fifth column reports that identification information, which are recorded are lot name, date and name of the operator. The sixth column states that the measurement is measured and recorded manually. In the second table the first column annonciates that person N.N. is responsible. The second column discovers that the measurement is length and unit is millimetre. The third column informs that measurement is done on the incoming inspection area. Whereas the fourth column reveals that every incoming lot is tested and finally the fifth column brings out that standard ISO2859, AQL 0,4 % and normal inspection level II is used.

4.2 Selection of tests and measures to focus

Measurements or tests that are already fixed so that there is no room for improvement are omitted from the data collection plan. For instance, if weight of a lot is already measured only once, it is quite difficult to reduce the number of measurements anymore if the weight measurement must be done for every lot.

What is left are the tests to study more intensively. Common factor for all selected tests is that there already is a sample plan in use, which is based on ISO standard 2859-1 Sampling procedures for inspection by attributes.

4.3 Combining of existing measurement data

All existing measurement data from quality and validation tests are collected to one MS Excel file. Thus, the process stability, the normality test and the capability test for each measurement are easier to perform as all data from each specific measurement are in the same file.

4.4 Performed measurement system analysis

Measurement system analysis is done covering gage run chart and gage R&R (Repeat & Reproducibility) Minitab tests. For that three persons are selected to measure twice three

components. A total of 15 units of each component are measured twice by each person. Total amount of measurements is $3 \times 3 \times 15 \times 2 = 270$ observations. Another option is to use less units but measure every unit three times. Number of required parts, appraisers (persons) and repeated readings depend upon the importance of the characteristic and upon the wanted confidence level (MSA 2010: 84). The persons are those who normally operate the measurement. The components are selected based on stability test, where the three components resulted the worst. The components are numbered to ensure that the same components are compared. Minitab recommends that measurements are done in random order (Minitab 2017). All operators use the same Vernier caliper in order to avoid variation due to measuring instrument. The measurements are entered in Minitab to get repeatability, reproducibility, standard deviation for the both, %R&R variation (measurement part from total variation) and P/T (precision to tolerance) ratio.

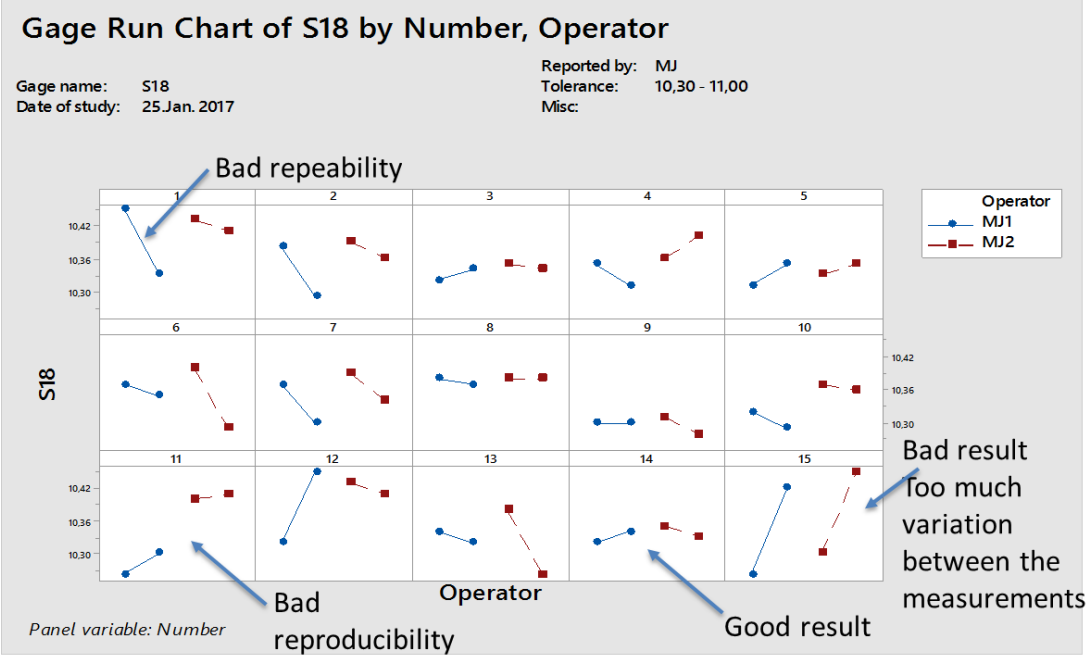


Figure 16. Gage Run Chart.

The Figure 16 shows the result of Minitab's Gage Run Test. There are two operators measuring the length of component S18. Totally 15 units are measured twice. The results are between 10,26…10,45 mm. The lines in each box should be horizontal. Whereas the lines between the boxes should be at different level to indicate the part-to-part variation. If the repeatability is good, then each measurement of the operator (same colour) should be at

the same level or very close together in the box. If the reproducibility is good, then each operator (different colour) should be at the same level or very close together in the box.

