YRKESHÖGSKOLAN NOVIA

Introduction of Process Capability Measuring at Oy Mapromec Ab

Marko Widjeskog

Thesis for Bachelor of Engineering

Industrial Management and Engineering

Vasa 2019

EXAMENSARBETE

Författare: Marko Widjeskog

Utbildning och ort: Produktionsekonomi, Vasa

Handledare: Kim Backman (Oy Mapromec Ab)

Mikael Ehrs (Yrkeshögskolan Novia)

Titel: Introduktion av processduglighetsmätningar vid Oy Mapromec Ab

Datum Sidantal 30 Bilagor 0

Abstrakt

Uppdragsgivaren för detta examensarbete, Oy Mapromec Ab, är ett företag som specialiserar sig på tillverkningen av delar för stora motorer med kolvtappar som huvudprodukt.

Processduglighetsstudier har under de senaste decennierna blivit en allt vanligare metod för organisationer att bedöma underleverantörers förmåga att tillverka godtagbara produkter. Flera av företagets kunder har efterfrågat processduglighetsstudier och syftet för detta examensarbete är att undersöka företagets nyckelprocesser samt räkna ut de mest använda processduglighetstalen.

Detta har utförts genom att samla in och analysera den data som normalt sparas i samband med en arbetsorder för två av företagets nyckelprodukter. De vanligaste processduglighetstalen räknas ut och presenteras för de fyra olika attribut som data finns tillgänglig för. Möjligheten att kombinera data från olika produkter för att analysera attributen på processbasis istället för produktbasis undersöks. I samband med analysen upptäcktes några trender i produktionen vilka också presenteras.

Språk: Engelska Nyckelord: Processduglighet, Statistisk processkontroll

BACHELOR'S THESIS

Author:	Marko Widjeskog
Degree Programme:	Industrial Management
Supervisor(s):	Kim Backman (Oy Mapromec Ab)
	Mikael Ehrs (Novia University of Applied Sciences)

Title:	Introduction of Process Capability Measuring at Oy Mapromec Ab

Date	Number of pages 30	Appendices 0

Abstract

This thesis was commissioned by Oy Mapromec Ab, a manufacturing company that specializes in the production of components for large-bore engines with piston pins being the main product.

Process capability studies have, during the past few decades, become successively more common as a method for organizations to assess the performance of their suppliers. As several of the company's customers have requested process capability studies to be conducted, the purpose of this thesis is to investigate the key processes and calculate the commonly used process capability indices.

This has been done by gathering and analyzing data that is normally recorded for work orders, focusing on two of the company's key products. Capability indices are calculated and presented concerning the four attributes for which there are available data. The possibility to combine data from different products to be analyzed jointly is examined. Also presented are some trends in the production output that were found as a result of the analysis.

Language: English　　Key words: Process capability, SPC

Table of contents

1 Introduction

How does one measure quality? The question whether a product can meet its requirements is as old as humankind's ability to produce goods. During post-industrial revolution times, however, the focus has successively moved from examining the individual product to investigating the process responsible for the production. A theoretical framework has emerged around the efforts to statistically quantify the capability of a process to produce goods that meet requirements during the past 40 years. This thesis is an attempt to apply these theories in practice and, to put it briefly, measure quality.

1.1 Oy Mapromec Ab

Oy Mapromec Ab is a medium size company founded in 1994 that employs 60-70 people at a given moment with a turnover of around 14 million euros. The company specializes in the manufacturing of engine components, mainly to be used in large-bore engines for ship and powerplant operations, its customers being the biggest manufacturers of said engines worldwide.

The main product of Oy Mapromec Ab are piston pins, the linking part between piston and connecting rod. A wide range of piston pins are manufactured at the company, the smallest weighing in at 5 kg and going all the way up to 250 kg. Piston pins are subjected to enormous amounts of stress during normal engine operations, and as such the precision required in their manufacturing is extraordinary and the quality of finished products of utmost importance.

1.2 Background

Process capability studies have become an increasingly common way of evaluating an organization's performance since the 1990s, in no small part due to the requirements set by the quality standard QS-9000 introduced by the "big three" American auto makers, General Motors, Chrysler and Ford, in 1994. As a result, capability studies are currently a standard, often required in order to assess suppliers. Several of Oy Mapromec Ab's customers have expressed a wish to ensure the quality of supplied goods by having process capability studies conducted.

1.3 Problem and purpose

In this thesis, the data normally recorded for each work order is analyzed in order to both produce key indicators that would satisfy the need to quantify process performance and to gain a better understanding of the processes in a continuous effort to improve quality. The most commonly used capability indices, C_p/C_{pk} and P_p/P_{pk}, will be calculated. This will be done separately for the four different attributes, diameter, length, surface roughness and cylindricity, for which data is available. The results from a recently undertaken measurement systems study are considered as part of this analysis.

Concluding that analysis results can be generalized to implicate all the process output would be beneficial in making the results of the study more widely applicable. Since different products are being analyzed in this study, a working hypothesis stating that surface roughness and cylindricity are process dependent rather than product dependent will be tested.

Previous attempts have been made at Oy Mapromec Ab in implementing statistical process control and the company has an ongoing, small scale pilot project aiming to test practical feasibility. Hopefully this thesis will provide information helpful in the development and implementation of a more statistics-based quality control.

1.4 Delimitations

Even though the company manufactures a wide range of products, this study will be limited to the two models of piston pins with the highest production volume. The data is gathered from records that are produced as a result of normal day-to-day operations. The four attributes for which there are continuous data are the focus of the analysis. Results from ultrasonic testing, magnetic particle inspection, hardness or cleanliness testing will not be included as they either represent discrete variables or processes which are not directly connected to the on-site manufacturing.

1.5 Thesis structure and confidentiality

The thesis is divided into the following headings:

1. Introduction
2. Theory – Some historical considerations and the theoretical framework for statistical tools used in the analysis.
3. Methodology – Presentation of the data and how it has been measured and recorded. Problematic features are given some consideration. This section will have some of the numbers omitted in the publicly available version.
4. Analysis – The results are presented. This section will be redacted in the publicly available version of this thesis.
5. Discussion – Central findings and recommendations.
6. Conclusions – Summary and closing thoughts.

2 Theory

In this chapter, the theoretical framework will be presented. The commonly used process capability indices will be explained along with a brief look at normal distributions, the cornerstone on which these indices rely. As some of the data analyzed in this thesis is non-normally distributed, there is a section discussing possible solutions.

