https://doi.org/10.26219/heal.aueb.6231

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
DEPARTMENT OF STATISTICS

Measurement System Analysis and Gauge Repeatability and Reproducibility Studies

By
Sofia Mylonakou

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfillment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece
September 2024

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

Ανάλυση Μετρητικών Συστημάτων και Μελέτες Επαναληπτικότητας και Αναπαραγωγιμότητας Εκτιμητών

Σοφία Μυλωνάκου

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Διπλώματος Μεταπτυχιακών Σπουδών στη Στατιστική

Αθήνα

Σεπτέμβριος 2024

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

To my brother

https://doi.org/10.26219/heal.aueb.6231

“ In any program of control we must start with observed data; yet data may be either good, bad or indifferent. Of what value is the theory of control if the observed data going in to that theory are bad? ”

Shewhart (1931)

https://doi.org/10.26219/heal.aueb.6231

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Prof. Stylianos Psarakis, both for the valuable lectures and for his guidance during the writing of this thesis. His helpful and constructive advice and remarks were decisive for its completion.

I am also highly obliged to my family for their constant encouragement, moral and financial support throughout my MSc degree in Statistics at the Athens University of Economics and Business (AUEB).

Lastly, I owe a special thank you to all my friends and everyone individually for their love, understanding, and support that helped me make this work a success.

https://doi.org/10.26219/heal.aueb.6231

ABSTRACT

Sofia Mylonakou

Measurement System Analysis and Gauge Repeatability and Reproducibility Studies

September 2024

Gauge R&R (Repeatability and Reproducibility) studies are a key part of Measurement System Analysis and they have gained notoriety in process improvement projects in the manufacturing sectors due to their essential applicability. They comprise a collection of statistical tools suitable for assessing the amount of variation in a measurement system that can be attributed to the instrument, the operators, and the interaction between them.

The main usage of Gauge R&R studies is to deal with the measurement system's odd error in order to provide the best possible and unbiased outputs. This enables practitioners in correct decision-making and consequently provides a profitable consulting to the corresponding company.

In conjunction with MSA, this thesis is referred also to the primary capability indices like C_p, C_{pk}, AC_p, and OC_p that are used to evaluate how capable a process is to generate products that meet the company's predetermined requirements.

A sophisticated strategy for achieving the best possible results is to combine these methodologies and approach this kind of projects from a more integrated perspective. Through this way we guarantee that the measurement system works properly for its intended purpose and the manufacturing process can adequately fulfill the demands for product quality.

Finally, the last part is dedicated to a computational experiment. We analyze based on the theory the performance of a hypothetical measurement system and we make inferences about the overall production process.

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

ΠΕΡΙΛΗΨΗ

Σοφία Μυλωνάκου

Ανάλυση Μετρητικών Συστημάτων και Μελέτες Επαναληπτικότητας και Αναπαραγωγιμότητας Εκτιμητών

Σεπτέμβριος 2024

Οι μελέτες επαναληψιμότητας και αναπαραγωγιμότητας (Ε&Α) των μετρητών/ εκτιμητών αποτελούν το βασικό μέρος της ανάλυσης συστημάτων μέτρησης και έχουν αποκτήσει μεγάλη δημοτικότητα κυρίως στα προγράμματα βελτίωσης διαδικασιών στον κατασκευαστικό τομέα λόγω της ουσιαστικής εφαρμογής τους. Περιλαμβάνουν μια συλλογή στατιστικών εργαλείων κατάλληλων για την αξιολόγηση του ποσοστού διακύμανσης ενός σύστημα μέτρησης που μπορεί να αποδοθεί στο όργανο, στους χειριστές και στην αλληλεπίδραση μεταξύ τους.

Η κύρια χρήση της μελέτης Ε&Α των μετρητών είναι η αντιμετώπιση του περιττού σφάλματος του συστήματος μέτρησης ούτως ώστε να λαμβάνονται όσο το δυνατόν καλύτερα και αμερόληπτα αποτελέσματα. Κάτι τέτοιο παρέχει μία καθαρή εικόνα της παραγωγικής διαδικασίας ωθώντας έτσι τους επαγγελματίες σε σωστά συμπεράσματα και αποφάσεις. Ως εκ τούτου, παρέχονται ωφέλιμες και κερδοφόρες συμβουλές στην εκάστοτε εταιρεία.

Στα πλαίσια της ανάλυσης μετρητικών συστημάτων, η παρούσα διπλωματική εργασία αναφέρεται και σε κάποιους βασικούς δείκτες ικανότητας όπως το C_p, το C_{pk}, το AC_p, και το OC_p που χρησιμεύουν στην αξιολόγηση των δυνατοτήτων μίας παραγωγικής διαδικασίας και κατά πόσο αυτή πληροί τις προκαθορισμένες απαιτήσεις.

Μια πιο εξελιγμένη στρατηγική για την επίτευξη των καλύτερων δυνατών αποτελεσμάτων, είναι ο συνδυασμός αυτών των μεθόδων που θα οδηγήσει σε μία πιο ολοκληρωμένη προσέγγιση του εν λόγω προβλήματος. Με αυτόν τον τρόπο εγγυόμαστε ότι το σύστημα μέτρησης λειτουργεί σωστά και παράλληλα ότι η διαδικασία κατασκευής μπορεί να ικανοποιήσει επαρκώς τις απαιτήσεις για την ποιότητα του προϊόντος.

Το τελευταίο μέρος της εργασίας αποτελείται από την υπολογιστική εφαρμογή ενός πειράματος. Αναλύουμε διεξοδικά την απόδοση ενός υποθετικού συστήματος μέτρησης εφαρμόζοντας το θεωρητικό κομμάτι που παραθέσαμε στα προηγούμενα κεφάλαια, εξάγοντας συμπεράσματα για τη συνολική παραγωγική διαδικασία.

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

Contents

Chapter 1: Introductory Theory ………….....……………….……………………..... 1

1.1 Introduction to Statistical Process Control (SPC) ……………….……..…. 1
1.2 Design of Experiment in Measurement System Analysis ……….……….... 4
1.3 Introduction to Gauge Capability Analysis …………………………….….... 6
1.3.1 Measurement System Capability Studies ………………........… 6
1.3.2 Gauge R&R Studies …………………………………..……..... 12
1.3.3 Measurement Capability Metrics ………………………......… 14
1.3.4 Misclassification Rates: False Defectives and Passed Defectives ………………………………………………........ 19
1.3.5 Analysis of Variance (ANOVA) Introduction Theory ……..... 24
1.3.6 ANOVA theory in Gauge R&R studies ……………….……….... 26

Chapter 2: Relations Between Process and Measurement System Capability Indices ……………………………………....………………………… 32

2.1 Introduction of the Process Capability indices C_p and C_{pk} ……..…...… 33
2.1.1 Relation Between P/T ratio and $SNR/\ DR$ ………………..…..... 35
2.1.2 Relation Between P/T ratio, SNR/ DR and $C_p/\ C_{pk}$ ..…..…..... 38
2.2 Relating Correlation ρ with other Quality Metrics …………………....…. 43
2.2.1 Correlation in Product Measurements ……………………....…. 43
2.2.2 Relation of Correlation ρ with Precision-to-Tolerance ratio *P/T* ………………………………………………....…….... 45
2.2.3 Relation of Correlation ρ with Precision-to-Tolerance ratio *P/T* and Process Capability C_p……………..…….…..... 47

Chapter 3: Relations Between the Actual AC_p and the Observed OC_p Process capability indices ………………..……………….……………...…..… 50

3.1 Critical value of Process Capability C_p index ………………………….... 50
3.2 The Actual AC_p and the Observed OC_p Process Capability ………...…. 51
3.2.1 The AC_p, OC_p and χ^2 Distribution …………………………… 53

Chapter 4: Computational Part ……………..……………………….....……. 59

4.1 Experimental Set-Up …………………………………………………..... 59

https://doi.org/10.26219/heal.aueb.6231

4.2 Statistical Analysis 61

4.2.1 Data Visualization 61

4.2.2 ANOVA Table Results 63

4.2.3 Further Analysis and Graph Illustration via "SixSigma" package 64

4.3 Process Capability Analysis Results 69

4.4 Statistical Analysis Conclusions 73

Chapter 5: Conclusions on Measurement System Analysis and Gauge R&R Studies and Further Research 75

Appendix 79

References 84

https://doi.org/10.26219/heal.aueb.6231

LIST OF CHARTS

Chart 1: DMAIC steps 3

Chart 2: Components of measurement system variability (Dev, Atul, Pankaj Kumar Jha, and Ved Parkash (2018)) 11

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

LIST OF TABLES

Table 1: Methods applicable in the automotive industry for evaluating gauge capability (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010). 11

Table 2: Gauge R&R parameters that describe the variation in the supervised procedure and the variation in the measurement system (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)). 15

Table 3: Table of data detection (one factor, α levels, n observations) (Zanobini, Andrea, et al. (2016)). 25

Table 4: Two-way ANOVA table for a standard model of a Gauge R&R experiment. 29

Table 5: 100(1 - α)% MLS Confidence Intervals for the Standard Gauge R & R Experiment. (Montgomery, Douglas C. (2019)) 30

Table 6: Definition of terms used in Table 5 (Montgomery, Douglas C. (2019)). 30

Table 7: The regions obtained from relating *P/T*, *SNR*/ *DR* and C_p/ C_{pk} capability indices (Abbas Al-Refaie and Nour Bata (2010)). 42

Table 8: Two-Way ANOVA table 63

Table 9: AIAG MSA manual acceptance guidelines. 64

Table 10: Contribution of the variance components. 68

Table 11: Study Variance of each component. 69

Table 12: Capability indices with confidence intervals. 70

Table 13: Capability indices with confidence intervals by appraiser. 71

https://doi.org/10.26219/heal.aueb.6231

LIST OF FIGURES

Figure 1: Upper and lower natural tolerance limits for a normal distribution (Montgomery, Douglas C. (2009)) 4

Figure 2: Repeatability (Afrooz Moatari Kazerouni (2009)) 12

Figure 3: Reproducibility (Afrooz Moatari Kazerouni (2009)) 13

Figure 4: Total Process Variability (Dev, Atul, Pankaj Kumar Jha, and Ved Parkash (2018)) 13

Figure 5: Precision-Accuracy concept : (a) high accuracy and high precision (b) high accuracy and low precision (c) low accuracy and high precision (d) low accuracy and low precision (Montgomery, Douglas C. (2009)). 19

Figure 6: Missed Fault (MF) and False Failure (FF) regions of a measurement system shown on a bivariate normal distribution contour. (D.C. Montgomery, (2009)) (Montgomery, Douglas C. (2009)). 22

Figure 7: Triangular representation of components variability ((Dalalah, Doraid, and Ali Diabat (2015)). 27

Figure 8: 36

Figure 8 (a): Regions of SNR

Figure 8 (b): Regions of P/T

Figure 8 (c): Combined regions of SNR and P/T

(Dalalah, Doraid (2023))

Figure 9: 38

Figure 9 (a): Regions of DR

Figure 9 (b): Regions of P/T

Figure 9 (c): Combined regions of DR and P/T

(Dalalah, Doraid (2023))

Figure 10: 39

Figure 10 (a): Regions of P/T, SNR and Cp criteria

Figure 10 (b): Regions of P/T, SNR and Cpk criteria

(Al-Refaie, Abbas, and Nour Bata (2010))

Figure 11: 41

Figure 11 (a): Regions of P/T, DR and Cp criteria

https://doi.org/10.26219/heal.aueb.6231

Figure 11 (b): Regions of P/T, DR and Cpk criteria
(Al-Refaie, Abbas, and Nour Bata (2010))
Figure 12: Comparison of P/T and correlation ρ gauge approval criteria. 46
(Majeske, Karl D., and Richard W. Andrews (2002))
Figure 13: Comparison of P/T, correlation ρ and capability index Cp gauge approval criteria (Majeske, Karl D., and Richard W. Andrews (2002)). 48
Figure 14: Distributions of estimated process capability indices $\widehat{AC}_p$ and $\widehat{OC}_p$. (Dalalah, Doraid (2023)) 53
Figure 15: The distribution of the actual ACp and observed OCp process capability indices. (Dalalah, Doraid (2023)) 54
Figure 16: The actual and the observed distributions for different values of SNR for a fixed sample size n and a reference value R. (Dalalah, Doraid (2023)) 55
Figure 17: The risk values of α and β in the χ^2distribution (Dalalah, Doraid (2023)). 57
Figure 18: Histogram of coil springs measurements with the empirical density and normal distribution curves. 61
Figure 19: Box plots grouped by operator and part. 62
Figure 20: Spring measurements box plots grouped by operator. 63
Figure 21: Bar plot of components variation at % percentage scale. 65
Figure 22: Range chart for each of the 3 appraisers. 66
Figure 23: Measures for each coil spring linked by their mean value. 66
Figure 24: Measures obtained by every appraiser linked by their mean value. 67
Figure 25: Line plot for the interaction effect between the coil spring and the appraiser. 68
Figure 26: Process capability analysis results of the whole measurement system. 71
Figure 27: Process capability analysis results from the 1st appraiser. 72
Figure 28: Process capability analysis results from the 2nd appraiser. 72
Figure 29: Process capability analysis results from the 3rd appraiser. 73

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

CHAPTER 1: Introductory Theory

Nowadays data has an unremitting huge impact in our lives. Both in industry and business we are using data more than ever before. Manufacturing companies collect and process a huge amount of information through measurements and inspections. Therefore, intensifying competition in the market requires cotemporary and effective methods to improve quality production and consequently increase customers' loyalty.

In order to cover these needs, the field of Statistical Quality Control (SQC) has been developed. More specifically, Walter A. Shewhart was the pioneered who established this theory in the early 1920s. It is a set of statistical principals and techniques applied at every stage of manufacturing production to monitor, analyze and control. SQC aims to verify that each product meets or exceeds consumer's requirements. It is really important to emphasize that the efficiency of any decision depends firstly on the data quality and secondly on a reliable measurement system.

1.1 Introduction to Statistical Process Control (SPC)

This chapter is an introduction to Statistical Process Control (SPC) and the philosophy behind the quality optimization. SPC is a methodology applied in manufacturing lines for monitoring and controlling the production process and verifies that the output satisfies the specification standards. Several statistical techniques are used to understand the process fluctuation in order to reduce variability and consequently improve the product quality. In other words, SPC provides data-driven insights that can help company to make decisions about process improvements, resource allocation and strategic planning.

The essential definition of quality relates to the desirable characteristics that a good, service, or procedure should have in order to satisfy customers' explicit and implicit needs and, by extension, to claim a competitive position in the market. In industrial context, quality characteristics refer to the attributes or features of a product, service, or process that determine its level of excellence and superiority.