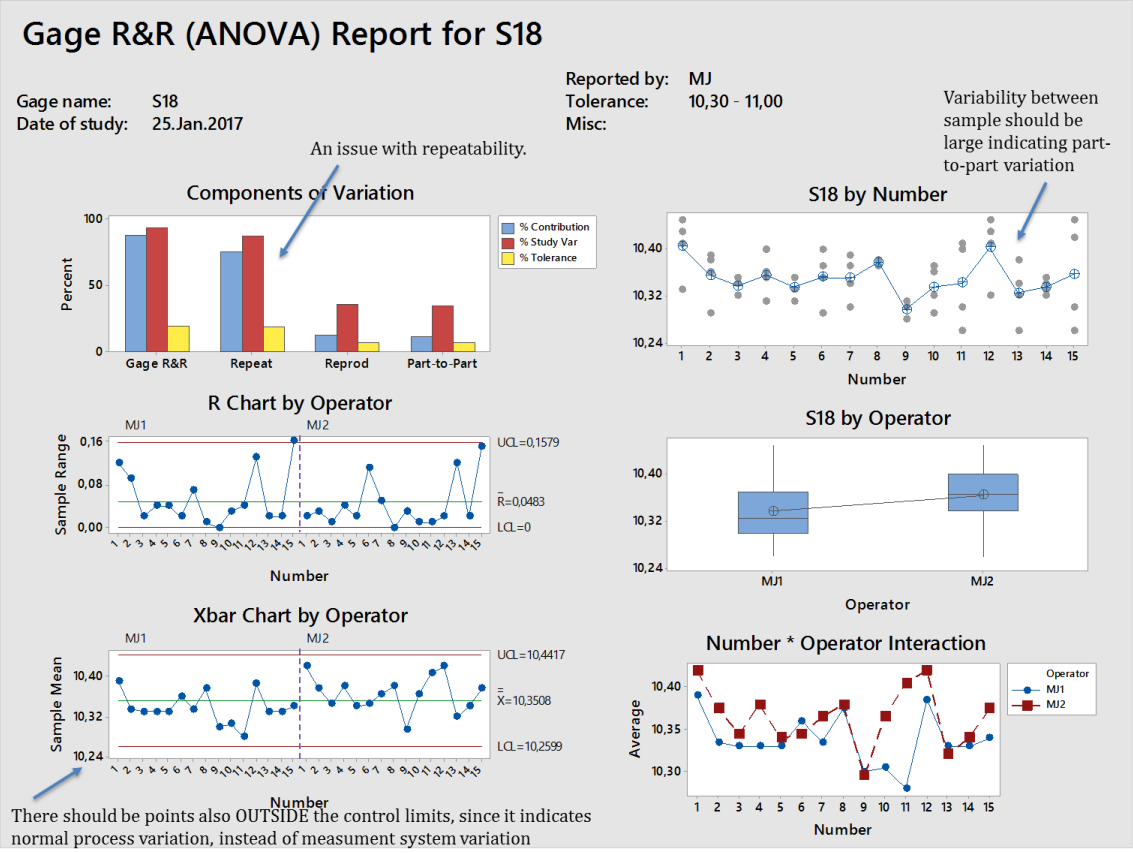


Figure 17. Gage R&R study.

The Figure 17 shows the result of Minitab's Gage Repeat & Reproducibility study test run with two operators and 15 units. The same measurements are used in the Figure 16. The upper left corner, component of variation shows the source of variation. Repeat and Reprod combine Gage R&R. The Gage R&R bars should be as small as possible whereas Part-to-Part should be 100% or close to that. That would indicate small measurement system variation compare to the process variation. In the Figure 17 the repeatability is causing the major part of the variation. On the middle left is presented sample range by operator. That shows the difference of each measurement for each item. The first operator MJ1 and then from middle the same for the operator MJ2. On the bottom left is presented the mean for each item by

the operator. This is the same case as with the range that first comes MJ1's measurements and then MJ2's measurements.

On the right-hand side top is all four measurements per items. Below that are averages of the both operator in the boxplot chart. Box includes 50% of measurements, the mean is marked with cross-circle and the median with the horizontal line inside the box. The bigger the box is the wider the data has spread. The graph bottom right shows average of each item by the both operator. (Minitab 2017)

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0,0022269	88,05
Repeatability	0,0019078	75,43
Reproducibility	0,0003191	12,62
Operator	0,0003191	12,62
Part-To-Part	0,0003022	11,95
Total Variation	0,0025291	100,00

Process tolerance = 0,7

$\%R\&R = \frac{\delta ms}{\delta total} x100$ = What percent of the Total Variation is taken up by measurement error. Notice the %Study Var's (excluding Total Variation %Study Var) do not sum to 100.