Special attention is given to estimating standard deviations, which is perhaps the most exacting aspect of the calculations. Statistical hypothesis testing and trend analysis will also be given a short introduction, these methods are employed in the analysis section to make inferences.

Finally, when it comes to analyzing measured results, some consideration should be given to the accuracy and precision of the measurements. The Gage Repeatability & Reproducibility-method for evaluating measurement systems is presented.

2.1 Statistical quality control

During the 20th century, a method for controlling quality by statistical means was developed, originating at the Bell Laboratories in the 1920s with the concept of *state of statistical control* and the introduction of the *control chart*. A process is in statistical control when no *special cause variation* is present, i.e. all the variation can be ascribed to properties inherent to the process, or *common cause variation*. (Bass 2007, 14-16). Control charts were developed in order to check whether processes are in control or not, to pick out signals from the noise. Note that a process which is in control is not necessarily stable, long-term change from internal properties may still occur.

Statistical Process Control (SPC) grew out of these early efforts. In a nutshell, SPC relies on the aforementioned control charts to weed out signals that indicate a process is drifting out of control. During normal, in-control operations, about half of the measurements should be above target and half below target. Even in a perfectly centered process, individual measurements will naturally be off target and adjusting the process will in such cases be counterproductive. Similarly, the opposite is true, not reacting to significant changes in a process is also detrimental. The control charts’ raison d’être is to statistically indicate when a process demands adjustment in order to avoid such over- or under-correction. (Bass 2007, 146-147).

As SPC was further developed, process capability analysis branched out as a method to assess the extent to which processes are expected to produce goods that conform to specifications (Qiu 2014, 102). A famous example is Motorola's Six Sigma concept, based on both capability studies and a variety of statistical improvement methods. The program resulted in savings of 11 billion USD and tripled the company's productivity worldwide (Deleryd 1998, 1). The concept was later spearheaded by General Electric and saw a spectacular rise and fall together with GE's stock price. Nevertheless, Six Sigma remains a familiar reference in quality jargon.

In conjunction with the introduction of the quality standard QS-9000 in 1994 within the automotive industry in America, capability studies have become a staple of assessing suppliers. QS-9000 requires that suppliers of a certified organization are using capability studies or are at least in the process of introducing them, which has resulted in more and more organizations adopting their use (Deleryd 1998, 2).

2.2 The normal distribution

The normal, or gaussian, distribution describes a continuous random variable that has the following probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where μ is the population mean and σ the standard deviation. The resulting function takes the shape of a bell, which is why it is sometimes also referred to as the bell curve. The normal distribution is important, because much statistical theory is developed under the assumption of normality and because many natural distributions can in fact be described fairly well by it (Qiu 2014, 15). The importance is further emphasized by the central limit theorem, which states that the sum of sample means from independent random variables will approach a normal distribution when the sample size grows larger, regardless of the population distribution. (Montgomery & Runger 2014, 243).

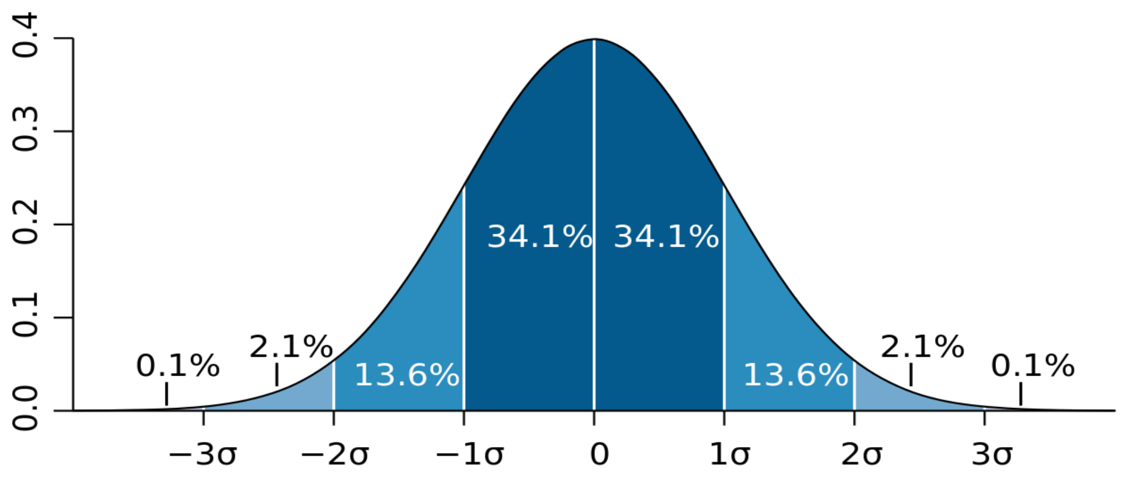

Figure 1: The normal distribution and corresponding area percentages at different levels of standard deviation (Wikipedia).

Figure 1 depicts a standard normal distribution. It is a special case of normal distribution where mean μ = 0 and standard deviation σ = 1, but the area percentages are the same regardless of mean and standard deviation. What it tells us is that in a normally distributed population, 68,27% of observations will fall within μ ± 1σ, 95,45% of observations will fall within μ ± 2σ and 99,73% of observations will fall within μ ± 3σ. (Salomäki 2003, 197; Oakland 2003, 89). The 3σ limit on both sides of the mean is sometimes referred to as the *natural tolerance limits*, an in-control process should mostly produce results within this range. Consequently, 6σ can be called the *total process spread*. (Pitt 1994, 322; Montgomery & Runger 2014, 689).

2.3 Process capability analysis

For a numerical representation of process capability, *process capability indices* have been developed as dimensionless measures based on output properties and product specifications. The first and simplest of these, C_p or capability index, was introduced by Joseph Juran in 1974 and is defined as:

$$C_p = \frac{USL - LSL}{6\sigma}$$

where USL is the upper specification limit, LSL the lower specification limit and σ is the standard deviation of process output. (Deleryd 1996, 25). In other words, C_p represents the allowed process spread divided by the actual process spread or rather 99,73% of the actual process spread in a normal distribution (Bass 2007, 175). It is thus important to note that in order to analyze process performance, the process output must follow a normal distribution reasonably well. The capability index does not consider whether a process is on target, only how its spread relates to its tolerancing and is therefore sometimes referred to as *potential process performance*.