According to Montgomery (2009), Statistical Process Control (SPC) is a collection of powerful problem-solving tools of which the most basic is Six-Sigma procedure and is about a general strategy we use to approach any project. The implementation of Six-Sigma includes the principal methodology of DMAIC, which is the acronym of the words: Define, Measure, Analyze, Improve and Control and correspond to the five-step sequence for implementing quality optimization. Below we give a brief description of each step:

https://doi.org/10.26219/heal.aueb.6231

1. **Define**: In this step the project goals are plainly defined. We focus on pinpointing the problem and potential opportunities of enhancement, clarifying the boundaries of the project and establishing the appropriate metrics to measure the success level. The main target is to identify and outline fundamental details like the scope, timeline, objectives and project's required resources.
2. **Measure**: Once we have settled on the problem, the next step is to measure the current outputs before any improvements are made and inspect the procedure or the system. We collect data on key process metrics and a starting point of performance is instated. Also, we check the validity of the measurement system. Due to this phase we understand the range of the problem and we define a reference point for comparison throughout the project.
3. **Analyze**: Here collected data are analyzed in order to distinguish the root causes of the problem that leads to procedure's inefficiencies. For this purpose various tools like charts and statistical analysis are used to identify abnormalities and potential sources of fluctuation. It is vital to understand through SPC analysis deeply the factors that influence the process in order to deal with them.
4. **Improve**: Here we concentrate on translating analysis findings into actionable solutions and implement them to improve the process performance. This usually involves generating ideas and applying of potential settlements in order to eliminate or moderate the root causes of the problem. Sometimes, in more complicated cases, when the procedure is influenced by multiple factors, Design of Experiment (DOE) is used to find the major influential factors and determine the optimal conditions under which the expected outcomes will be achieved.
5. **Control**: This is the final stage of DMAIC methodology and here the main target is to consolidate in long-term an ongoing improvement. So, control measures and system monitoring are applied to ensure that the process remains stable and the products satisfy the requirements. Moreover, it is important to develop out-of-control plans in case of emergency.

It is worth noting that in Measure step of DMAIC method, checks are carried out in the measurement system as well in order to ensure the credibility. In particular, measuring is a vital part of Six-Sigma methodology. Techniques like Measurement System Analysis (MSA) are applied extensively to evaluate its stability, precision and accuracy and verify that the collected data are valid and genuine as much as possible. More details on this field will be given in subsequent chapters.

Practitioners apply consecutively statistical techniques to manage all the product cycle stages and ensure that the output meets consistently the specification limits. Through pre-manufacturing development they quantify and analyze process fluctuation against product predefined requirements. Control charts, histograms/

https://doi.org/10.26219/heal.aueb.6231

probability plots and designed experiments are primarily used for this purpose. The eventual goal is to eliminate or greatly reduce variability. This general activity is called Process Capability Analysis (PCA) and is mainly carried out in Analyze phase of DMAIC.

The flow chart presents in summary the main goals in every phase of DMAIC:

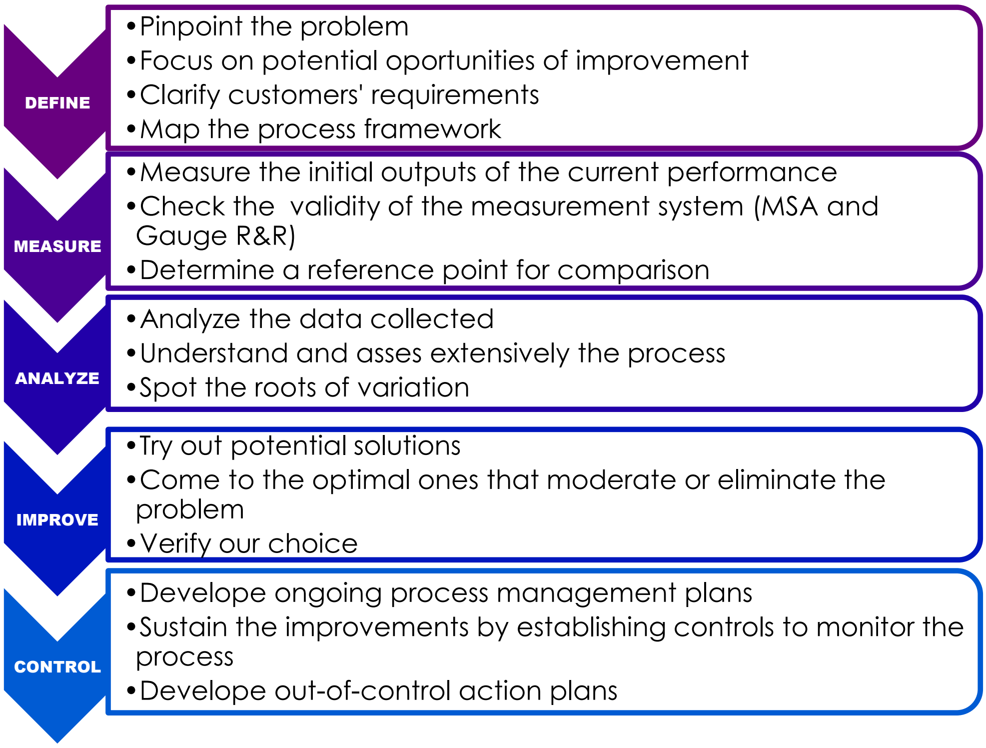

Chart 1: DMAIC steps

It is generally accepted to use the Six-sigma spread in the product quality characteristic distribution as a proxy for process capability. If we consider a normal distribution with mean μ and standard deviation σ, then upper and lower **natural tolerance limits** as illustrated in Figure 1 fall at $\mu + 3\sigma$ and $\mu - 3\sigma$ respectively (Montgomery, Douglas C. (2009)), i.e.

$$UNTL = \mu + 3\sigma$$

$$LNTL = \mu - 3\sigma$$

Considering a normal distribution, 97.73% of the variable is included in the natural tolerance limits, or put another way, only 0.27% of the process output will fall outside of the natural tolerance limits. The 0.27% of the variable that exceeds these limits is actually a remarkable quantity because it corresponds to 2700 non-conforming parts per million. Also, if the distribution of the process output is not normal, then the percentage of output falling outside $\mu \pm 3\sigma$ may vary significantly from 0.27%.

https://doi.org/10.26219/heal.aueb.6231

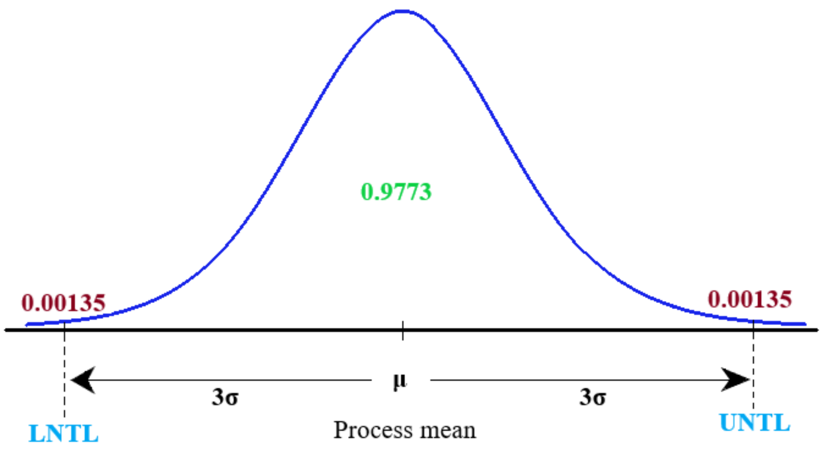

Figure 1: Upper and lower natural tolerance limits for a normal distribution.

In Process Capability Analysis (PCA) and Measurement System Analysis (MSA) these bounds are commonly referred as upper and lower specification limits, USL and LSL respectively, due to the predefined specifications/ requirements given from the practitioners.

1.2 Design of Experiment in Measurement System Analysis (MSA)

Design of Experiments (DOE or DOX) is a thorough and rigorous systematic approach for planning, conducting, analyzing and interpreting control experiments in order to assess the factors (also known as independent variables or inputs) that may influence a specific output (also known as response or dependent variable). Generally, this strategy helps analysts to recognize cause-and-effect interactions through meticulous exploration in order to apply the appropriate settings and achieve the best possible results. Moreover, DOE is used to pinpoint and control the sources of variability and lead the process to a steadier condition (Montgomery, Douglas C. (2019), Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)).

In addition, DOE is integral to MSA. It is crucial to guarantee the precision and accuracy of measurement system which in turn will ensure us credible outputs to analyze. Below we list the reasons in more detail:

- **Quantifying Measurement System Variability:** One of the core uses of DOE is to identify and quantify the sources of variability during the production process and then divide it into its corresponding components. Interventions and changes at different fields such as instruments, parts and operators help us to monitor and consequently understand better the production environment. In this way, we isolate the percentage of variation

 https://doi.org/10.26219/heal.aueb.6231

which is due to the measurement system thus evaluating its precision and accuracy.

- **Repeatability and Reproducibility Evaluation:** Another fundamental application of DOE in MSA is Gauge Repeatability and Reproducibility field. DOE comprises tools and strategies capable to evaluate the Repeatability (the variability that results when the same operator measures the same part multiple times under identical conditions) and the Reproducibility (the variability that results when the same part is measured from different operators or under changing conditions). Thus, we ensure that fluctuations are restricted as much as possible and the measurement system provides credible and consistent outputs across different operators and conditions.
- **Interaction Effects Inspection:** Through DOE we can identify and asses the interaction effects within the measurement system. For example, it is common different operator/ measuring instrument gives significant different results. More specifically, a certain combination can lead to higher variability levels and this is something that the practitioners should be aware of.
- **Optimizing of Measurement Conditions:** By studying DOE resulting feedback we are able to determine the optimal conditions for the measurement system, such as ideal environmental conditions (i.e. temperature or humidity) and best possible settings (i.e. suitable calibration) and effective training programs for the operator. This has as a consequence a more dependable and stabilized manufacturing process.
- **Measurement System Capability Improvement:** A well-conducted design experiment is capable through investigation to spot the weak points and limitations and propose at the same time the best possible solutions. Since we have distinguished the dominant factors that influence the measurement variation, we can implement targeted actions to enforce the system's capability. Thus, we validate that it meets the required archetypes of accuracy and precision.
- **Identifying Linearity and Bias:** Moreover DOE can be utilized to asses linearity (potential fluctuations on perceived accuracy or/ and precision experienced over the entire range of measurements made by the system) and bias (systematic error of the measurement). It assists in identifying any systematic deviation that requires rectifying inspection and helps to determine whether the measurement system produces consistent readings across the entire range by testing it at several known reference values.
- **Providing Statistical Rigor:** In MSA context, DOE provides statistical rigor during the whole analysis procedure. Statisticians rely on controlled experiments and systematic data analysis in order to draw conclusions based on solid and reliable statistical evidences and decreasing simultaneously the probability of erroneous decisions. So in this way, they guarantee robust validation for the measurement system and by extension for the validity of the results.

https://doi.org/10.26219/heal.aueb.6231

Therefore, using Designed Experiments in Measurement System Analysis is a requisite action to comprehend properly the measurement system and build a foundation to improve it. It provides an analytical structured approach to study and understand extensively the main factors that affect measurement system variability and decide if it is capable to produce consistent products in manufacturing line.

1.3 Introduction to Gauge Capability Analysis

Measurement is the cornerstone of decision-making. Apparently, a sufficient and necessary condition to have accurate and reliable results depends on the quality of the readings we will receive. Ideally, the measurement system would generate data that exactly represents the geometry of the part. In the manufacturing sector efforts are often directed for zero defect production by reduction of variability as much as possible.

If the data are contaminated with errors, this will surely lead to erroneous decisions that may have detrimental effect. For instance, when the measuring equipment is inadequate or ineffective, it can have a harmful impact on the linked process performance as well as on business and administrative decisions, which can cause a considerable economic loss to the related organization. Therefore, accurate and precise measured data is a recurring concern for contemporary industry.

1.3.1 Measurement System Capability Studies

The last years, within industrial context, many quality control strategies have been evolved for this purpose. **Measurement System Analysis** (MSA) is a set of statistical techniques that has been highly developed since is a fundamental tool of modern industrial production as it helps in making informed decisions. It constitutes a comprehensive set of means, like Gauge R&R study, that we use to monitor and evaluate the performance of a measurement system under stable conditions. The main target of MSA is to verify that the results taken from the equipment are accurate, steady and credible (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)).

The assurance and determination of measurement system capability is an important precondition before we move on to quality and process improvement actions. Since

https://doi.org/10.26219/heal.aueb.6231

every manufacturing process generates products with specific properties (features) and each property should be measured, there are two types of measurement methods able to validate quality and quantify performance (W.H. Woodall, C.M. Borror, (2008)). These are:

- measurement of procedure and
- measurement of product

Both of them demand credible approach. Thus, we should consider the actual influence of measurement errors throughout the entire production cycle to avoid making wrong decisions. As a matter of fact, inadequate measurement equipment causes difficulties in monitoring, controlling, optimizing and managing properly the procedure.

Measurement tools, also known as gauges or gages, are used to evaluate products' quality. It is worth noting that measurement systems, as far as the methodology is concerned, are conceptually far more sophisticated than the measuring equipment utilized for readings.

A measurement system typically comprises an instrument (gauge) as well as the operator(s) and the conditions or different time points under which the device is utilized, and the external environment in which the measure is obtained. Specifications, procedures, setting up and measuring techniques for the parts, fixtures and tooling for locating and orienting the object being measured, software for performing intermediate calculations and producing the result, and presumptions used to quantify a unit of measure or the entire process for obtaining measurements are also included (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)).

However, the main question that arises and must be answered responsibly is the following:

"Is the measurement system capable to distinguish between the bad and the good units?"

In general, the efficacy of the measurement system relies on both accurate gauges and proper gauge use. To ensure reliability in data gathering, the gauging tools and procedures need to be well-managed and appropriate for the tasks at hand (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)).

A perfect measurement system under ideal conditions always produces true values. Nevertheless, in actual reality, something like this is impossible, that is, resulted observations contain systematic and random errors. So, practitioners use the tools of measurement system capability analysis to minimize these errors and control the deviation.