Source	StdDev (SD)	Study Var (3 × SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0,0471904	0,141571	93,84	20,22
Repeatability	0,0436784	0,131035	86,85	18,72
Reproducibility	0,0178642	0,053593	35,52	7,66
Operator	0,0178642	0,053593	35,52	7,66
Part-To-Part	0,0173844	0,052153	34,57	7,45
Total Variation	0,0502906	0,150872	100,00	21,55

%P/T = = What percent of the tolerance is taken up by measurement error. Tolerance = USL-LSL

Number of Distinct Categories = 1

Figure 18. Gage R&R study report.

The Figure 18 tells the results of the same test run as in previous figure in numeric form. VarComp column shows the variance of each source. Those can be summed

Repeatability + reproducibility = Total Gage R&R

Part-To-Part + Total Gage R&R = Total Variation

But StdDev columns cannot be summed to get the total variation as standard deviation is square root of variance and thus it is nonlinear.

$a + b \neq \sqrt{a^2 + b^2}$ for instance, 3+4=7 but $\sqrt{3^2 + 4^2} = 5$

The results show that the measurement system is causing 93,8% of the total variation and 20% from the tolerance. In other words, in this case the measurement system is bad and

should be improved. Improvement action could be e.g. training of the operators or better instructions for the measurements. Total Gage R&R should be less than 30% and less than 10% would be ideal (Minitab 2017).

Number of Distinct Categories shows how many distinct categories of parts the measurement system can distinguish. The higher Total Gage R&R, the lower this number will be. In this case number is 1 and shows that process is incapable of distinguishing parts properly. The number should be at least 5 (Minitab 2017).

5 OTHER MINITAB TESTS THAT ARE DONE FOR THE COMPANY

5.1 Statistical process control and process stability

Control charts are done for each component to see is the process in control and stabile. Each production lots are considered. In a case that a component has had more than one supplier only the components produced by the last supplier are taken into tests.

5.2 Process normality

To see that the processes have normally populated data processes normality are tested. Minitab shows normality with the same test than stability or also separate normality test can be used.

5.3 Capability Analysis

Each component is tested to see is the process capable to produce components within the tolerance limits. Every component has the both, lower and upper specification limits.

5.4 Hypothesis test and power

Hypothesis tests are done to see are production lots statistically similar. Mean and standard deviation are tested.

5.5 Regression test

The regression test is used to see correlation between the predictor(s) X(s) and the response variable Y. Specially interesting cases are those where the response is not as expected.

6 RESULTS

It was found out that the smallest sample size could be reached when the test results can be read from a variable scale. This is especially true when the standard deviation of the process is known. Hence, it is essential that the process performance is constantly recorded and kept in a form that possible trends can be seen. If variable data cannot be used and attribute data methods are used, then sequential sampling gives the smallest sample size on average. The sample sizes are always depending on the accepted quality limit, the lot size, how big sampling risk can be tolerated and how well the process in known. Likewise, the acceptance and the rejection criteria could differ. Therefore, individual test case sample sizes are not presented in this thesis. Instead, the ISO standards provide tables for each combination.

The risk level of accepting a lot with nonconforming items is not increasing or decreasing if the sample plan is changed between single, double, multi, sequential or variable sampling plans.

It was found out that bulk material sample size calculation can be complicate when homogeneity is tested but also a simple version is found.

One finding was that the company would benefit from using a better statistical software tool. Video instructions on how to use basic features of Minitab 17 is made. This is done with open-source screen recording program Webinaria. The reasons for Webinaria selection are that it is free of charge, it works as specified and it has voice narration capture feature. Each main feature of Minitab is recorded as own video.

7 CONCLUSION

The main goal of this thesis was to find the suitable sampling systems for BonAlive Biomaterials Ltd. During the journey, it came clear that sample systems evaluations are needed for tests providing attribute data, for variable data and as well for a bulk material.

The thesis work was carried out in accordance with the original schedule and the research questions were answered. The thesis writer leant that statistic tools are not always needed when sample size is determined but also ISO standards could be used. The quality persons of the company were able to learn about Minitab and got more insight to ISO standards.

Major finding was the company could save time and money by changing sampling system from attribute systems to variable systems when the inspection of the sample provides variable data. It was also found out that the number of currently inspected items of bulk material meets the ISO standard requirements and can be even reduced.

More validity to test cases, introduced in Minitab test case sections, can be accomplished by using the data from the future inspections. More inspected lots will allow also to move from S-method variable inspection to σ-method variable inspection with considerable smaller sample size.

Data collection plan turned out to be a good method to understand what kind of measurements and tests are done in the company and see all measurements in the one sheet. With the plan, it was easy to discuss with the staff and check are all measurements and details of them covered.