Since the capability index is quite limited in describing how a process behaves, it was complemented in 1984 by Victor Kane in the development of the C_{pk}-index, which factors in whether a process deviates from the middle of its specification range. The calculation is twofold, split into upper and lower capability indices. These are defined as:

$$CPU = \frac{USL - \mu}{3\sigma} \qquad CPL = \frac{\mu - LSL}{3\sigma}$$

where μ is the mean process output. C_{pk} is the lesser of these values or:

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}; \frac{\mu - LSL}{3\sigma}\right)$$

An alternative way of calculating C_{pk} is to multiply C_p by a factor (1-k), a scaled distance that indicates the extent to which the process is off-center:

$$k = \left|\frac{T - \mu}{\frac{1}{2}(USL - LSL)}\right|$$

where T represents the target value of the process. C_{pk} is then acquired by:

$$C_{pk} = (1 - k)C_p$$

It follows that $C_{pk} \leq C_p$ since $0 \leq k \leq 1$. In other words, in a perfectly centered process, C_{pk} and C_p are equal. (Bass 2007, 181; Kane 1984, 45-46).

Figure 2 illustrates the meaning of the indices. Depicted above are the probability density functions of two versions of a perfectly centered process, the blue line depicts output from a process that is described by a standard normal distribution, i.e. it has mean = 0 and standard deviation = 1. In this case, both C_p and C_{pk} = 1.

$$C_p = \frac{3 - (-3)}{6 \cdot 1} = 1 \qquad C_{pk} = min\left(\frac{3 - 0}{3 \cdot 1}; \frac{0 - (-3)}{3 \cdot 1}\right) = 1$$

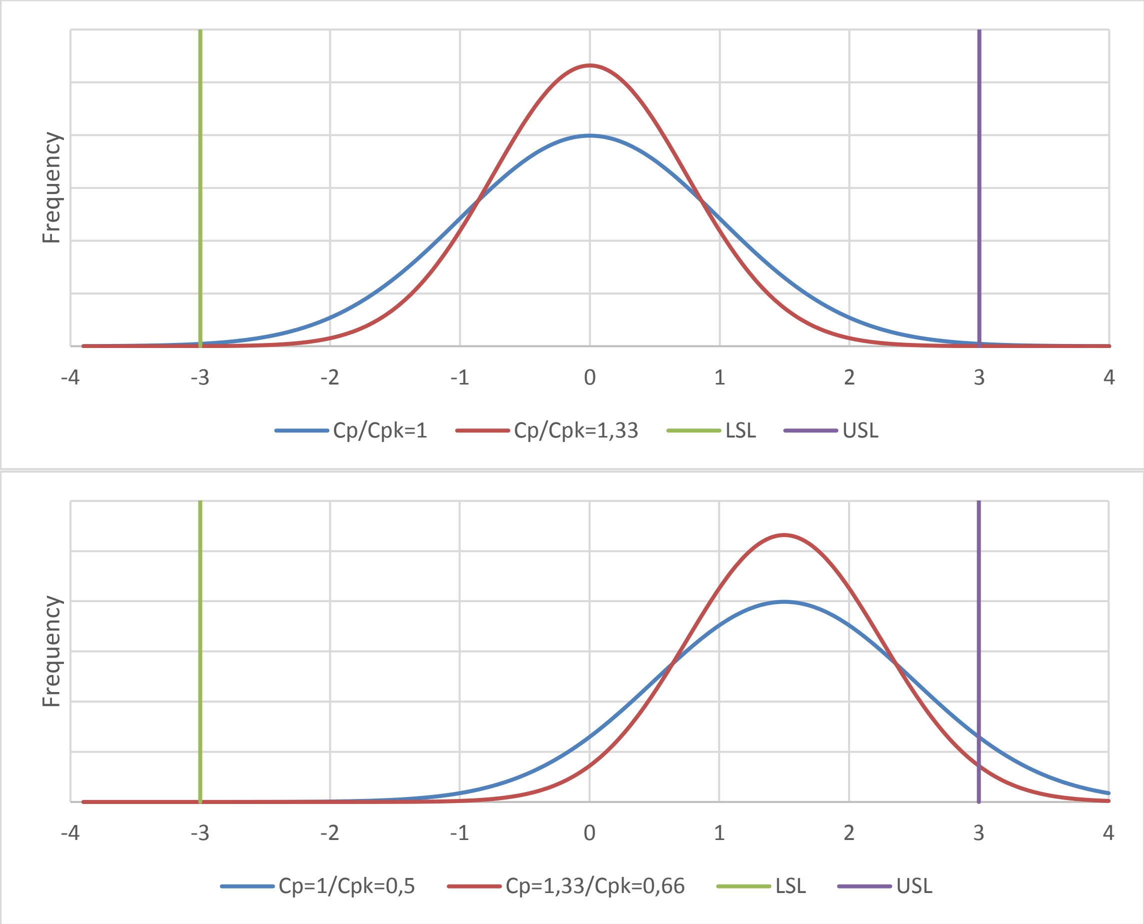

Figure 2: Idealized versions of process outputs – centered (above) and off-center (below)

If the process spread is under better control and the standard deviation decreases to 0,75, we arrive at the process depicted with the red line. As the spread decreases, the values of C_p and C_{pk} increase, in this case to 1,33.

In figure 2, depicted below are the probability density functions for a process that is off-center with the same standard deviations as above but with a mean of 1,5. Both C_p-values stay the same as the index is location independent but C_{pk} decreases as the process moves further from center. A tighter spread still yields a better result, the blue line with standard deviation = 1 has a C_{pk}-value of 0,5 and the red line with standard deviation = 0,75 has a C_{pk}-value of 0,66.

In the case of one-sided specifications, the C_p-value becomes useless but C_{pk} is still valid (Salomäki 2003, 195; Oakland 2003, 265). In these cases, the relevant part of the C_{pk}-calculations is sufficient, i.e. either CPU or CPL, as in the situation above where CPU was used to calculate C_{pk}.

As was discussed earlier, 99,73% of observations fall within a ± three sigma interval in a normal distribution. Since the standard process capability indices require a relatively normal set of data, the three-sigma level corresponds to a C_{pk} of 1, i.e. the process mean ± three standard deviations are within specification limits. A standard practice within industry has become a C_{pk}-requirement of 1,33 (Bergquist & Albing 2006, 966) which corresponds to a four-sigma level of conformity to specifications or 99,994%. Other numbers used to describe what sigma-level or C_{pk} refers to in practice are failures per million opportunities and time wasted per 720 h (a month) as shown in table 1. (Bass 2007, 20; Salomäki 2003, 205; Oakland 2003, 357). In addition, the so-called Six Sigma quality program, mentioned in 2.1, refers to a C_{pk}-level of 2 with a theoretical rate of one failure out of 500 million opportunities.