For a company, the knowledge of variance is one of the most powerful "weapons" in the pursuit of improvement. There are two sources of variability that we should take into consideration when a process's output is measured:

- **part-to-part variability** that refers to the inherent variability among the separate parts being measured and
- **measurement system variability** that refers to the variability that results from the measurement device we use

In the second case, MSA techniques are applied in order to spot and discriminate the variation components and asses how much of the total observed variability is due to the gauge. So, given that all measurements contain error, we can signify the observed value as follows:

$$Observed\ value = True\ value + Measurement\ Error$$

The corresponding simple mathematical model for measurement system capability studies can be developed as follows:

$$y = X + \varepsilon \qquad (1.1)$$

where y stands for the total observed measurement, X refers to the actual true value of the product, and ε is the measurement error. We assume that X and ε are normally and independently distributed random variables with means μ and 0, and variances(σ_P^2) and (σ_{Gauge}^2) respectively, that is, $X \sim N(\mu, \sigma_P^2)$, $\varepsilon \sim N(0, \sigma_{Gauge}^2)$. Denote that the measurement error mean μ_ε is considered the measurement bias, which is often assumed to be 0.

Thus, the variance of the total observed measurement y is given by the formula below:

$$\sigma_{Total}^2 = \sigma_P^2 + \sigma_{Gauge}^2 \qquad (1.2)$$

where σ_{Total}^2 is the total variance of the observed measurements, σ_P^2 is the variance of the process, and σ_{Gauge}^2 is the variance of the measurement process. Statistical methods like control charts can be used to split σ_{Total}^2 to its individual parts, as well as to give an evaluation of gauge capability. Remark that the last component σ_{Gauge}^2 is usually broken up into two other components corresponding to Repeatability and Reproducibility, i.e. $\sigma_{Gauge}^2 = \sigma_{Repeatability}^2 + \sigma_{Reproducibility}^2$. As a result, the complete expression of the total variance is:

$$\sigma_{Total}^2 = \sigma_P^2 + \sigma_{Repeatability}^2 + \sigma_{Reproducibility}^2 \qquad (1.3)$$

More details on this field are provided in unit 1.3.2 of this chapter.

An alternative definition of MSA is the way of comprehending and controlling the gauging error. It is a crucial task that needs to be implemented prior to any

https://doi.org/10.26219/heal.aueb.6231

improvement actions, because measurement error distorts or masks procedure's actual capability (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010), Montgomery, Douglas C. (2019), Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery(2005)).

As we will see in the next chapter, one auditing tool that could be used to provide input for improving measuring techniques is the gauge R&R analysis. Repeatability and Reproducibility (R&R) outputs reveal how much of the variability in the production process is due to dispersion of the measurement device. There are several ways to calculate and assess an instrument's R&R index and the most common ones are presented below (Kazerouni, Afrooz Moatari (2009)):

1. **Range method:** It is used to determine a quick estimate of measurement system variation, but this method is not able to divide these variances into the two fundamental components: repeatability and reproducibility.
2. **Average and Range method:** This is a computational way to represent measurement system repeatability and reproducibility estimates. Unlike to range method, this one can segregate the two aforementioned components of variability.
3. **Average and Standard Deviation:** This one has similar characteristics as the Average and Range method.
4. **Analysis of Variance (ANOVA):** Repeatability and reproducibility can also be distinguished through this strategy. These divisions pertain to differences in operators and instruments. 'ANOVA' and 'Average and Range' are the most prevalent and important methods compared to the other two.

Since we have presented the main definitions of MSA, it is necessary to discuss about the components/ causes of the total measurement system variation. They are separated into four categories: **stability**, **bias**, **repeatability** and **reproducibility**. Each of these elements is isolated to be quantified, because only then is attainable to reduce the contribution of each one of these error components.

In general measurement error resources are divided in two categories: the regular errors and the accidental errors. Regular errors in an instrument can typically be measured by its constituent or by the calibration laboratory. Conversely, accidental errors arise due to variations among measurement devices, operators, temporary biases in instrument, environmental changes, and different adjustments. As a result, it would be required to analyze measuring capacity using definitions of accuracy and precision.

Accuracy is the ability of a gauge instrument to measure the actual value correctly in average. The measuring system will be more accurate and the likely average of distribution measurement will be closer to the true value the smaller the difference in measurement (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010), Montgomery, Douglas C. (2019)). Accuracy can be divided into three categories:

https://doi.org/10.26219/heal.aueb.6231

- **Stability** is defined as different levels of variability under different operating regimes, which is possibly due to uneven operator performance, warm-up effects, environmental factors, and inadequate operating procedures. Also, this statistical tool is able to quantify bias evolution over time. In other words, Stability Study evaluates measurement system condition across the timeline.
  Obviously, a measurement system after regular and repeated use is more likely to exhibit fluctuations due to natural wear and tear and, as a result, go out of stability. Hence, when the gauge needs to be calibrated, a stability test can be a very useful indicator.
- **Bias** is a difference between an observed measurement and the true/reference value obtained from a master or gold standard or from a different measurement technique that produces accurate values. Usually, Bias is ascribed to either an instrument error that adjusts (adds or subtracts) a constant value at each reading. This can be due to a worn out device or to appraiser's incorrect handling.
  From a statistical point of view, Bias can be defined as the statistically significant and systematic error of the measurement result from its true master value that is called the accuracy of a measurement.
- **Linearity** refers to the ability of a measurement system to generate readings that are exactly proportionate to the actual values over the whole measurement range. In essence, linearity assesses measurement system accuracy and consistency across various points on the measurement scale. The ideal linear system would show no deviation from the true values regardless of the position on the scale. In other words, linearity evaluates the difference between the measured and the actual value (bias) across the measurement range. Maintenance and calibration issues are the most common sources of linearity problems.

Precision is the variability that appears when using the same instrument to measure the same part. If the variance decreases, the precision of the measurement system may increase. Precision can be divided into two categories (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010), Montgomery, Douglas C. (2019)):

- **Repeatability** represents the basic inherent precision of the gauge itself where precision refers to the measure of the measurement system's inherent random variation. In a simpler form, it is the readings fluctuation due to instrument error.
- **Reproducibility** is the variability arising from external factors like operators and their unique techniques, setups and environmental changes over time. It is typically attributed to fluctuations among the operators who obtain various results.

Denote that the guidelines from specific automotive companies and MSA-based software developers contain similar methods for assessing measurement system.

 https://doi.org/10.26219/heal.aueb.6231

These methods can be put in particular order, as shown in Table 1. This arrangement shows how they are used practically, beginning with the initial inspection of a newly installed system (including its stability determination). The final step of the evaluation is about R&R study of the measurement system in the context of a real production environment.

Step	Examination	Applied Statistical Method
1	Stability	Range & Average Analysis
2	Bias	Range & Average Analysis
3	Linearity	Regression Analysis
4	Repeatability & Reproducibility	Range & Average Analysis and ANOVA analysis

Table 1: Methods applicable in the automotive industry for evaluating gauge capability.

In the subsequent Chart 2 we can see analytical the decomposition of the observed process variability into the aforementioned factors:

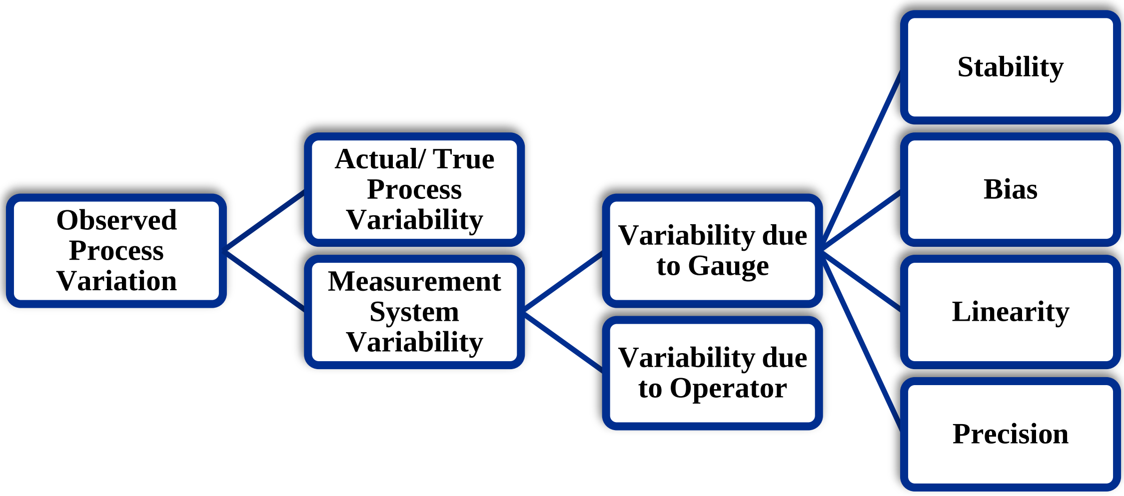

Chart 2: Components of measurement system variability.

To sum up, MSA is a modern quality assurance infrastructure used before data-based decision making, such as Statistical Process Control (SPC), Correlation and Regression Analysis, and Design of Experiment (DOE).The eventual goal is to inspect the measurement devices and verify that they give consistent and accurate readings, as well as being able to adequately discriminate between the good and the bad parts.

https://doi.org/10.26219/heal.aueb.6231

1.3.2 Gauge R&R Studies

Gauge Repeatability and Reproducibility (Gage R&R) is a particular methodology within the broader context of MSA that focuses on measurement system variation. It is used in industries such as pharmaceuticals, automotive, aerospace and manufacturing to guarantee the quality and reliability of their products. Analysts decently utilize it both as an audit tool and as a source of feedback to monitor and optimize the measurement system's performance.

As mentioned above, total process variation is a combination of part-to-part variation plus measurement system variation. The primary objective of Gauge R&R studies is to determine the amount of variability that is due to the measurement system. Repeatability and Reproducibility are the two fundamental components that asses measurement system's variation. Below we provide a comprehensive definition of these R's (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010), Montgomery, Douglas C. (2019), Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery (2005)):

Repeatability (or equipment's variation): The variability caused by the measurement tool/ gauge when it is used to measure the same part repeatedly or multiple times with the same operator or in the same time period. More specifically, it quantifies the precision under a controlled repeatable environment. Stated differently, it describes the deviation that is detected when a single gauge instrument is used several times by a single operator in order to measure the identical attribute on the same part. This can also be conceptualized as "within operator" error (one appraiser, one instrument). Denote that, to estimate sufficiently Repeatability, every operator measures every attribute at least twice.

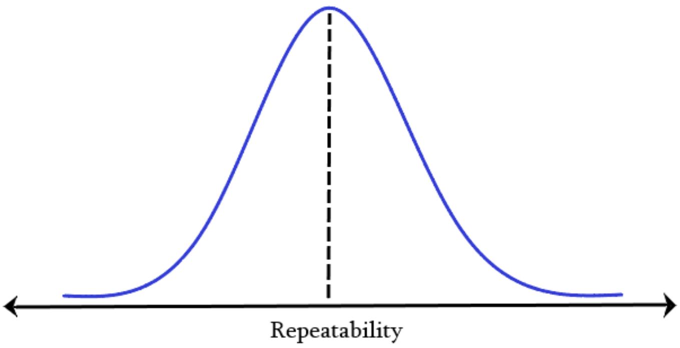

Figure 2: Repeatability

Reproducibility (or appraiser's variation): The variability arising through several operators, setups or time intervals. It quantifies measurement system's ability to produce consistent and credible results across different operators. More particular, it refers to the variation observed when the identical attribute of the same part is measured by various appraisers that utilize the same gauging tool. Another approach to think about is as "between operator" error (many appraisers, one instrument). To estimate sufficiently Reproducibility, the parts must be measured at least by two operators. Also, operators should measure the parts in random order, and the selected ones should represent the possible range of measurements.

In measurement system capability the combination of Repeatability and Reproducibility is known as Gauge R&R studies.

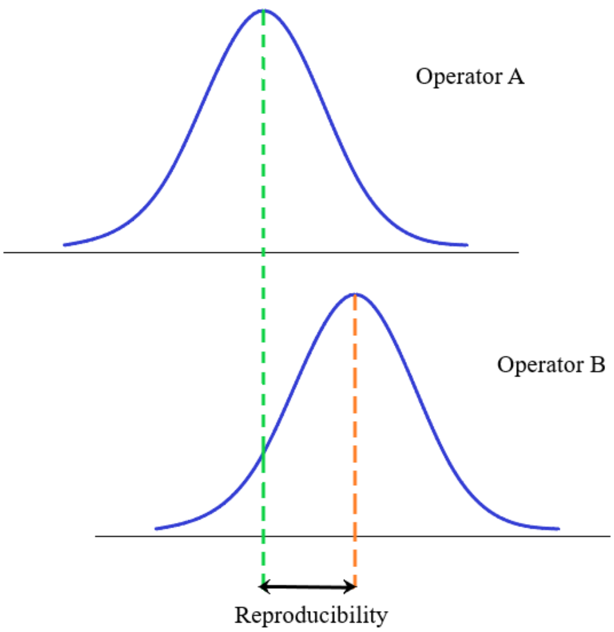

Figure 3: Reproducibility

Figure 2 and Figure 3 illustrate Repeatability and Reproducibility respectively.

The combined estimated variability from repeatability and reproducibility is known as the total Gauge R&R. Figure 4 illustrates a graphical representation of the total process variability in the production line under the assumption of a normal population.

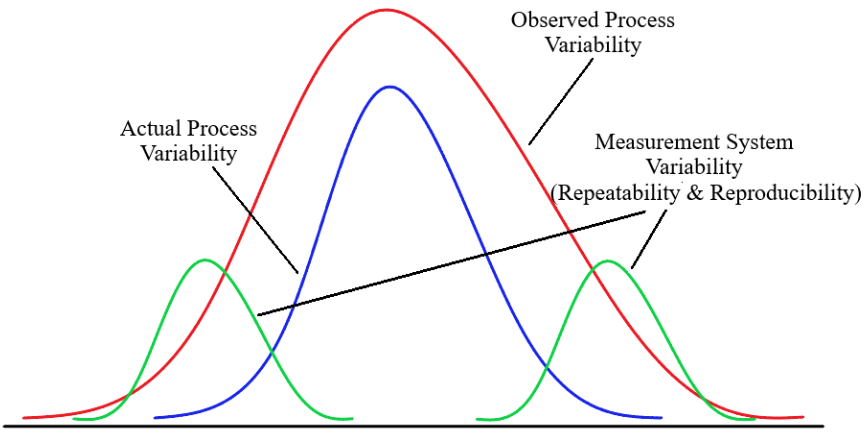

Figure 4: Total Process Variability

https://doi.org/10.26219/heal.aueb.6231

Note of that the observed process variation is always higher than the actual/ true, since we take into consideration the measurement system variation too (Dev, Atul, Pankaj Kumar Jha, and Ved Parkash (2018)).

When repeatability outweighs reproducibility, the gauge must be redesigned, and equipment must be properly maintained. Conversely, when reproducibility outweighs repeatability, indicates that the operator needs further training in the proper use and reading of gauge instrument (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010)).

The total variance of the measurement system is the sum of the overall Gauge R&R variability with part-to-part variability, where part is the variability between the separate pieces.