Retrospectively could be said that in measurement system analysis also couple of components among the best result group should have been selected. Thus, the result of the measurement system analysis would have represented the entire component group better. The measurements used are variable data measurements. The measurement system analysis could have been done also for measurement systems of attribute data. Minitab has own attribute agreement analysis feature for that.

This thesis work could be a DMAIC project where D stands for project definition, the phase where the issue or the opportunity is clarified. M means process measurement. A means process analysis. In M and A phases the most important process inputs are identified. M and A phases clarify how the process is working and what are the key contributors of the

process. Whereas I means process improvement, the phase where the most important process inputs are optimized. C means process control, where controls of the process are implemented to sustain the gains. Due to time constrains I and C phases were not covered in this thesis, but surely the company continues with the processes improvements and controls.

REFERENCES

AIAG 2005, Statistical process control, Reference Manual, DaimlerChrysler Corporation, Ford Motor Company, GM Corporation, 2. Edition, ISBN 978-1605341088

Apostol, Tom 1969, Calculus volume II, Wiley International Edition, 2. edition, ISBN 0-471-00008-6

BS 6001-0, Sampling procedures for inspection by attributes – Part 0: Introduction to the BS 6001 attribute sampling system, 1996, BSI – British Standards Institution

Durivage, Mark 2016, How to establish samples sizes for process validation using statistical tolerance intervals. Pharmaceutical online, accessed 22. Feb. 2017 https://www.pharmaceuticalonline.com/doc/how-to-establish-sample-sizes-for-process-validation-using-statistical-tolerance-intervals-0001

Eckes, George 2000, The Six Sigma Revolution: How General Electric and Others Turned Process Into Profits, John Wiley & Sons, ISBN: 978-0-471-38822-7

FDA, U.S. Food & Drug Administration, Silver Spring, Maryland 20993, USA, accessed 1. Mar. 2017 http://google2.fda.gov/search?client=FDAgov&site=FDAgov&lr=&proxystylesheet=FDAgov&requiredfields=-archive:Yes&output=xml_no_dtd&getfields=*&ie=UTF-8&ulang=fi&&access=p&sort=date:D:L:d1&entqr=1&entqrm=0&wc=200&wc_mc=1&oe=UTF-8&ud=1&q=%22sample+size%22+inmeta:search_topic%3DMedical%2520Devices&dnavs=inmeta:search_topic%3DMedical%2520Devices

ISO 2859-1, Sampling procedures for inspection by attributes -- Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection, 1999, the International Organization for Standardization ISO

ISO 2859-5, Sampling procedures for inspection by attributes -- Part 5: System of sequential sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection, 2005, the International Organization for Standardization ISO

ISO 3534-1, Statistics -- Vocabulary and symbols -- Part 1: General statistical terms and terms used in probability, 2006, the International Organization for Standardization ISO

ISO 3534-2, Statistics -- Vocabulary and symbols -- Part 2: Applied statistics, 2006, the International Organization for Standardization ISO

ISO 3951-1, Sampling procedures for inspection by variables -- Part 1: Specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL, 2013, the International Organization for Standardization ISO, 2. edition

ISO 10725, Acceptance sampling plans and procedures for the inspection of bulk materials, 2000, the International Organization for Standardization ISO

Minitab 17, statistical software, Minitab Inc. State College Pennsylvania USA, accessed 13. Feb. 2017 https://www.minitab.com/en-us/products/minitab/

MSA, Measurement System Analysis, Reference Manual, 2010,4th edition, Chrysler Group LLC, Ford Motor Company, GM Corporation, ISBN 978-1-60-534211-5, accessed 20. Feb. 2017
https://mechanical-engg.com/applications/core/interface/file/attachment.php?id=2289

SFS-EN ISO 9000, Quality management systems. Fundamentals and vocabulary, 2015, Suomen Standardoimisliitto SFS, 3. Edition

Smithson, Michael 2003, Confidence intervals, Sage Publications, Inc. Ca USA, ISBN 0-7619-2499-X

Tennant, Geoff 2001. SIX SIGMA: SPC and TQM in Manufacturing and Services. Aldershot, UK: Gower Publishing, Ltd. ISBN 0-566-08374-4.

United States District Court, District of New Jersey, 1993, Civil Action No. 92-1744, accessed 24. Feb. 2017

http://academy.gmp-compliance.org/guidemgr/files/1-23-1--WOLIN_JUDGEMENT.PDF

Wickens, Christopher & Hollands Justin 2000, Engineering Psychology and Human Performance, Prentice Hall, New Jersey, 3. edition, ISBN 0-321-04711-7

Webinaria, version 2,1, Open-source, accessed 6. Mar. 2017
http://www.webinaria.com/index.php