Table 1: Sigma levels and corresponding factors

σ	Conformity %	C_{pk}	PPM	Time wasted / 720 h
±1	68,26 %	0,33	317400	228,5 h
±2	95,46 %	0,67	45500	32,8 h
±3	99,73 %	1	2700	1,94 h
±4	99,994 %	1,33	63	2,74 min
±5	99,99994 %	1,67	0,57	1,49 min
±6	99,9999998 %	2	0,002	0,005 s

2.4 Other indices

In addition to C_p and C_{pk}, which are the most widely used (Bergquist & Albing 2006, 966), additional indices have been developed. Among the better known are the performance index and the machine capability index. The P_p/P_{pk} or the performance index is calculated using the mean of all observations and overall standard deviation. There is no need to adjust the standard deviation with constants (Salomäki 2003, 200). Special causes of variation are included in the results, as part of the process. The index can never exceed C_p/C_{pk} unless the process has in fact improved (Oakland 2003, 267).

C_m/C_{mk} or the machine capability index is calculated exactly like C_p/C_{pk}, the difference being that the sampled products should be manufactured in sequence, in as little time as possible under as similar circumstances as possible. This is done in order to minimize all variation except that which arises from the machine or process being scrutinized. A minimum of 50 samples is recommended in order to measure C_m/C_{mk}, and the resulting indices have higher requirements, with 1,33 – 1,67 regarded as barely capable, ≥ 1,67 as capable (Salomäki 2003, 199-200).

2.5 Estimating means and standard deviations

It is usually not possible to know the actual process mean and standard deviation. When sampling, the sample mean is denoted with $\bar{x}$ while the mean of sample means, or the process mean, is denoted $\bar{\bar{x}}$. The letter s is used to denote a sample standard deviation. Particularly with small sample sizes, the calculated standard deviation has a tendency towards being underestimated. That is why the sum of the squared deviations is divided by sample size minus one instead of the actual sample size. (Oakland 2003, 84-87; Salomäki 2003, 179-181).

Since the process standard deviation tends to be an unknown, it cannot be accurately calculated nor replaced by a sample standard deviation. An estimated standard deviation, denoted $\hat{\sigma}$, is used to calculate process capability indices. Different methods are applied in these calculations and since, in this thesis, data from different variables is gathered at varying intervals, three separate methods will be employed:

1) For variables where the data for all units is available, there is no need for estimates. P_p/P_{pk} is calculated straight from the population mean and standard deviation while C_p/C_{pk} uses $\bar{\bar{x}}$ and the mean of subgroup standard deviations.

2) For variables with 10% sampling, a pooled standard deviation will be used to calculate the within-subgroup standard deviation. The formula for this is:

$$\hat{\sigma}_{within} = \frac{Sp}{C_4(d+1)} \qquad Sp = \sqrt{\frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{\sum_i (n_i - 1)}}$$

Above, x_{ij} represents the j:th observation in the i:th subgroup while $\bar{x}_i$ is the mean of the i:th subgroup and n_i the number of observations in the i:th subgroup. C_4 is an unbiasing constant, which can be found in tables, for the value (d+1), the sum of subgroup sizes plus one. The value for C_4 approaches one as the sum increases. The numerator for Sp is calculated by summing the squared differences between individual values and their respective subgroup mean. The denominator is the sum of subgroup sizes minus one.

3) For variables with sample size = 1, the within subgroup-calculations need another approach. An average moving range-method is employed:

$$\sigma_{\bar{x}} = \frac{\bar{R}}{d_n}$$

Where $\bar{R}$ is the mean range or the sum of differences between consecutive values divided by the number of observations – the width of the moving range + 1. The result is then divided by the unbiasing constant d_2 for the value of the range width. The moving range method is also applied in the $\bar{x}$/R-control chart. (Steiner, Bovas & MacKay 1997, 5).

2.6 Non-normal distributions

Because inferring proportions of nonconformity from process capability indices assumes a normal distribution of results, having non-normal data can result in a very large proportion of nonconformity even for values of C_{pk} that would, under the assumption of normality, yield acceptable results (Deleryd 1996, 143; Levinson 2011, ix). It is quite common that process output has a non-normal, often skewed, distribution, especially in the case of one-sided specifications where one of the ends is limited by a natural, in practice unreachable, boundary. One such example would be surface roughness (Deleryd 1996, 31-32).

While the whole matter of dealing with non-normal distributions is discussed at some length by e.g. Levinson (2011) as the issue is complex, for the purposes of this thesis a brief overview of the procedures employed should suffice. There does not seem to be a consensus regarding the proper method of dealing with non-normality (Deleryd 1998, 27).

Broadly speaking, there are two different approaches to take. The first is to find another distribution that closely matches actual process output and conduct the analysis with the underlying assumption that the distribution reflects reality (Levinson 2011, 61-63). The second approach is to transform the data to resemble a normal distribution more closely (Deleryd 1996, 147-149; Levinson 2011, 57-61).

The upside of using transformed data is that a standard process capability analysis can be performed, while fitting the data to another distribution model entails calculating P_p and P_{pk} not from the mean and standard deviation but based on the parameters for the particular distribution that is being modeled. Values for C_p and C_{pk} cannot be obtained this way (Bass 2007, 200-201).

Common methods of transforming data include using the base-10 or the natural logarithm of values, or carrying out a Box-Cox transformation according to:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} \; if\ \lambda \neq 0 \\ \log y \;\; if\ \lambda\ = 0 \end{cases}$$

After which all values for -5 ≤ λ ≤ 5 are tried and the best approximation for a normal distribution selected. This is quite laborious to do manually but most statistical software include an option for the transformation. (Bass 2007, 195).

2.7 Statistical hypothesis testing

The testing of statistical hypotheses begins with the assumption of the *null hypothesis* (H_0) with obvious parallels to the presumption of innocence in criminal trials. The null hypothesis cannot be proven to be true, accepting it simply means that there isn't enough evidence to support the *alternative hypothesis* (H_1) (Levinson 2001, 6). There is always a risk present that a hypothesis is rejected even when it is actually true, or on the contrary that a hypothesis is not rejected even when it is actually false. These are called *alpha*- and *beta errors* respectively and they are inversely related (Bass 2007, 123; Montgomery & Runger 2014, 308-309). Usually testing is performed with a five percent *significance level* and a 95% *confidence interval*. Significance level, denoted α, indicates a risk percentage of making an alpha error i.e. rejecting a true null hypothesis. A confidence interval signifies a range of values within which an estimation can, with the level of confidence indicated, be assumed to be contained (Triola 2014, 351-352).