An insufficient measurement system renders it extremely challenging to monitor, regulate, improve, or handle a process correctly. As Douglas C. Montgomery (2009) has stated: It's somewhat analogous to navigating a ship through fog without radar—eventually you are going to hit the iceberg! As excessive measurement variation contributes to overall product variation, numerous other process improvement activities are adversely affected. This includes necessitating larger sample sizes in comparative or observational studies, requiring more replication in designed experiments aimed at enhancing processes, and demanding more extensive product testing.

Gauge R&R studies are primarily focused on estimating repeatability and reproducibility, the two fundamental components of measurement system variability, in order to determine whether the gauge is capable or not of its intended use.

1.3.3 Measurement Capability Metrics

As part of Gauge's R&R studies, there are several criteria we utilize to determine how capable a measurement system is. This field is known as Measurement Capability Metrics. This section focuses on capability metrics that compare measurement variation to total or partial variation. Below is listed a few of these methods:

1. R&R criterion
2. Precision-to-Tolerance ratio (P/T)
3. Ratio of process/ part variability to total variability (ρP)
4. Ratio of measurement system variability to total variability (ρM)
5. Signal-to-noise ratio (SNR)
6. Number of Distinct Categories (ndc)

 https://doi.org/10.26219/heal.aueb.6231

7. Discrimination ratio *(DR)*

Prior to delving into these methods, it is imperative to acknowledge the parameters as they characterize the variability in both the supervised procedure and the measurement system within a Gauge R&R study. These parameters are presented in Table 2.

Parameter	Definition of variation
γP	Variance of the monitored process (often referred as part-to-part variability)
γM	Variance of the measurement system (repeatability and reproducibility or gauge variably)
$\gamma T = \gamma P + \gamma M$	Total variance of the response variable
$\rho P = \gamma P/\gamma T$	Proportion of Total variance due to the process
$\rho M = \gamma M/\gamma T$	Proportion of Total variance due to the measurement system
$\gamma = \rho P/\rho M$	Ratio of process variance to measurement system variance

Table 2: Gauge R&R parameters that describe the variation in the supervised procedure and the variation in the measurement system.

% R&R criterion

Gauge R&R analysis provides us the estimated variance components. One basic criterion that determines measurement system's acceptability is $\sqrt{\hat{\rho} M}$ or $\sqrt{\gamma M/\gamma T}$, commonly referred to as the Total Gauge R&R %Study Var, and it should be suitably small. It should be clarified that the title '%Study Var' refers to the per cent (%) study variation which involves ratios of standard deviations and not ratios of variances as the name % Study Var may lead the reader to believe.

According to AIAG[1] a value of $\sqrt{\hat{\rho} M} < 0.1$ indicates that the measurement system is considered to be acceptable. If $\sqrt{\hat{\rho} M}$ takes values that lies between **0.1** and **0.3** that makes the gauge possibly accepted depending on factors such as the importance of the application, the cost of the measurement device, and the cost of the repair. Values exceeding **0.3** are typically regarded as unacceptable and it is suggested that the measurement system be improved to the fullest extend possible.

[1]AIAG (Automotive Industry Action Group) is a non-for-profit association founded in 1982 by executives from well-known automotive manufacturers such as Ford, General Motors, and Chrysler. It provides guidelines and standards for assessing the quality and reliability of measurement systems used in manufacturing processes since it helps companies to evaluate manage their accuracy, precision and consistency.

https://doi.org/10.26219/heal.aueb.6231

Precision to Tolerance ratio (P/T)

Comparing the estimated gauge capability to the specifications' width/ tolerance band for the part that is being measured is a fairly typical (though not always wise) procedure. The Precision-to-Tolerance ratio ($\boldsymbol{P/T}$) applies when the quality characteristic has a two-sided design specification. Designating the lower standard limit as **LSL** and the higher specification limit as **USL**, we commonly define the Tolerance as:

$$\boldsymbol{TOL = USL - LSL}$$

The formula ($\boldsymbol{P/T}$)that serves this purpose is a function of γM:

$$\frac{\boldsymbol{P}}{\boldsymbol{T}} = \frac{\boldsymbol{k\sqrt{\gamma M}}}{\boldsymbol{USL - LSL}} \times \mathbf{100\%} = \frac{\boldsymbol{k\sqrt{\gamma M}}}{\boldsymbol{TOL}} \times \mathbf{100\%} \qquad (1.4)$$

where USL is the upper specification limit, USL is the lower specification limit, $\sqrt{\gamma M}$ is the gauge standard deviation and k is a predefined constant. Most popular choices for k are k=5.15 and k=6. The value of k=5.15 corresponds to the limiting value of the number of standard deviations between bounds of 95% tolerance interval that contains at least the 99% of a normal population, and k=6 corresponds to the number of standard deviations between the usual natural tolerance limits of a normal process. Remark that the precision to tolerance ratio P/T is used when the quality characteristic has two-sided design specification or tolerance.

As Montgomery (2009) has stated and AIAG's MSA Manual has confirmed, if Precision-to-Tolerance ratio (P/T) takes values of 0.1 or less, that indicates sufficient gauge capability. This is founded on the generally used rule that requires a measurement device to be calibrated in units one-tenth as large as the accuracy desired in the final measurement. On the other hand, if PTR takes values of 0.3 or greater, then the gauge is incapable with an indistinct margin left between these limits.

However, we should use caution in applying this general rule of thumb in all cases. P/T is not a reliable indicator of how well a measurement system operates, since a process with high capability is possible to accept a measurement system with higher P/T than a process with lower capability. A gauge needs to be able to measure precisely and accurately the products, and generate reliable data that will help the statistician to make the correct decisions. So, $P/T < 0.1$ may not always be a good reference value.

Because of this, we use other metrics to determine the measurement system's capability. The ratio of process (part) variability to total variability, that corresponds to the function $\boldsymbol{\rho P}$, is one of them. Another one is the function $\boldsymbol{\rho M}$, the ratio of

https://doi.org/10.26219/heal.aueb.6231

measurement system variability to total variability, which can also be expressed as $\boldsymbol{\rho M = 1 - \rho P}$.

Signal to Noise Ratio (SNR):

Another indicator that describes measurement system adequacy is a function of ρP that is called Signal to Noise Ratio and it has the following form:

$$\boldsymbol{SNR} = \sqrt{\frac{\boldsymbol{2\rho P}}{\boldsymbol{1 - \rho P}}} \quad (1.5)$$

where ρP is the proportion of total variance due to the process. According to AIAG, Signal to Noise ratio (SNR) is the number of distinct levels or categories that can be reliably obtained from the measurements. A value of 5 or greater is recommended and a value less than 2 indicates insufficient gauge capability.

Number of Distinct Categories (ndc):

An extra metric that we use in the context of MSA is the Number of Distinct Categories (*ndc*) in order to assess the precision and resolution of the gauge. It indicates how well a measurement system can differentiate between different levels of the characteristics being measured. The main formula given by AIAG is:

$$\boldsymbol{ndc = 1.41\sqrt{\hat{\gamma}}} \quad (1.6)$$

If ndc takes values greater than or equal to five ($ndc \geq 5$), that points out that the level of precision is usually adequate for most industrial applications. The number of ndc is truncated to give an integer. Obviously, in this form ndc is equivalent to SNR.

For one that prefers to use the ratio of variances, this is equivalent to the requirement that $\sqrt{\hat{\rho}M} < 27.14\%$ or, equivalently $\rho M < 7.4\%$. So, we can also calculate the value of ndc from the total gage R&R %Study Var using the following relationship:

$$ndc = 1.41\left(\frac{1}{\%Study\,Var} - 1\right)^{1/2} = 1.41\left(\frac{1}{\hat{\rho}M} - 1\right)^{1/2} \quad (1.7)$$

If $\sqrt{\hat{\rho}M} < 0.1$, then the value of ndc is at least 14. Such value is much more in live with the common 10-1 (or 'ten-bucket') rule of AIAG than a minimum ndc value of five. As we mentioned previously, this rule indicates that the measuring equipment should be able to discriminate one-tenth of the process variation (Woodall, William H., and Connie M. Borror (2008)).

Discrimination Ratio (D_R):

 https://doi.org/10.26219/heal.aueb.6231

Still, another indicator that evaluates gauge capability is the Discrimination Ratio. It is also a function of ρP and it can be expressed in the following way:

$$\boldsymbol{DR} = \frac{\mathbf{1} + \boldsymbol{\rho P}}{\mathbf{1} - \boldsymbol{\rho P}} \qquad (1.8)$$

Discrimination Ratio (D_R) is used to determine the number of categories that a measurement system is capable to distinguish. The number of data categories is often referred to as the discrimination ratio, because it specifies how many classifications can be consistently distinguished given the observed process variability. According to W. H. Woodall and C. M. Borror, it is safer to work on the measurement process when D_R exceeds 4 (preferably $DR \geq 5$). This suggests that the gauge capability is sufficient.

Since there is a close relationship between D_R and ndc, as Wheeler and Lyday stated, the last metric can also take the following form if we take into account 1.41 as an approximation of $\sqrt{2}$:

$$\boldsymbol{DR} = (\boldsymbol{ndc}^{\mathbf{2}} + \mathbf{1})^{\mathbf{1/2}} \qquad (1.9)$$

It must be remembered that none of these indicators accurately captures measurement system or gauge capability in a way that is directly interpretable. All the quantities mentioned above have been criticized for being overly subjective. In contrast to SNR, ndc and D_R measures, misclassification rates provide a more impartial assessment of the system's performance. Unlike the number of categories indicators, the misclassification rates rely on procedure's specification limits, which is something much more operatable.

Moreover, in this section the main focus was on gauge precision and not on gauge accuracy. The Figure 1 below illustrates the difference between these two concepts. In this figure the red mark in the center of the cycles/ targets is considered to be the true value of the measured characteristic, or μ the mean of X in equation 1 (according to assumption that $X \sim N(\mu, \sigma_P^2)$).

Accuracy refers to the ability of the instrument to measure correctly on average the true value. Precision has to do with measurement system's inherent variability. In scenario (a) of Figure 5 the gauge is accurate and precise, since the green dots are concentrated and fairly close to the red mark. In scenario (b) the gauge is accurate but not precise, because even though the dots are around the red mark, they are quite scattered. In scenario (c) the gauge is precise but not accurate, since even though the dots are close together, they are far away from the red mark. In the last scenario (d) the gauge is neither accurate nor precise, because as we can see the dots are quite scattered and far away from the red mark. Obviously scenario (a) represents the desirable gauge.

https://doi.org/10.26219/heal.aueb.6231

It is frequently necessary to utilize a standard in order to evaluate the accuracy of a measurement system, for which the actual value of the measured attribute is known.

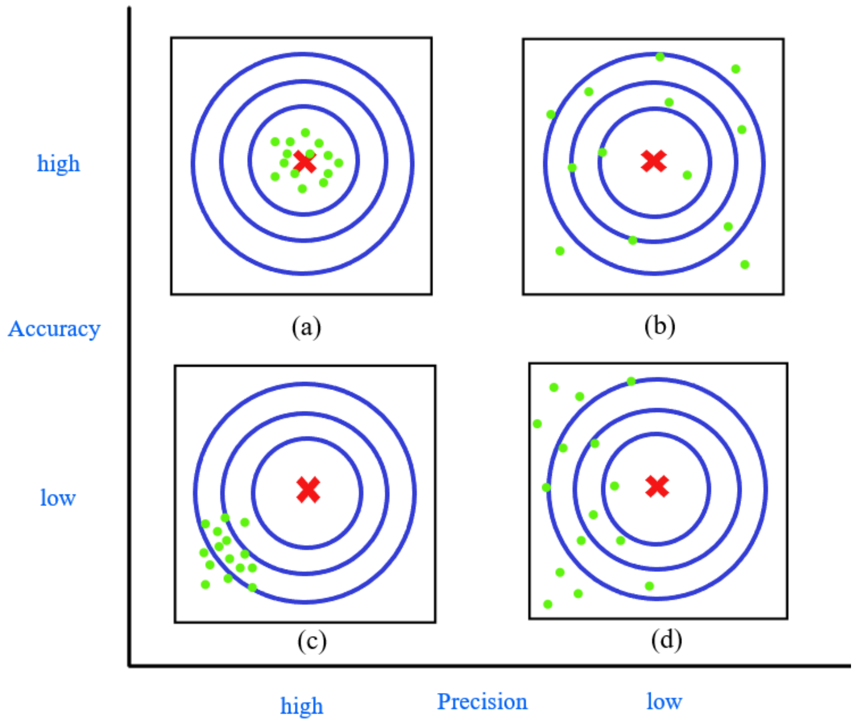

Figure 5: Precision-Accuracy concept : (a) high accuracy and high precision (b) high accuracy and low precision (c) low accuracy and high precision (d) low accuracy and low precision.

1.3.4 Misclassification Rates: False Defectives and Passed Defectives

The P/T ratio, the signal-to-noise ratio (SNR), number of distinct categories (ndc), the discrimination ratio (DR), and ρP and ρM are some of the metrics we've analyzed earlier that are used to condense the capabilities of the gauge instruments. However, none of these quantities can describe gauge capability in a directly interpretable form. The best way to characterize the efficacy and capability of a measurement system is to look at how well it can distinguish between the good and the bad parts. A more objective and reliable criterion we use to examine a gauge is the misclassification rate also known as misclassification probability. These probabilities offer trustworthy, practical and simple to understand information, because they are predicated on measurement process's actual performance.

Misclassification rates are explained by considering the model first proposed in equation (1):

https://doi.org/10.26219/heal.aueb.6231

$$y = x + \varepsilon$$

where x is the true value of the measurement, ε is the measurement error and y is the total observed measurement. We work under the assumption that x and ε are normally distributed and independent random variables with means μ_P and 0 and variances σ_P^2 and σ_{Gauge}^2 respectively, i.e. $x \sim N(\mu_P, \sigma_P^2)$ and $\varepsilon \sim N(0, \sigma_{Gauge}^2)$.Note that the mean of measurement error μ_M usually is setted equal to 0, since μ_M is consider to be the bias and we assume as a default that ε is unbiased (the measurement system is impartial). As a result y is also a normally distributed random variable with mean $\mu_y = \mu_P$and variance $\sigma_y^2 = \sigma_P^2 + \sigma_{Gauge}^2$.