When interpreting results of hypothesis tests, statistical significance can be inferred from the resulting p-value. If the p-value ≤ α, the null hypothesis must be rejected (Bass 2007, 126-127).

The *two-sample t-test* is used to determine whether two independent groups are different based on a sample of both groups. The null hypothesis for these tests states that the samples are equal and the alternative hypothesis that a difference between them exists. The test assumes normal distribution of populations, but when the size of the samples is large (30+) the test is quite robust to violations of normality. (Minitab LLC. 2017).

Equivalence tests can be used to determine whether two samples can be considered equivalent. Contrary to the two-sample t-test, the burden of proof is placed on proving equivalence. The null hypothesis is in other words that the sample mean differs from a reference. When testing for equivalence, a range must be specified that is within an acceptable distance from the reference to still be considered equal. (Minitab LLC. 2019a).

2.8 Trend analysis

In this thesis, *simple linear regression* will be used to analyze trends. In linear regression one or more predictor or independent variables and a dependent response variable are considered. If there is only one independent variable, the method is called simple linear regression (Montgomery & Runger 2014, 431). The procedure finds a function for a straight line that predicts, as accurately as possible, the dependent variable as a product of the independent variable. This is achieved by the principle of least squares, i.e. minimizing the sum of squared vertical distances between the function and the data points (Johnson 2018, 327-330; Allen 2007, 345-346).

A confidence interval can be plotted for a regression model, in this case it indicates, within a certain level of confidence, possible configurations for the fitted line. Similarly, a prediction interval displays the probable width of the data points' spread. (Johnson 2018, 338-345).

When analyzing results, the R^2-value indicates the percentage of variation in the data that is explained by the linear relation (Johnson 2018, 348). A p-value can also be calculated, its function equivalent to hypothesis testing, i.e. a p-value of $\leq$ 0,05 indicates statistical significance (Triola 2014, 559).

2.9 Gage R&R

For any conclusions inferred from data to be relevant, the reliability of the data must be assessed. All data analyzed in this thesis are results from measurements and it follows that an evaluation of the measurement systems should be conducted in order to verify that the data represents actual variation in the measured quantities and not variations in the measurement systems.

In order to assess whether the measurement system in use is adequate, a commonly used method is the Gage Repeatability and Reproducibility test, also known as Gage R&R or simply GRR. The test is performed by having different appraisers measure a series of test pieces on multiple occasions where the total of number of measurements, i.e. number of appraisers multiplied by the number of test pieces should be a minimum of 15 (AIAG 2017, 104).

The results of the GRR-test splits variation into three categories:

1) Within appraiser or repeatability represents variation that arises when the same appraiser measures the same test piece using the same equipment.

2) Between appraisers or reproducibility represents variation between the different appraisers while measuring the same test piece using the same equipment.

3) Part-to-part represents the actual variation between the measured test pieces. (AIAG 2017 54-56).

For the actual formulas to calculate GRR, refer to AIAG (2017) or Salomäki (2003). The resulting percentages are judged as:

Acceptable < 10% Marginal < 30% Unacceptable (AIAG 2017, 78; Levinson 183)

The observed process variance is the sum of the actual process variance and the measurement system variance:

$$\sigma_{Obs}^2 = \sigma_{Act}^2 + \sigma_{Meas}^2 \quad \rightarrow \quad \sigma_{Act} = \sqrt{\sigma_{Obs}^2 - \sigma_{Meas}^2}$$

A poor measurement system is subsequently compensated for by improving the capability index accordingly, since some of the process variation is attributed to measurement system variation (Levinson 2011, 185-186). When GRR is a known quantity, the relation between actual C_p and observed C_p is as follows:

$$C_{p(Act)} = \frac{C_{p(Obs)}}{\sqrt{1 - GRR^2}} \qquad GRR = \frac{k\sigma_{Meas}}{6\sigma_{Obs}}$$

The constant k refers to the level of coverage to which the measurement error is represented. Historically, a value of k = 5,15 has been used to represent a 99% spread, while the process spread, i.e. 99,73%, is attained with a value of k = 6. (AIAG 2017, iv, 198-199). In this thesis, the process spread is used.

3 Methodology

The production of piston pins at Oy Mapromec Ab is done in batches of between one and 250 units with the average estimated to be around 50. Measuring process capability from batch production is something of a challenge since products of different dimensions, tolerances and sometimes other defining characteristics (e.g. piston pins with or without oil holes) are machined at the same production line, using the same process.

For this analysis, to keep the scope manageable, two of the most commonly manufactured high-quantity products are focused on. These were chosen as higher quantities should present more reliable statistics and because these products, constituting nearly 50% of the company's piston pin production, have a higher relevance for the bottom line than any other pair of products.

A third product was initially included in the study, but as its production had recently been moved from one line to another there was a lack of current data. It also seemed prudent to limit the study to the one production line at which both products under review are being manufactured at. The products will henceforth be referred to as products A and B, product A being the company's current best seller.

Another approach to the analysis was briefly considered; instead of choosing to focus on two products, all the products would have had their data normalized to e.g. percentiles of their respective specification ranges. It did not take very long to realize that this approach was quite fruitless as the resulting numbers varied widely between different products. The reason for this was mainly the different specification limits, even though the machine produces similar results, large discrepancies could be seen in the indices due to the sometimes much stricter requirements. Another reason is that the process does in fact seem to produce somewhat different results depending on the dimensions being worked as can be seen in the non-interchangeability between product data for length and diameter (see 4.4 for more details).

An attempt was made to find a suitably homogenous, large number of consecutively manufactured units for a C_m/C_{mk}-analysis but this proved to be a tall order. If the units should all be contained within a single heat treatment batch, the product must also be on the smaller side for 50 units to be included in a single batch. This undertaking was finally abandoned but would certainly make for interesting results if a dedicated experiment was to be set up.

3.1 Gathering the data

The target was to have 25 batches of each product to analyze. Hence, with the higher volume and pace of production in the case of product A, the time frames from which the data has been gathered are dissimilar. In the end, the data for product A consists of a total production volume of 2351 units and product B 1523 units.

- Data for product A has been gathered over the period 11/2018 – 7/2019
- Data for product B has been gathered over the period 8/2018 – 8/2019

The exception to this is results regarding cylindricity, in order to obtain a higher number of observations, product A goes back to 7/2018 and product B to 11/2017. While doing comparisons between the two products, the time period used is discussed case-by-case, as the results may differ in drifting (experiencing a long-term change in the mean output) processes.

3.2 About the attributes

For each work order, a document is compiled in which four attributes, diameter, length, surface roughness and form, are recorded. The document is often included in deliveries as per customer request, or simply stored for future reference. The analysis was conducted on data gathered from these documents. The methods used to measure and analyze the attributes are presented below.

Table 2: Example of raw data from a piston pin measurement record

MM	NO:	1	2	3	LENGTH	CYLINDRICITY	Ra
18M1	05549	14	15	13	75		0,06
18M1	05526	8	11	10	75		
18M1	05554	13	15	14	75		
18M1	05531	9	11	11	75		
18M1	05538	10	11	8	75		
18M1	04534	10	12	11	75		
18M1	04318	10	12	10	75		
18M1	04517	11	13	11	75	0,003	0,08

3.2.1 Diameter

In figure 3, step 20 of the manufacturing process entails measuring all dimensions and surfaces. This includes measuring the diameter of pins at three locations, top, center and bottom, for 100% of the batch. The measurements are performed by the machine operators using a snap gage with a digital dial indicator.

The tolerances for the diameters are nominal – [redacted] μm (product A) and nominal – [redacted] μm (product B), therefore the measurements are recorded as amounts of microns less than nominal. In cases such as this the target value naturally becomes not the nominal but the middle point of the tolerance range, i.e. -[redacted] μm and -[redacted] μm respectively. For the sake of practicality, however, these are converted into absolute values.

As three distinct, albeit usually closely related and optimally identical, values are recorded for each pin, the question arises which of these values should be used to describe the diameter of a given piston pin. From a perspective that focuses on product functionality, the choice would probably be the value closest to nominal, i.e. the largest diameter. After a discussion with the manager of quality inspection, the mean value of the three measurements was selected as best representing a measure of overall quality.

3.2.2 Length

As with diameter above, length is measured simultaneously and for 100% of the batch using a digital caliper. The tolerances are substantially wider than those regarding the other recorded variables, nominal – [redacted] mm. To keep things simple, measurements have been converted to be displayed in the same manner as diameter i.e. the values are subtracted from nominal. Of some reason, the values are recorded rounded to the nearest 0,05 mm although the measuring resolution is 0,01 mm.

Plans are in place to update the process in this regard to include a length adjustment at step 15 of the manufacturing process in figure 3. Currently, the somewhat unpredictable effect of heat treatment sets the length, and if nonconformity is detected at the post heat treatment measuring, short pieces are scrapped while long pieces are reworked during step 15. Length is the one attribute analyzed here that is independent of step 16, grinding and lapping. In other words, the other variables are all, to a degree, measurements of the part-process where the outer surface is finished while length measures an altogether different part-process, which would be the pre heat treatment turning and the subsequent shifting of form during heat treatment.

3.2.3 Surface roughness

A 10% sampling rate is required for surface roughness. This is, once again, performed during the measuring phase together with diameter and length using a portable surface roughness tester except for a single sample, which is measured by a quality inspector in conjunction with form measurement.

The result is recorded with a resolution of 0,01 Ra, which might provide a problematic distribution, especially in the case of a one-sided tolerance. On the other hand, recording further decimals seems like wasted effort since surface roughness is, by its nature, quite impossible to measure accurately. Placing the tester in different locations of a test piece often yields different results, and the specification limit itself is supposed to represent at mean of five measurements, unless explicitly otherwise stated. As a rule of thumb, if the first result is ≤ 70% of the specification limit, the measurement is accepted as such. If that’s not the case, there is a somewhat convoluted process to follow which will not be described here, see Mitutoyo Engineers’ Reference Book for further details.

While analyzing the data, it was found that results regarding surface roughness differed significantly between measurements taken by the machine operators and by the quality inspectors. While the reason for this discrepancy is unknown, the analysis will proceed under the assumption that both measurements are internally consistent but mixing of these measurements should be avoided. Results gathered by the quality inspectors were therefore omitted, having an order of magnitude more results from the machine operators tipped the scale towards using their measurements even though it might well be that the measurements taken by the quality inspectors are in fact closer to the true values. The omitted measurements had slightly higher values than those in the datasets used for analysis. Since a GRR-analysis has not been performed regarding surface roughness, it would be sensible to do so using both testers that the results are based on.

3.2.4 Form

In most cases, including products A and B, form testing entails testing for error in cylindricity.

As the measuring of cylindricity is a rather slow procedure, depending on the method 10-20 minutes, only one piston pin per production batch is taken aside, measured by a quality inspector and recorded with a resolution of 0,001 mm. Once again, the resolution could be higher for a better distribution. Cylindricity is somewhat arbitrary in that the operator can manipulate results by using different measurement settings and methods for analysis. A true value is, then, open to interpretation. As long as the settings are internally consistent though, the data should be valid for analysis. This has been the case with LSC (least square cylinder) method being utilized with a 50 UPR low pass gaussian filter.

Five measurements of product A and one measurement of product B were omitted (7/2018 period) due to a clear case of special cause variation.

3.3 Problems associated with the data

There are a few problematic features in the data that has been gathered for this study, chiefly that units which are deemed to not conform with specifications are not recorded. Hence, the calculated indices are not actually representative of process performance but instead the quality of delivered goods, or what Oakland (2003, 269-270) calls $C_{pk(delivery)}$. The situation is not exactly what is referred to by Oakland, who describes circumstances where a bad batch is detected by sampling and consequently the entire batch is discarded. There is no discarding of entire batches in this case, non-conformity is usually found in the 100% measurement categories within the rare single workpiece, which is then either scrapped or reworked if possible. Overall, results ought not be significantly altered by the miniscule amount of data missing. For future reference though, it would be quite simple to include nonconforming units in the measurement record, e.g. in a dedicated worksheet. This would improve data quality for further similar undertakings.

There is a lack of granularity in the data, most notably in the cases of length and cylindricity, but none of the attributes can be reliably tested for distribution. The assumption is made that values are in fact normally distributed, which seems almost certain in the case of diameter and probable in the case of length and cylindricity. The exception is surface roughness, which is clearly not normally distributed. In the case of cylindricity, and arguably surface roughness, recording an additional decimal would improve any analysis in the future. Length should equally be recorded with a 0,01 mm resolution.