The joint probability density function of y and x, say $f(x,y)$, is a bivariate normal with mean vector $[\mu_P, \mu_P]'$ and covariance matrix

$$\begin{bmatrix} \sigma_{Total}^2 & \sigma_P^2 \\ \sigma_P^2 & \sigma_P^2 \end{bmatrix}$$

Now we have two additional scenarios to take into consideration:

Scenario 1: A unit of product or part is in conformance to the specifications, if $LSL < x < USL$

Scenario 2: The measurement system will pass a product or part as non-defective, if $LSL < y < USL$

There are two types of possible misclassifications of a product or part:

1. If scenario 1 is true but scenario 2 is false, that means that even though a product is within specifications is not passed. In other words, the part has been incorrectly failed and this misclassification type is called false failure. This probability is also known as producer's risk.
2. If scenario 1 is false but scenario 2 is true, that means that even though a product is defective, i.e. it does not meet the specifications, it is considered good and it passes. In other words, the part has been incorrectly passed and this misclassification type is called missed fault. This probability is also known as consumer's risk.

The producer's risk δ is defined as the conditional probability that the measurement system will fail a part when the part conforms to the specifications. On the other hand, the consumer's risk β is defined as the conditional probability that the measurement system will pass a part when the part does not conform to the specifications. Notice that both types of misclassification rates, either false failure or missed fault, can be pretty costly for the company.

The expressions for calculating these two conditional probabilities are available below:

https://doi.org/10.26219/heal.aueb.6231

False failure probability/ producer's risk:

$$\delta = P\big((LSL < x < USL)\ and\ (y < LSL\ or\ y > USL)\big)$$

$$\delta = P((LSL < x < USL\ and\ y < LSL) or\ (LSL < x < USL\ and\ y > USL))$$

$$\delta = \int_{LSL}^{USL} \int_{-\infty}^{LSL} f(y,x)\,dy\,dx + \int_{LSL}^{USL} \int_{USL}^{+\infty} f(y,x)\,dy\,dx \qquad (1.10)$$

Missed fault probability/ consumer's risk:

$$\beta = P((x < LSL\ or\ x > USL)\ and\ (LSL < y < USL))$$

$$\beta = P((x < LSL\ and\ LSL < y < USL) or\ (x > USL\ and\ LSL < y < USL)$$

$$\beta = \int_{-\infty}^{LSL} \int_{LSL}^{USL} f(y,x)dydx + \int_{USL}^{+\infty} \int_{USL}^{LSL} f(y,x)dydx \qquad (1.11)$$

Mader, Prins and Lampe (1999) expressed the conditional probabilities as:

$$\delta_c = \frac{\delta}{\pi} \quad (1.12) \quad \text{and} \quad \beta_c = \frac{\beta}{1-\pi} \quad (1.13)$$

where

$$\pi = \int_{LSL}^{USL} f(x)\,dx \quad (1.14)$$

is the probability of a good part and $f(x)$ is the marginal probability density function of the random variable x that follows the normal distribution with mean μ_P and variance σ_P^2. Therefore, the completed equations are:

$$\delta_c = \frac{\int_{LSL}^{USL} \int_{-\infty}^{LSL} f(y,x)\,dy\,dx + \int_{LSL}^{USL} \int_{USL}^{+\infty} f(y,x)\,dy\,dx}{\int_{LSL}^{USL} f(x)\,dx} \qquad (1.15)$$

$$\beta_c = \frac{\int_{-\infty}^{LSL} \int_{LSL}^{USL} f(y,x)dydx + \int_{USL}^{+\infty} \int_{USL}^{LSL} f(y,x)dydx}{1 - \int_{LSL}^{USL} f(x)dx} \qquad (1.16)$$

Stated differently, δ_c is the conditional probability that a part fails given that is good, and β_c is the conditional probability that a part passes given the part is bad.

Figure 6 illustrates the regions false failures (FF) and missed faults (MF) on a density contour of a bivariate normal distribution. Thus, equations (1.10) and (1.11) can be used to compute the quantities δ and β for given values of μ_P, σ_P^2 and σ_{Total}^2.

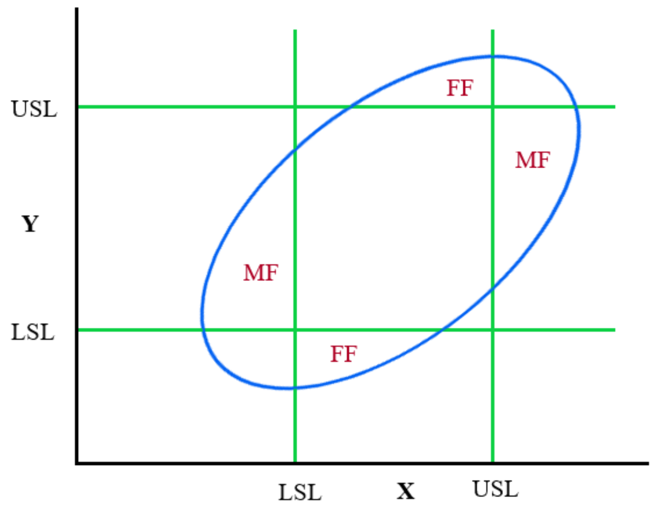

Figure 6: Missed Fault (MF) and False Failure (FF) regions of a measurement system shown on a bivariate normal distribution contour. [From D.C. Montgomery, Introduction to Statistical Quality Control 6th edition (2009)].

The measurement system is considered as incapable if δ, β or both are unacceptably large. Unfortunately, there are not standard acceptable levels to which these values can be compared, but only lower and upper values for a given level of confidence. However, R. K. Burdick proposed to compare the aforementioned misclassification rates to the ones that would be obtained by employing a chance process. Let's assume that the value π in equations (1.12) an (1.13) is known. A chance process for classifying parts is to randomly π of the parts as good and 1-π as bad, without taking any measurements. Hence, the false failure unconditional probability for this chance process is:

$$\delta^* = P((LSL < x < USL) \; and \; (Part \; classified \; as \; bad))$$

$$= P(LSL < x < USL) \times P(Part \; classified \; as \; bad)$$

$$= \pi(1-\pi) \qquad (1.17)$$

In the same way, the missed fault unconditional probability is:

$$\beta^* = P\big((x < LSL \; or \; x > USL) \; and \; (Part \; classified \; as \; good)\big)$$

$$= P(x < LSL \; or \; x > USL) \times P(Part \; classified \; as \; good)$$

$$= (1-\pi)\pi \qquad (1.18)$$

Therefore, in order for a measurement system to be considered acceptable, the misclassification rates δ and β should be less than the values $\delta^*$and β^*, that one could

https://doi.org/10.26219/heal.aueb.6231

get by chance. An equivalent argument can be made by using the conditional probabilities δ_c and β_c. In this instance, $\delta_c^* = 1 - \pi$ and $\beta_c^* = \pi$ are the conditional probabilities that could be obtained by chance.

Burdick provided confidence intervals for the ratio between the chance process's misclassification rates and the current process's misclassification rates. More particularly, confidence intervals are built on the ratios:

$$\delta_{index} = \frac{\delta}{\delta^*}$$

$$= \frac{\delta}{\pi(1-\pi)} \qquad (1.19)$$

and

$$\beta_{index} = \frac{\beta}{\beta^*}$$

$$= \frac{\beta}{(1-\pi)\pi} \qquad (1.20)$$

If the confidence interval encompasses the value of unity, this indicates that the current measurement system may not be any better than the one where the parts are classified by chance. As an illustration, if the confidence interval on $^{\delta}/_{\delta^*}$ contains the unity, then there is possibility for $^{\delta}/_{\delta^*}$ to be equal to 1, or equivalently $\delta = \delta^*$. So, the anticipated outcome is the entire confidence interval that covers a range of values less than 1.

In addition, let's remark that $^{\delta_c}/_{\delta_c^*} = \delta_{index}$ and $^{\beta_c}/_{\beta_c^*} = \beta_{index}$, which means that these indexes can be applied for both the unconditional and the conditional definitions of misclassification rates.

The misclassification rates in equations (1.19) and (1.20) can be used to compute the pivotal quantities δ and β for given values of μ_P, σ_P^2, σ_{Total}^2, LSL and USL, by integrating bivariate normal distributions as shown in equations (1.10) and (1.11) (Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery (2003)). But in practice we do not know the true values of μ_P, σ_P^2 and σ_{Total}^2 and we have to estimate them from the readings obtained in the R&R analysis. However, if point estimates are all that are used, the calculation ignores the uncertainty in the estimations. When calculating δ and β parameters it would be highly beneficial to provide confidence intervals as well. A common method is by computing δ and β under different scenarios suggested by the confidence intervals on the variance components (Senvar, Ozlem, and Seniye Umit Oktay Firat (2010), Montgomery, Douglas C. (2019),

https://doi.org/10.26219/heal.aueb.6231

Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery (2005)). For example, we can consider one pessimistic and one optimistic scenario:

- **Pessimistic scenario:** In this case, we assume the worst possible performance for the measurement system and the worst possible capability for the manufacturing process at the same time. To do this, we set σ_P^2 equal to the upper bound of the confidence interval for σ_P^2 and we solve for the value of σ_{Total}^2 that provides the lower bound on ρ_P. This is done by computing δ and β using upper bound on σ_P^2 and lower bound on ρ_P.
- **Optimistic scenario:** In this case, we assume the best possible performance for the measurement system and the best possible capability for the manufacturing process at the same time. Accordingly now we set σ_P^2 equal to the lower bound of the confidence interval for σ_P^2 and we solve for the value of σ_{Total}^2 that provides the upper bound on ρ_P.

Usually, confidence intervals tend to be quite wide due to the small number of operators that AIAG suggests. Researchers have demonstrated that in Gauge R&R study, more operators are needed than the commonly used three when the operator effect is deemed random.

Zappa and Dedlossi (2009) showed how to determine the number of components, operators and replications that control for the conditional probability errors as well as how to define the key levels of measurement capability indices by taking advantage of the misclassification rates.

1.3.5 Analysis of Variance (ANOVA) Introduction Theory

As we know, Analysis of Variance or ANOVA is a powerful and extensively used tool that consists of a series of techniques originating from the theory of inferential statistics that can be applied in order to evaluate and compare the variability of the data. Its use in Gauge R&R studies is crucial since it provides a particular and robust statistical analysis for ensuring the accuracy and reliability of measurement systems. Variability can be attributed to shifts, operators, equipment, gauges, parts, and chance (also known as residual effects) or even interactions between these factors. More specifically, ANOVA, as a technique of parametric statistical inference that relies on hypothesis tests, is used to estimate the differences between the sample means of two or more populations, by analyzing the corresponding variances. It evaluates whether or not such differences are random by comparing two or more distinct distributions.

Let's assume an experiment where the effects of factors are taken into consideration with different levels (or Treatments) for each factor. Then, a response variable would

https://doi.org/10.26219/heal.aueb.6231

be the result of this experiment. Analyzing each factor's impact as a source of variability requires taking into account 'α' levels and gathering 'n' random observations (replicated responses for each level. This leads to the extraction of a data collection for each factor, that can be represented by an $\alpha \times N$ dimension matrix as shown in Table 3.

Level	Observations				Total	Expected values
1	y_{11}	y_{12}	$\cdots$	y_{1n}	y_1	$\bar{y}_1$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	y_2	$\bar{y}_2$
…	…	…	…	…	…	…
α	$y_{\alpha 1}$	$y_{\alpha 2}$	$\cdots$	$y_{\alpha n}$	y_α	$\bar{y}_\alpha$

Table 3: Table of data detection (one factor, α levels, n observations)

In the context of ANOVA the total data variability is divided into two main components: within-group (or within) variability and between-group (or between) variability. **Within** variability refers to the variability within each group and measures how much individual observations within each group deviate from their respective group mean. **Between** variability refers to the variability due to the differences between the group means and measures how much the group means deviate from the overall mean (grand mean) (Zanobini, Andrea, et al. (2016)).

This framework allows for the consideration of different ANOVA types based on the number of factors included in the experiment. In particular:

- **One-way ANOVA** involves one independent variable/ factor with multiple levels and it determines if there are statistically significant differences in the means across the groups. Obviously, it does not account for interactions since there is only one factor.
- **Two-way ANOVA** involves two independent variables/ factors, each with multiple levels and it determines if there are statistically significant differences in the means for each factor independently. Moreover, evaluates the interaction effect between the two factors, revealing if the effect of one factor depends on the level of the other factor.

One-Way ANOVA is more appropriate for simpler experiments with one factor, while Two-Way ANOVA is used for more complex experiments involving two factors and their potential interaction effects.

1.3.6 ANOVA theory in Gauge R&R studies

 https://doi.org/10.26219/heal.aueb.6231

In the context of Gauge R&R studies, accuracy which is a parameter that indicates the measurement process variability, can be decomposed into **repeatability** (the inherent precision of the gauge itself) and **reproducibility** (variability due to different operators using the gauge, or different conditions in general). According to this method, the total variability can be portioned into:

$$\sigma^2_{Measurement\ Error} = \sigma^2_{Gauge} = \sigma^2_{Repeatability} + \sigma^2_{Reproducibility} \quad (1.21)$$

where $\sigma^2_{Repeatability}$ is the variability associated to repeatability and $\sigma^2_{Reproducibility}$ is the variability associated to reproducibility.

Subsequently, we consider α randomly selected parts and b randomly selected operators, and each operator measures every part n times. Denote that $i = part$, $j = operator$ and $k = measurement$. Then the measurements can be represented by the following model:

$$\boldsymbol{y_{ijk}} = \boldsymbol{\mu} + \boldsymbol{P_i} + \boldsymbol{O_j} + (\boldsymbol{PO})_{\boldsymbol{ij}} + \boldsymbol{\varepsilon_{ijk}} \quad (1.23)$$

$$\text{where} \quad \begin{cases} i = 1,2, \dots, p \\ j = 1,2, \dots, o \\ k = 1,2, \dots, n \end{cases}$$

where the model parameters P_i, O_j, $(PO)_{ij}$ and ε_{ijk} are all independent random variables. In particular, $\boldsymbol{\mu}$ is the general mean measurement (constant), $\boldsymbol{P_i}$ represents the effect of the i-th part, $\boldsymbol{O_j}$ represents the effect of the j-th operator, $(\boldsymbol{PO})_{\boldsymbol{ij}}$ refers to the interaction effect between the i-th part and the j-th operator, $\boldsymbol{\varepsilon_{ijk}}$ is the random error (measurement error or repeatability) and $\boldsymbol{y_{ijk}}$ is the observed measurement for k-th trial on the i-th part by the j-th operator (Montgomery, Douglas C. (2019), Zanobini, Andrea, et al. (2016)). Equation (1.23) outlines **a random effects model analysis of variance (ANOVA)** which is also known as standard model for a gauge R&R experiment and furthermore is a two-way ANOVA model. We hypothesize that the random variables P_i, O_j, $(PO)_{ij}$ and ε_{ijk} are normally distributed with null mean and variances given by $V(P_i) = \sigma^2_P$, $V(O_i) = \sigma^2_O$, $V\left[(PO)_{ij}\right] = \sigma^2_{PO}$ and $V\left(\varepsilon_{ijk}\right) = \sigma^2$. As a result, any observation's variance is:

$$V\left(y_{ijk}\right) = \sigma^2_P + \sigma^2_O + \sigma^2_{PO} + \sigma^2 \quad (1.24)$$

where σ^2_P, σ^2_O, σ^2_{PO} and σ^2 are the partial variance components that need to be estimated. For this purpose we use analysis of variance methods. The process entails dividing the overall measurement variability into the subsequent component parts:

$$SS_{Total} = SS_{Parts} + SS_{Operators} + SS_{Parts\times Operators} + SS_{Error} \quad (1.25)$$

where SS symbol refers to the sum of squares. Here, it is noteworthy to point out a triangular way to represent the variability components decomposition (Fig. 7) by utilizing the Pythagorean Theorem (Dalalah, Doraid, and Ali Diabat (2015)).