Chronology is another problem. Within a subgroup (or a work order), measurements are not in order of production but somewhat randomly entered. It has never been deemed important that the correct order be maintained, and it has probably never caused any problems before someone wanted to analyze trends in the output. When comparing different subgroups, the chronology is largely correct but not 100%, since there is always the odd piece out that e.g. needs to be reworked and included in the next batch instead. This, however, is in the grand scheme of things a minor annoyance and should not significantly alter results. Trend analysis in in other words best performed with batch means instead of single unit measurements although in the end, there was very little difference between the two.

4 Analysis

This chapter is redacted in the publicly available version of the thesis.

4.1 Gage study results

Table 3: GRR results

4.2 Inferences from the data

Figure 4: Results from equivalence test / cylindricity

Figure 5: Results from equivalence test / surface roughness

4.3 Trends

Figure 6: Surface roughness as a product of time

Figure 7: Diameter of product A as a product of time

4.4 Capability analysis

4.4.1 Diameter

Figure 8: Histograms illustrating process spread of diameter

Table 4: Diameter, key values and calculated indices with GRR-adjusted values included

4.4.2 Length

Figure 9: Histograms illustrating process spread of length

Table 5: Length, key values and calculated indices

4.4.3 Surface roughness

Figure 10: Process capability report for surface roughness using data transformed to its logarithm of 10, specification limits set for product B

Figure 11: Process capability report for surface roughness based on a three-parameter gamma distribution model, specification limits set for product B

4.4.4 Form

Figure 12: Process capability report for cylindricity, specification limit set to product B

Figure 13: Predicted index values for given specification limits / cylindricity

4.5 Analysis summary

5 Discussion

There are several factors presented in this thesis which indicate that measuring process capability within the circumstances outlined is a less than optimal solution. First and foremost, is there any benefit in conducting a study on attributes with 100% measurements? The answer is both yes and no. From a scholarly perspective, no new information is gained by the analysis. All defective units will, in theory, be found. The actual population means and standard deviations can be calculated and the amount of defective units is known. There are no estimations to be made. On the upside, we have gained knowledge of our position on an arbitrary scale, which in practice might be of some value since said scales represent normalized, comparable values that customers are interested in. From a process improvement perspective, the tangible value in conducting an analysis in the 100% measurement categories would be in shifting focus from a binary way of looking at output in the categories “OK” or “defective” and instead highlighting the importance of process variation and its improvement.

Drifting processes

Another feature in the analysis were the drifting processes. The question arises whether producing indices from these processes is a fair assessment. Does the resulting index reflect reality in any sense? Although it was outside the scope of this thesis, the output of these processes could be normalized to follow a straight line in order to better analyze the momentary process spread and thus separate the two problems of long-time process drift and short-time variation. This is, to an extent, already achieved by the C_{pk}-index as opposed to the P_{pk}-index by differentiating between subgroups. The problem with P_{pk} is the time frame under analysis. For results to be representative of actual performance, the period should comprise the whole interval during which the process drifts, e.g. if the deterioration of a machine part causes a process to drift, the representative P_{pk}-index should comprise of data gathered during the lifetime of said machine part. (Steiner et.al. 1997, 10-11).

Regarding the trends discussed in 4.3, it seems obvious that something is being worn down and having a detrimental effect on surface roughness, which is normal in a process such as this. Maintenance should be scheduled with some haste as it would seem like the process in its current state is just barely keeping up with the requirements for product B. There are several products produced in smaller quantities that have even stricter requirements and chances are that entire batches will be borderline nonconforming. Because the two sets of measurements, taken by the quality inspectors and the machine operators respectively seem to disagree to an extent, conducting a Gage R&R-study regarding surface roughness could shed some light on the issue.

The diameter drift, on the other hand, defies explanation and should be closely monitored until such a time when it can be reliably assessed whether a problem exists and what can be done about it. While tools naturally wear out and produce drifts such as this, the fact that two products that undergo the same process do not exhibit the same kind of drift is peculiar. The production volume of product A being greater may be a contributing factor, but a definite explanation would involve a more in-depth study.

Improving the reliability of measurements

The Gage R&R-results show that the method of measuring diameters with a snap gage is not necessarily precise enough. For the products under analysis results are marginal, for products with stricter specifications, results quickly shift toward the unacceptable range. If the process is not fundamentally changed there is not much that can be done about the fact, as some research has been made into more precise gages without much luck. On the other hand, measuring the piston pins in a coordinate measuring machine (CMM) would certainly provide adequate precision but measuring 100% would be excruciatingly slow. While possible in theory by building a dedicated robot cell for measuring, a more realistic approach would be to move toward sampling and measuring the samples in a CMM.

This approach would also necessitate a ubiquitous shift towards an SPC-based system with implementation of suitable control charts to monitor process performance, emphasizing prevention instead of correcting problems retroactively. All in all, it would entail much work to make the necessary changes, but a successful implementation would both improve the reliability of measurements and more importantly eliminate the need to manually measure 100% of the units. As an additional advantage, measurement resolutions would increase, which would benefit any future undertakings in statistical quality control.

There are already plans in place for a database in which all the pertinent data from single piston pins would be stored. The implementation of such a database would also be a boon in the introduction of SPC as, hopefully, data could be mined directly from the database for control charts. The measurement records would obviously be compiled using data derived from the database as well.

Machine capability

To gain additional information about the process, a dedicated machine capability study could be set up quite easily. The main problem is finding a large batch that is homogenous enough, i.e. it should be composed of units manufactured from the same material batch and that have undergone heat treatment simultaneously. The heat treatment requirement is the difficult one as the smaller of the products analyzed undergoes heat treatment in batches of 35. A compromise could be made and include two batches or alternatively select another, smaller, product for this experiment. This series of units should then be machined in sequence without making any adjustments during their run, preferably in the middle of a larger batch so that the machines would be warmed up when processing the experimental batch. These could subsequently be measured using a CMM for the most reliable, high resolution data.

Evaluating the predictions

An interesting conclusion to this study is comparing the predicted numbers of defects with the actual defects found in production within the 100% measurement categories. Keeping it simple, the predicted amounts will be derived only from P_{pk}.

Table 6: Predicted vs. found defective units for corresponding time frame

		Diameter	Length
Product A	Predicted (pcs.)	11,5	93,0
	Actual (pcs.)	8 (+3)	13
Product B	Predicted (pcs.)	4,2	3,9
	Actual (pcs.)	4 (+3)	2

Numbers in brackets indicate defects by apparent special cause variation that should not be included, the predictions are based on defects that should arise as a natural consequence of the common cause variation inherent to the process. As can be seen in table 8, the predictions are quite good approximations for product B and the diameter is a good match for product A, but there is a huge discrepancy between prediction and reality concerning length of product A. The length of product A was also the index with the poorest values of all. If there is a conclusion to be drawn, it would be that something has gone wrong with the data, but just looking at the histogram that displays the process spread does make one wonder how the results can stay within the specification limits as well as they do. Data resolution may be the main problem and the distribution's tails could look significantly better when dividing the outermost categories further, which would also indicate that the results regarding length presented in this thesis are not valid. As there seems to be no good reason to arbitrarily round off the measured values and it comes at no additional cost, the recording of length should be done with the highest possible resolution.

6 Conclusions

This thesis has been an attempt to measure process capability at Oy Mapromec Ab. As the company manufactures various products in batches of different sizes, the task was not straightforward. A delimitation was made to concentrate the efforts on the company's main product, piston pins, and therein to only the two models which constitute the highest revenue for the company.

By default, data is recorded in four subcategories for batches of piston pins. These records have been used to conduct the analysis while also taking into account a recent Gage Repeatability and Reproducibility study carried out at the company.

A working hypothesis in this thesis was that the process under investigation produces similar results regarding surface roughness and cylindricity regardless of which product is being manufactured. After testing failed to reject this hypothesis, the attributes in question were analyzed jointly. In the process of performing the analysis, two apparent trends in the production output were revealed. These are also briefly discussed.

Data quality presented some problems while conducting analyses for this thesis. The data is recorded at a lower than optimal resolution and measurements of nonconforming units are not present in the datasets. This has an impact on the reliability of results and there is not much that can be done about it, except recommending improvements in the recording of data for more reliable analyses in the future.

The most commonly used capability indices, $C_{p,}/C_{pk}$ and P_p/P_{pk} were calculated in order to illuminate the company's ability to produce goods that conform to specifications. As normally distributed data is a prerequisite in order to calculate these indices, one of the problems encountered during this work was that some of the data did not follow a normal distribution. Consequently, the methods employed were data transformation and analysis following the parameters of a gamma distribution. Additional research into analysis of non-normally distributed attributes could be advantageous as this is covered only superficially.

It is important to remember that the findings presented here are only a beginning, in order to reach quantifiable improvement, they could represent a reference point. In a study conducted within Swedish industry, some half of the respondents claimed to conduct some manner of capability studies. The commonly given answers when asked how long it would take for all processes to be capable was 5-20 years. Some respondents even concluded that all processes would never be deemed capable since requirements were made stricter at the same pace as the processes were improving. (Deleryd 1996). Never reaching the goal is not necessarily a bad thing. In order to stay competitive, complacency is the enemy and always having a goal to strive for is the mark of progress.

7 References

Allen, T., 2006. *Engineering Statistics and Six Sigma.* London: Springer.

Automotive Industry Action Group (AIAG), 2017. *Measurement Systems Analysis.* 4 ed. s.l.:Automotive Industry Action Group.

Bass, I., 2007. *Six Sigma Statistics with Excel and Minitab.* s.l.:McGraw-Hill.

Bergquist, B. & Albing, M., 2006. Statistical Methods – Does Anyone Really Use Them?. *Total Quality Management,* 17(8), pp. 961-972.

Deleryd, M., 1996. *Process Capability Studies in Theory and Practice. Licentiate thesis.* Luleå: Luleå University of Technology. Department of Business Administration and Social Sciences Division of Quality Technology and Statistics.

Deleryd, M., 1998. *Enhancing the Industrial Use of Process Capability Studies. Doctoral thesis.* Luleå: Luleå University of Technology. Department of Business Administration and Social Sciences Division of Quality Technology and Statistics.

Johnson, R. A., 2018. *Miller & Freund's Probability and Statistics for Engineers.* 9 ed. s.l.:Pearson Education.

Kane, V., 1986. Process Capability Indices. *Journal of Quality Technology,* 18(1), pp. 41-52.

Levinson, W., 2011. *Statistical Process Control for Real-World Applications.* Boca Raton: CRC Press.

Minitab LLC., 2017. *2-Sample t-Test.* [Online]
Available at: https://support.minitab.com/en-us/minitab/18/Assistant_Two_Sample_t.pdf
[Accessed 24 9 2019].

Minitab LLC., 2019. *Test for 2-Sample Equivalence Test.* [Online]
Available at: https://support.minitab.com/en-us/minitab/18/help-and-how-to/statistics/equivalence-tests/how-to/2-sample-equivalence-test/interpret-the-results/all-statistics-and-graphs/test/
[Accessed 24 9 2019].

Minitab LLC., 2019. *Unbiasing constant c4'().* [Online]
Available at: https://support.minitab.com/en-us/minitab/18/quality-and-process-improvement/control-charts/how-to/time-weighted-charts/ewma-chart/methods-and-formulas/unbiasing-constant-c4/
[Accessed 16 9 2019].

Minitab LLC., 2019. *Unbiasing constants d2(), d3(), and d4().* [Online]
Available at: https://support.minitab.com/en-us/minitab/18/quality-and-process-improvement/control-charts/how-to/time-weighted-charts/ewma-chart/methods-and-formulas/unbiasing-constants-d2-d3-and-d4/
[Accessed 16 9 2019].

Mitutoyo Scandinavia AB Finnish Branch, n.d. *Metrologinen hakuteos.* s.l.:s.n.

Montgomery, D. & Runger, G., 2014. *Applied Statistics and Probability for Engineers.* 6 ed. Hoboken, NJ: John Wiley & Sons, Inc..

Oakland, J., 2003. *Statistical Process Control.* 5 ed. Oxford: Butterworth-Heinemann.

Pitt, H., 1994. *SPC for the Rest of Us - A Personal Path to Statistical Process Control.* Reading, MA: Addison-Wesley Publishing Company.

Qiu, P., 2014. *Introduction to Statistical Process Control.* Boca Raton: CRC Press.

Salomäki, R., 2003. *Suorituskykyiset prosessit - Hyödynnä SPC.* Tampere: Metalliteollisuuden keskusliitto MET.

Steiner, S., Abraham, B. & MacKay, J., 1997. *Understanding Process Capability Indices,* Institute for Improvement in Quality and Productivity research report. Waterloo, Ontario: University of Waterloo, Department of Statistics and Actuarial Science.

Triola, M., 2014. *Elementary Statistics Using Excel.* 5 ed. s.l.:Pearson Education.

Wikipedia, 2019. *Standard deviation.* [Online]
Available at: https://en.wikipedia.org/wiki/Standard_deviation
[Accessed 10 9 2019].