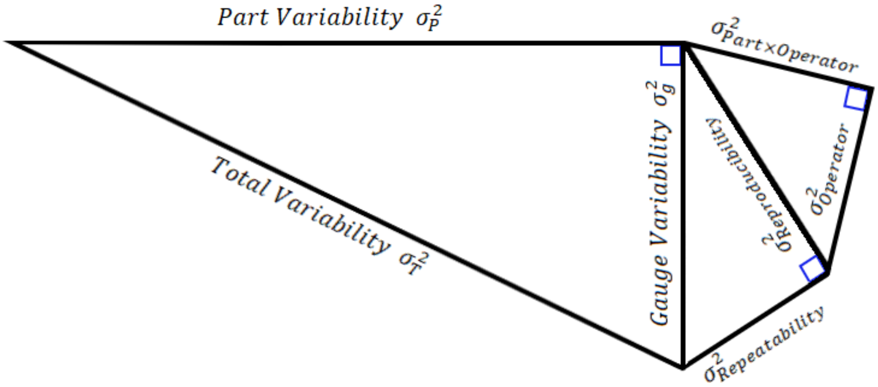

Figure 7: Triangular representation of components variability.

A computer software package is required for the computation of sum of squares. As a next step, mean of squares are produced by dividing each sum of squares component by the corresponding degrees of freedom (Montgomery, Douglas C. (2019), Zanobini, Andrea, et al. (2016)).

$$MS_P = \frac{SS_{Parts}}{p-1}$$

$$MS_O = \frac{SS_{Operators}}{o-1}$$

$$MS_{P\times O} = \frac{SS_{Parts\times Operators}}{(p-1)(o-1)}$$

$$MS_E = \frac{SS_{Error}}{po(n-1)}$$

We can show that the expected values of the mean squares are as follows:

$$E(MS_P) = bn\sigma_P^2 + n\sigma_{PO}^2 + \sigma^2$$

$$E(MS_O) = an\sigma_O^2 + n\sigma_{PO}^2 + \sigma^2$$

$$E(MS_{P\times O}) = n\sigma_{PO}^2 + \sigma^2$$

$$E(MS_E) = \sigma^2$$

Another important metric of ANOVA table is the F-statistic or F-ratio, a value that shows the ‘between-group variability’ compared to ‘within-group variability’. Generally, the larger this quantity is, the more significant ‘between-group variability’ would be. Below we represent the F-ratio for the parts, operators and their interaction respectively:

 https://doi.org/10.26219/heal.aueb.6231

- $F_P = \frac{MS_P}{MS_{P \times O}}$ which tests if we have statistically significant difference in the variability between parts and the variability resulting from the interaction between parts and operators.
- $F_O = \frac{MS_O}{MS_{P \times O}}$ which tests if we have statistically significant difference in the variability between operators and the variability resulting from the interaction between parts and operators.
- $F_{P \times O} = \frac{MS_{O \times P}}{MS_E}$ which tests if we have statistically significant difference in the variability between parts and operators and the variability resulting from the measurement error (repeatability). In essence, a high $F_{P \times O}$ indicates that the measurement variability depends on a specific combination of part and operator.

The most crucial part of an ANOVA table is the quantity of p-value p that indicates the statistical significance of the main effects of each factor. Statistical hypothesis tests are used to determine if the observed differences are due to random chance or not. In Gauge R&R experiment we examine the following hypothesis tests:

- For the main effect of Parts:
  $\boldsymbol{H_0}$: The means of different parts are equal
  $\boldsymbol{H_1}$: At least one part mean is different
- For the main effect of Operators:
  $\boldsymbol{H_0}$: The means of different operators are equal
  $\boldsymbol{H_1}$: At least one operator mean is different
- For the interaction effect Part×Operator:
  $\boldsymbol{H_0}$: There is no interaction effect between parts and operators
  $\boldsymbol{H_1}$: There is an interaction effect between parts and operators

Researchers can determine whether to reject the null hypothesis for the main and interaction effects and determine whether these factors have a significant impact on the measurement variability by comparing p-values to a significance level α, commonly α = 0.05. The smaller the p-values is, the more important the corresponding factor is. The interpretation of the corresponding p-values is:

- If $\boldsymbol{p_P} < a$ we reject the null hypothesis H_0, which means that there are statistically significant differences between the parts.
- If $\boldsymbol{p_O} < a$ we reject the null hypothesis H_0, which means that there are statistically significant differences between the operators.
- If $\boldsymbol{p_O} < a$ we reject the null hypothesis H_0, which means that there are statistically significant interaction between parts and operators.

Table 4 summarizes the ANOVA measurements that one can obtain using the ANOVA method.

 https://doi.org/10.26219/heal.aueb.6231

Source of Variation	Sum of Square (SS)	Degrees of Freedom (df)	Mean of Square (MS)	Expected Mean of Square	F-ratio (F)	P-value (p)
Parts	SS_P	$n-1$	MS_P	$E(MS_P) = bn\sigma_P^2 + n\sigma_{PO}^2 + \sigma^2$	$\frac{MS_P}{MS_{P\times O}}$	p_P
Operator	SS_O	$p-1$	MS_O	$E(MS_O) = an\sigma_O^2 + n\sigma_{PO}^2 + \sigma^2$	$\frac{MS_O}{MS_{P\times O}}$	p_O
Parts×Operator	$SS_{P\times O}$	$(n-1)(p-1)$	$MS_{P\times O}$	$E(MS_{P\times O}) = n\sigma_{PO}^2 + \sigma^2$	$\frac{MS_{O\times P}}{MS_E}$	$p_{P\times O}$
Error	SS_E	$np(k-1)$	MS_E	$E(MS_E) = \sigma^2$		
Total	SS_T	$npk-1$				

Table 4: Two-way ANOVA table for a standard model of a Gauge R&R experiment.

By equating the computed numerical values of the mean squares from the analysis of variance to their expected values and solving for the variance components, one can estimate the variance components:

$$\hat{\sigma}_P^2 = \frac{MS_P - MS_{P\times O}}{on}$$

$$\hat{\sigma}_O^2 = \frac{MS_O - MS_{P\times O}}{pn}$$

$$\hat{\sigma}_{PO}^2 = \frac{MS_{P\times O} - MS_E}{n}$$

$$\sigma^2 = MS_E$$

In general, we consider the random error σ^2 as the repeatability variance component and the **reproducibility** as the sum of the operator's variability σ_O^2 and the part×operator variability σ_{PO}^2, i.e.

$$\sigma_{Repeatability}^2 = \sigma^2$$

$$\sigma_{Reproducibilty}^2 = \sigma_O^2 + \sigma_{PO}^2$$

Consequently,

$$\sigma_{Gauge}^2 = \sigma_{Repeatability}^2 + \sigma_{Reproducibilty}^2$$

or

$$\sigma_{Gauge}^2 = \sigma^2 + \sigma_O^2 + \sigma_{PO}^2 \qquad (1.26)$$

As a result the total variability is:

$$\sigma_{Total}^2 = \hat{\sigma}_P^2 + \sigma_{Gauge}^2 = \hat{\sigma}_P^2 + \sigma^2 + \sigma_O^2 + \sigma_{PO}^2$$

https://doi.org/10.26219/heal.aueb.6231

If we do not have interaction effect between parts and operators, then the sum of square, degree of freedom, and mean square should be added to error term.

The Gauge R&R study and the ANOVA procedure we analyzed in this section leaded to point estimates of the experimental model variance components and for σ^2_{Gauge}, $\sigma^2_{Repeatability}$ and $\sigma^2_{Reproducibilty}$. Therefore, construction of confidence intervals for these parameters would be highly informative and helpful in Gauge R&R analysis. Although several methods for confidence interval construction have been discussed, the Modified Large Sample (MLS) is the prevalent and most usable one. It provides the best possible results and also is relatively easy to be implemented for the standard gauge capability experiment. The MLS confidence interval formulas that are typically involved in MSA are presented in Table 5. Table 6 is collateral since it provides definitions used in Table 5. Remark that the percentage point $F_{a,df,\infty}$ of the F distribution defined in Table 6 is equivalent to $\chi^2_{a,df}$.

Parameter	Lower Bound	Upper Bound
σ^2_P	$\hat{\sigma}^2_P - \frac{\sqrt{V_{LP}}}{on}$	$\hat{\sigma}^2_P + \frac{\sqrt{V_{UP}}}{on}$
σ^2_{Gauge}	$\hat{\sigma}^2_{Gauge} - \frac{\sqrt{V_{LM}}}{pn}$	$\hat{\sigma}^2_{Gauge} + \frac{\sqrt{V_{UM}}}{pn}$
σ^2_{Total}	$\hat{\sigma}^2_{Total} - \frac{\sqrt{V_{LT}}}{pon}$	$\hat{\sigma}^2_{Total} + \frac{\sqrt{V_{UT}}}{pon}$
ρP	$L_P = \frac{pL^*}{pL^*+o}$	$U_P = \frac{pU^*}{pU^*+o}$
ρM	$1 - L_P$	$1 - U_P$

Table 5: 100(1 - α)% MLS Confidence Intervals for the Standard Gauge R & R Experiment.

Term	Definition
V_{LP}	$G_1^2 MS_P^2 + H_3^2 MS_{PO}^2 + G_{13} MS_P^2 MS_{PO}$
V_{UP}	$H_1^2 MS_P^2 + G_3^2 MS_{PO}^2 + H_{13} MS_P^2 MS_{PO}$
V_{LM}	$G_2^2 MS_O^2 + G_3^2 (p-1)^2 MS_{PO}^2 + G_4^2 p^2 (n-1)^2 MS_E^2$
V_{UM}	$H_2^2 MS_O^2 + H_3^2 (p-1)^2 MS_{PO}^2 + H_4^2 p^2 (n-1)^2 MS_E^2$
V_{LT}	$G_1 p^2 MS_P^2 + G_2^2 o^2 MS_o^2 + G_3^2 (po-p-o)^2 MS_{PO}^2 + G_4^2 (po)^2 (n-1)^2 MS_E^2$
V_{UT}	$H_1 p^2 MS_P^2 + H_2^2 o^2 MS_o^2 + H_3^2 (po-p-o)^2 MS_{PO}^2 + H_4^2 (po)^2 (n-1)^2 MS_E^2$
G_1	$1 - 1/F_{1-\frac{a}{2}:p-1,\infty}$
G_2	$1 - 1/F_{1-\frac{a}{2}:o-1,\infty}$
G_3	$1 - 1/F_{1-\frac{a}{2}:(p-1)(o-1),\infty}$

https://doi.org/10.26219/heal.aueb.6231

G_4	$1 - 1/F_{1-\frac{a}{2}:po(n-1),\infty}$
H_1	$1/F_{1-\frac{a}{2}:p-1,\infty} - 1$
H_2	$1/F_{1-\frac{a}{2}:o-1,\infty} - 1$
H_3	$1/F_{1-\frac{a}{2}:(p-1)(o-1),\infty} - 1$
H_4	$1/F_{1-\frac{a}{2}:po(n-1),\infty} - 1$
G_{13}	$\frac{(F_{1-\frac{a}{2}:p-1,(p-1)(o-1)} - 1)^2 - G_1^2 F^2_{1-\frac{a}{2}:p-1,(p-1)(o-1)} - H_3^2}{F_{1-\frac{a}{2}:p-1,(p-1)(o-1)}}$
H_{13}	$\frac{(1 - F_{\frac{a}{2}:p-1,(p-1)(o-1)})^2 - H_1^2 F^2_{\frac{a}{2}:p-1,(p-1)(o-1)} - G_3^2}{F_{\frac{a}{2}:p-1,(p-1)(o-1)}}$
L^*	$\frac{MS_P - F_{1-\frac{a}{2}:p-1,(p-1)(o-1)} MS_{PO}}{p(n-1)F_{1-\frac{a}{2}:p-1,\infty} MS_E + F_{1-\frac{a}{2}:(p-1)(o-1)} MS_O + (p-1)F_{1-\frac{a}{2}:p-1,\infty} MS_{PO}}$
U^*	$\frac{MS_P - F_{\frac{a}{2}:p-1,(p-1)(o-1)} MS_{PO}}{p(n-1)F_{\frac{a}{2}:p-1,\infty} MS_E + F_{\frac{a}{2}:(p-1)(o-1)} MS_O + (p-1)F_{\frac{a}{2}:p-1,\infty} MS_{PO}}$

Table 6: Definition of terms used in Table 5

https://doi.org/10.26219/heal.aueb.6231

CHAPTER 2: Relations between process and measurement system capability indices

As we mentioned already, a measurement system may not always yield the part's exact dimension, instead, it may yield measurements that are somewhat off from the real value. When a product turns out to be nonconforming, it is typically asserted that the variation is due to the process, and consequently, the requisite enhancing actions are implemented to optimize process capability. However, efforts like these might not always lead to actual improvements. It is possible that the process is already sufficiently capable, but there is no way of proving this due to inadequate measurement system. Furthermore, even if the measurement system is completely capable, the measurement error may still be deemed unacceptable in light of process variability. Therefore, prior to implementing any further improvement initiatives, it is imperative to investigate the variability of both a production process and a measuring system.

In this chapter we present some principal process capability indices and we discuss their relationship with different Gauge R&R metrics as well as their contribution to the overall optimization of the production process. A process capability can be determined regardless of the accuracy of the observed data. For this reason, we will examine the impact of measurement errors and we will estimate how they affect the computed process capabilities. Initially, we will use Precision-to-Tolerance and Signal-to-Noise ratio to identify the resulting distribution of the process capability, which should aid in identifying the critical values of C_p and C_{pk} indices to better prevent higher risks of measurement errors. Then we will combine both the gauge R&R and the process capability indices in the two-dimensional (σ_g, σ_P) space, in order to study and evaluate simultaneously the performance of the measurement system and the production process. Hence, this provides us a comprehensive image about the overall performance and subsequently a more direct way to pinpoint and control the potential issues.

 https://doi.org/10.26219/heal.aueb.6231

2.1 Introduction of the process capability indices C_p and C_{pk}

In performing the GR&R study, most industries are using the acceptance criteria of Precision to Tolerance (*P/T*) ratio as stipulated by QS9000[2]. On the other hand, the process capability is usually assessed using potential and actual process capability indices, C_p and C_{pk}, respectively. Meanwhile, it is highly dubious whether the same acceptance criteria are adequate for various manufacturing processes.

Process Capability Indices (PCI) consist a family of quality metrics for evaluating a manufacturing process. In particular, they express its ability to generate parts that satisfy the predetermined level of production tolerance. The potential process capability index C_p is widely used in industry, and it measures the spread of the specifications relative to the Six-Sigma spread, or stated differently, the relationship between engineering requirements and the process performance (Montgomery, Douglas C. (2019)). Under the assumptions that the process of quality characteristic follows the normal distribution, $X \sim N(\mu, \sigma_P^2)$, this capability index is expressed as:

$$\boldsymbol{C_p = \frac{USL - LSL}{6\sigma_P} = \frac{TOL}{6\sigma_P}} \qquad (2.1)$$

where μ is the mean, σ_P is the standard deviation of the process and USL and LSL are the upper and lower specification limits, respectively.

The confidence interval of this index can be computed as follows:

$$CI(C_p) = \left(C_p \sqrt{\frac{\chi^2_{a/2.n-1}}{n-1}}, C_p \sqrt{\frac{\chi^2_{1-a/2.n-1}}{n-1}} \right)$$

where $\chi^2_{a/2.n-1}$ and $\chi^2_{1-a/2.n-1}$ are the chi-squared critical values with $n-1$ degrees of freedom, n is the sample size, and a is the desired significance level.

A production process is approved by suppliers and manufacturers if $C_p \geq C_0$. Automakers and suppliers who adhere to AIAG guidelines utilize $C_0 = 1.67$ as their threshold, or in other words, a process with $C_p \geq 1.67$ is accepted. For a given C_0 we construct the following inequality (Al-Refaie, Abbas, and Nour Bata (2010)):

[2] QS9000, or Quality System requirements, was a quality standard developed by a consortium of the 'Big Three' American automakers, General Motors, Chrysler and Ford. It was introduced to the industry in 1994. The purpose of this standard was to guarantee the automotive industry's continual supply of high-quality and dependable products. Its principles and requirements were later integrated into the globally recognized ISO/TS 16949 standard, and subsequently IATF 16949, reflecting the evolving needs of the international automotive market. QS900 was useful in PCA and MSA because it offered standardized, comprehensive requirements for assessing and enhancing the quality of processes and measurement systems.

 https://doi.org/10.26219/heal.aueb.6231

$$C_p \geq C_0$$

$$\frac{TOL}{6\sigma_P} \geq C_0$$

$$\sigma_P \leq \frac{TOL}{6C_0}$$

However, the capability index C_p does not take into consideration the location of the process mean according to the predetermined specifications. Hence, there is another capability ratio, C_{pk}, that is used extensively, and takes into account both the process deviation from the target and the magnitude of process variability. This index is defined as:

$$\boldsymbol{C_{pk} = min\left\{\frac{USL-\mu}{3\sigma_P}, \frac{\mu-LSL}{3\sigma_P}\right\}} \qquad (2.2)$$

where μ represents the process mean. If $C_p = C_{pk}$ indicates that the process is centered at the midpoint of the specifications, while $C_p > C_{pk}$ indicates an off-center process.

The confidence interval of this index can be computed as follows:

$$CI(C_{pk}) = \left(C_{pk} - z_{a/2}\sqrt{\frac{1}{9n} + \frac{C_{pk}^2}{2n-1}}, C_{pk} + z_{a/2}\sqrt{\frac{1}{9n} + \frac{C_{pk}^2}{2n-1}}\right)$$

where $z_{a/2}$ is the critical value from the standard normal distribution for a two-tailed test at significance level a, and n is the sample size (Montgomery, Douglas C. (2019)).

For both C_p and C_{pk} indices, a general acceptance value is **1.33**. Values greater than **1.67** is an excellent indication about the process, whiles a value below **1** show that the process is not meeting specification limits consistently and is not capable.

The potential capability ratio C_p is easier to be calculated compared to the actual capability ratio C_{pk}, since the first one requires only the specification limits USL and LSL, and the process standard deviation σ_P, while the second one demands extra data regarding the quality characteristic center value μ.

https://doi.org/10.26219/heal.aueb.6231

2.1.1 Relation between P/T ratio and SNR/ DR

After the variance components have been estimated using ANOVA, it is straightforward to evaluate the capabilities of the production process and measuring system. As already mentioned in section 1.3.3 of Chapter 1, *P/T* ratio, Discrimination ratio (*DR*) and Signal-to Noise ratio (*SNR*), are some of the most important quality measures we use in MSA.

According to Al-Refaie, Abbas, and Nour Bata (2010) and Dalalah, Doraid (2023) , if we define the *P/T* equivalently as in eq. (1.4) by setting k=6, $USL - LSL = TOL$, and $\sqrt{\gamma M} = \sigma_g$, then we have the formula:

$$P/T = \frac{6\sigma_g}{USL - LSL} = \frac{6\sigma_g}{TOL} \qquad (2.3)$$

Therefore, the gauge variability σ_g can be expressed in terms of *P/T* criterion as:

- $\sigma_g \leq \frac{0.1TOL}{6}$ if $P/T \leq 0.1$, the gauge is capable (2.3.1)
- $\sigma_g > \frac{0.3TOL}{6}$ if $P/T > 0.3$, the gauge is incapable (2.3.2)

In the same way, since $\rho P = \gamma P/\gamma T$ and $\rho M = \gamma M/\gamma T$, by considering $\sqrt{\gamma P} = \sigma_P$ and $\sqrt{\gamma M} = \sigma_g$, from eq. (1.5) we can define *SNR* as:

$$SNR = \frac{\sqrt{2}\sigma_P}{\sigma_g} \qquad (2.4)$$

This formula, as we can see below, provides a relationship between gauge variability σ_g and part variability σ_P:

- $\sigma_P \geq \frac{5}{\sqrt{2}}\sigma_g$ if $SNR \geq 5$, the gauge is capable (2.4.1)
- $\sigma_P < \frac{2}{\sqrt{2}}\sigma_g$ if $SNR < 2$, the gauge is incapable (2.4.2)

Equivalently, we can express *DR* as:

$$DR = \sqrt{\frac{2\sigma_P^2}{\sigma_g^2} + 1} \qquad (2.5)$$

Now, if we utilize this form of *DR* instead of *SNR* for assessing the measurement capability, the relation between σ_g and σ_P becomes:

https://doi.org/10.26219/heal.aueb.6231

- $\sigma_P \geq \sqrt{\frac{15}{2}}\sigma_g$ if $DR \geq 4$, the gauge is capable (2.5.1)
- $\sigma_P < \sqrt{\frac{3}{2}}\sigma_g$ if $DR < 2$, the gauge is incapable (2.5.2)

The following diagrams in Figures 8 (a) and (b) demonstrate how the different values of *SNR* and *P/T*, plotted in space (σ_g, σ_P), can create distinct zones/ regions that correspond to measurement systems with different attributes. By combining the two thresholds, according to inequalities (2.3.1), (2.3.2) and (2.4.1), (2.4.2), more regions will result as shown in Figure 8 (c) (Dalalah, Doraid (2023)).

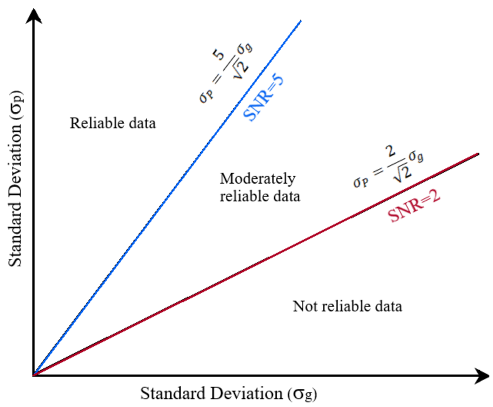

Figure 8 (a): Regions of SNR

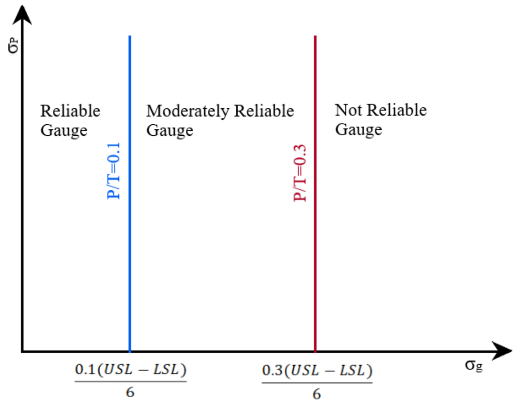

Figure 8 (b): Regions of P/T

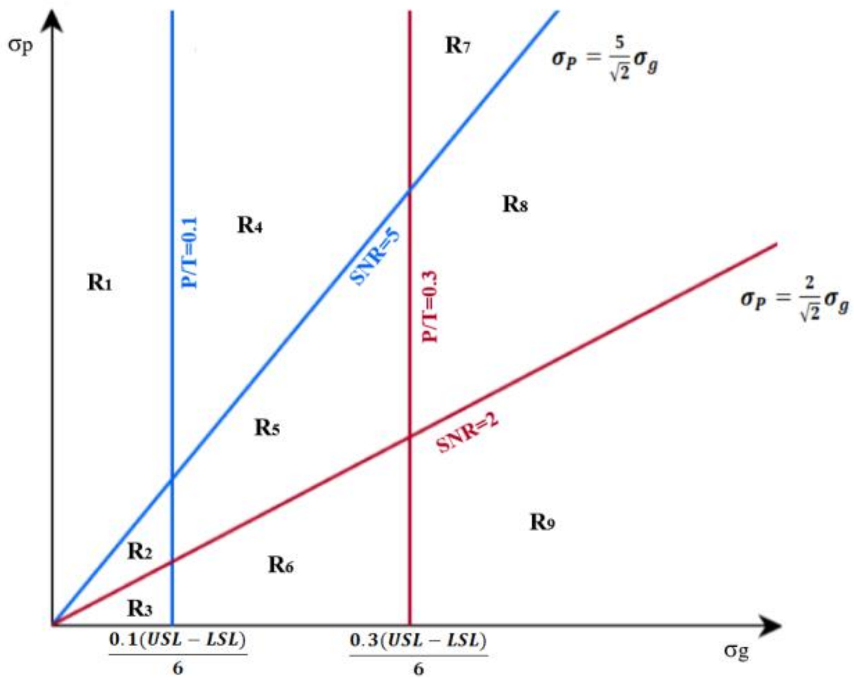

Figure 8 (c): Combined regions of SNR and P/T

Fig. 8 (a) which shows the regions of *SNR* in the two-dimensional space (σ_g, σ_P), indicates that above the blue line, i.e. where *SNR* takes values above 5, we have reliable data obtained from the measurements. The area between the blue and the red line is controversial since we have moderately reliable data, which requires further

https://doi.org/10.26219/heal.aueb.6231

investigation. The area below the red line, i.e. where $SNR \leq 2$, the data reliability is very poor. Fig. 9 (a) that shows the distinct areas for *DR*, has exactly analogous representation as Fig 8 (a).

Fig. 8 (b) and 9 (b) which show the regions of P/T ratio in space (σ_g, σ_P), confirms as we have already mention, that for values smaller than 0.1 the gauge is acceptable, for $0.1 \leq P/T \leq 0.3$ the gauge is partially accepted, and for $P/T > 0.3$ the gauge is unacceptable.

Fig. 8 (c) and Fig. 9 (c) is a combination of the other two corresponding figures, that associate both *SNR/ DR* and *P/T* distinct zones, and we extract the following inferences:

- The *SNR* (*DR*) criterion is satisfied in regions R_1, R_4 and R_7, whereas the *P/T* criterion is satisfied in regions R_1, R_2 and R_3. Hence, only in area R_1 among all regions the measurement system is accepted as capable, since both criteria are met there.
- Region R_2 meets the *PTR* requirement, but the *SNR* (*DR*) can be increased further by decreasing the gauge variability σ_g.
- In region R_3 the *P/T* ratio is valid, but the *SNR* (*DR*) criterion is not satisfied because the σ_g is still large in comparison to the σ_P. It is therefore necessary to take efforts to enhance the gauge's accuracy (e.g. recalibration of the gauge).
- In R_4 area the gauge variability $\sigma_g$should be reduced in order to meet the *P/T* requirements. However, the *SNR* (*DR*) criterion is satisfied since part variability σ_P is larger than σ_g.
- R_5 demands also reduction of σ_g to improve the gauge capability. Both *P/T* and *SNR* (*DR*) are insufficiently satisfactory.
- The gauge in region R_6 is not capable at all since neither *P/T* nor *SNR* (*DR*) takes the desirable values, so it must be changed in order to improve the measurement system efficacy.
- In R_7 the *SNR* (*DR*) criterion is satisfied since σ_P is large compared to σ_P. However, *P/T* ratio is quite high, which makes the measurement system inadequate. As a result, reduction of both σ_g and σ_P is required in order to optimize both process capability metrics *P/T* and *SNR* (*DR*).
- The measurement system in regions R_8 and R_9 deem inadequate since *PTR* and *SNR* (*DR*) criteria are not satisfied at all. Consequently, the gauge must me change in order to have better and more reliable results.

It is evident from the aforementioned findings that the *P/T* alone is an inadequate quality metric for assessing the measurement system performance. Additionally, *SNR* and *DR* provide almost identical results. As a result, *P/T* will be utilized in conjunction with each of these two quality measures to evaluate the capabilities of the measurement system and offer a relationship between gauge and parts variances.

https://doi.org/10.26219/heal.aueb.6231

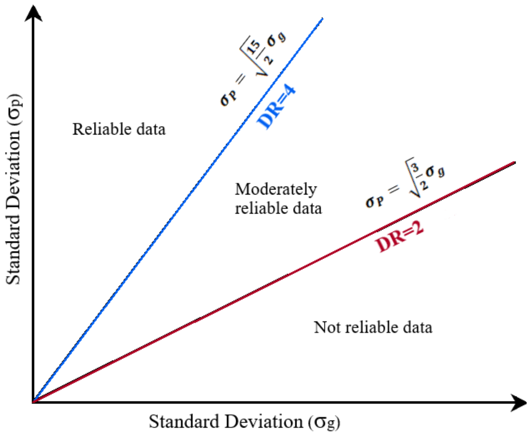

Figure 9 (a): Regions of *DR*

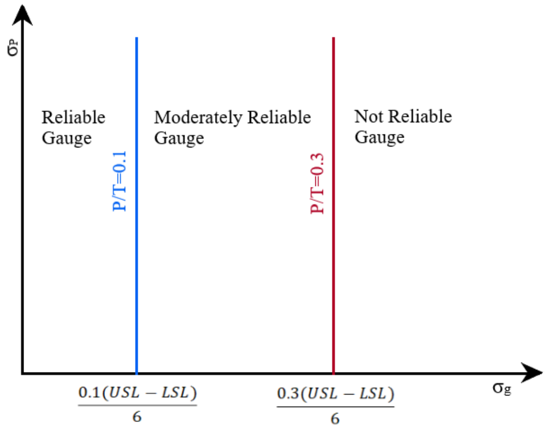

Figure 9 (b): Regions of *P/T*

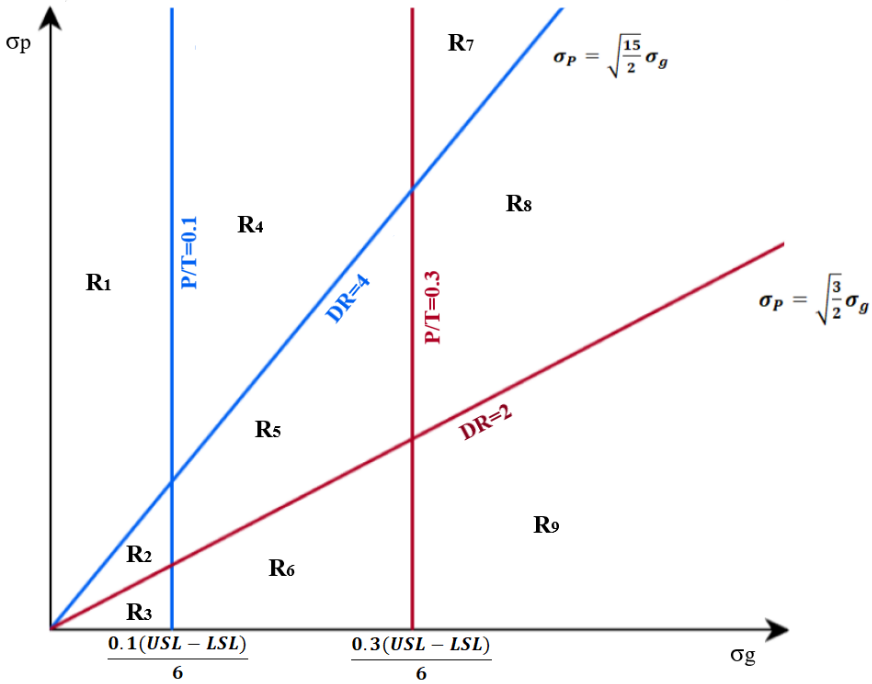

Figure 9 (c): Combined regions of *DR* and *P/T*

2.1.2 Relation between P/T ratio, SNR/ DR and C_p/ C_{pk}

As we already know, a sufficient and necessary condition for the approval of a production process is to meet the quality index C_p given by eq. (2.1). The part variability σ_P can be expressed in terms of C_p and TOL as follows:

https://doi.org/10.26219/heal.aueb.6231

- $\sigma_P \leq \frac{TOL}{6C_p}$, for a capable production process (2.6.1)
- $\sigma_P > \frac{TOL}{6C_p}$, for an incapable production process (2.6.2)

In the same way, σ_P can be expressed in terms of C_{pk} given by eq. (2.2) as follows:

- $\sigma_P \leq min\left\{\frac{USL-\mu}{3C_{pk}}, \frac{\mu-LSL}{3C_{pk}}\right\}$, for a capable production process (2.7.1)
- $\sigma_P > min\left\{\frac{USL-\mu}{3C_{pk}}, \frac{\mu-LSL}{3C_{pk}}\right\}$, for an incapable production process (2.7.2)

Therefore, to assess the capabilities of the measurement system and manufacturing process, in Fig. 10 (a) and (b), the two-dimensional space (σ_g, σ_P) incorporates the inequalities (2.6.1) and (2.6.2) in plots analogous to those above. In a similar vein, Fig. 11 (a) and (b) illustrate the corresponding space (σ_g, σ_P) using the *DR* criterion instead of *SNR* and the relations (2.7.1) and (2.7.2) respectively.

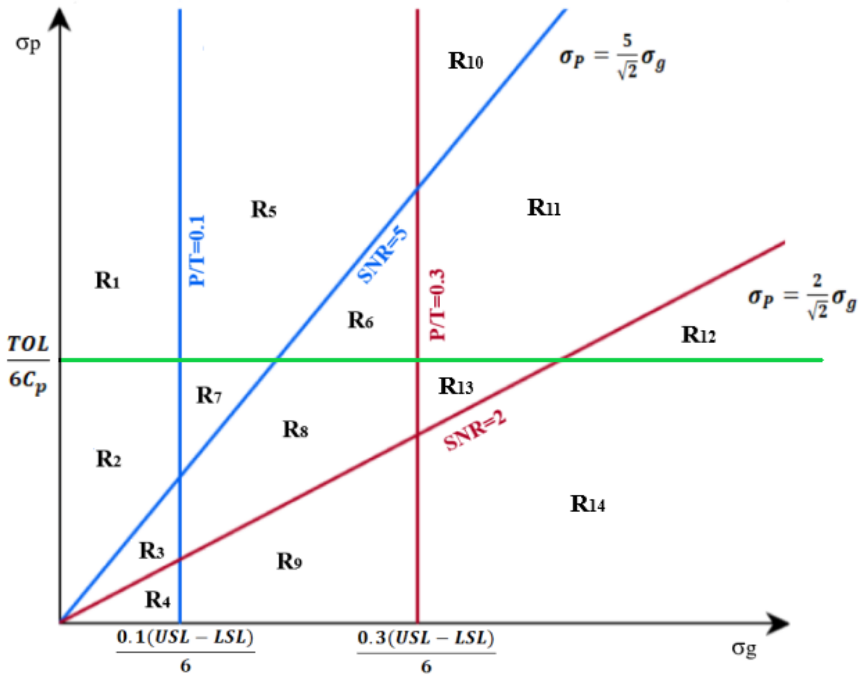

Figure 3 (a): Regions of *P/T*, *SNR* and C_p criteria

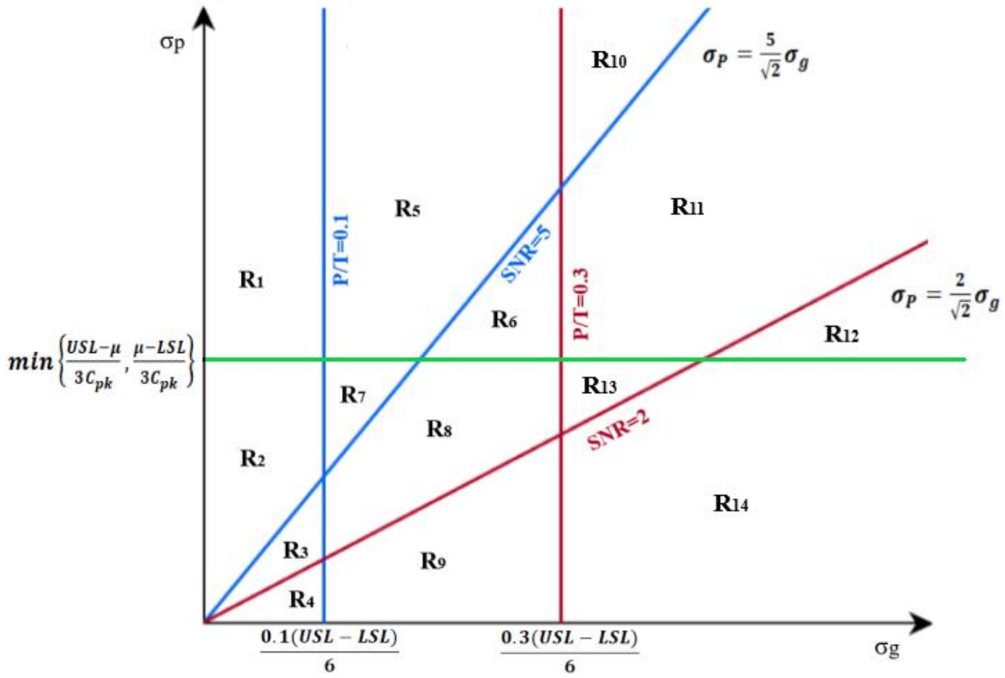

Figure 10 (b): Regions of *P/T*, *SNR* and C_{pk} criteria

Consequently, from Fig. 10 (a) and 10 (b) and equivalently from Fig. 11 (a) and 11 (b) we can draw the following conclusions ((Al-Refaie, Abbas, and Nour Bata (2010)):

- Although region R_1 satisfies the *PTR* and *SNR* (*DR*) requirements, the process is still unable due to the huge part variability σ_P. So, the reduction of σ_P is recommended.
- R_2 is the only region where all the three quality measures are satisfied and as a result both the gauge and the process capabilities are adequate there.
- Region R_3 meets the *P/T* criterion, since it falls at the left side of the blue line that indicates $P/T < 0.1$, but *SNR* (*DR*) can be optimized by decreasing the gauge variability σ_g.
- In area R_4 even though the P/T ratio is met, steps should be taken to increase the gauge accuracy since *SNR* (*DR*) are not satisfied at all.
- For regions R_5 and R_{10} both σ_g and σ_P should be reduced in order to improve *P/T* and the process capability and maintain the index *SNR* (*DR*) at the same level.
- According to region R_6, decreasing σ_g and σ_P will help to improve *PTR*, *SNR* (*DR*), and process capability.
- In regions R_7 and R_8, the reduction of σ_g is recommended in order to optimize the gauge's accuracy.
- If we want a better measurement system, the gauge in area R_9 must be updated or modified because it is not adequate at all to measure the product.
- Regions R_{11} and R_{12} propose that in order to determine the most significant parameters influencing the process and the gauge capabilities, the measurement system should be changed through a designed experiment.

 https://doi.org/10.26219/heal.aueb.6231

- As indicated by areas R_{13} and R_{14}, the process in capable, however the measurement system should be changed in order to improve gauge capability.

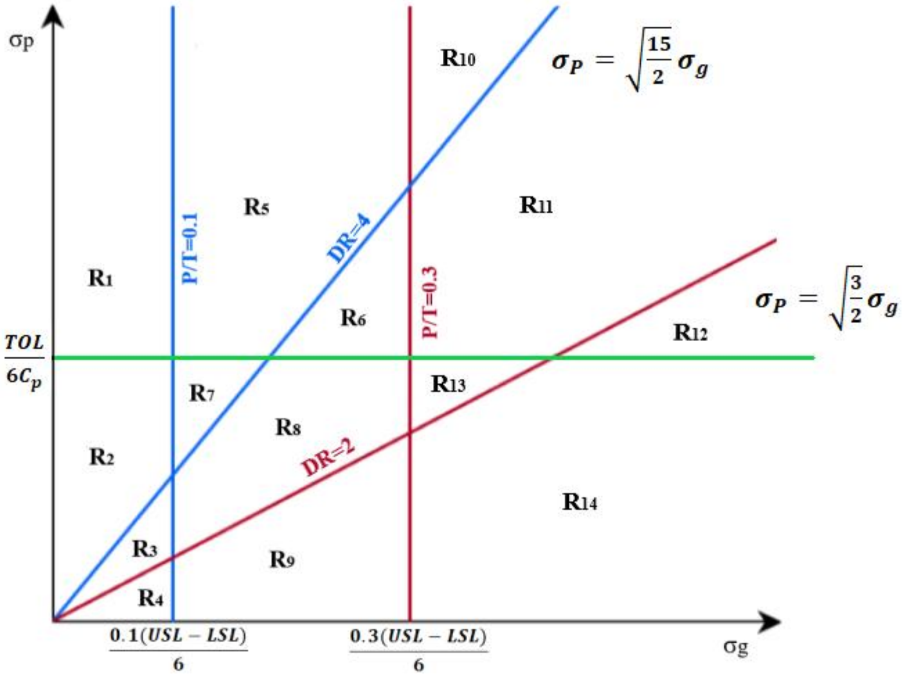

Figure 4 (a): Regions of *P/T*, *DR* and C_p criteria

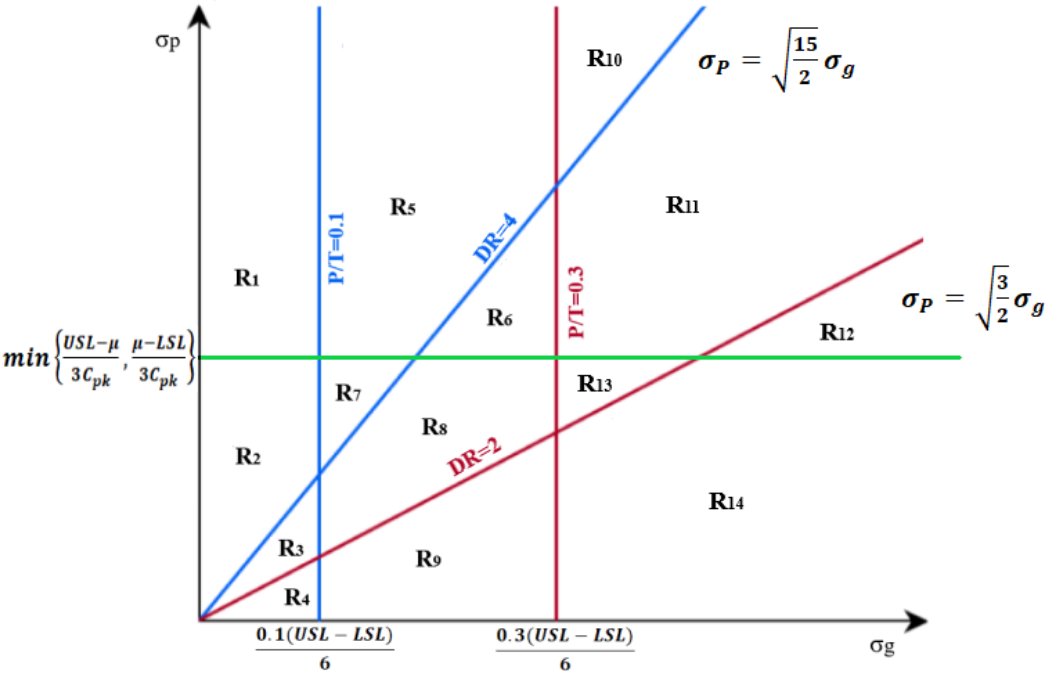

Figure 11 (b): Regions of *P/T*, *DR* and C_{pk} criteria

Table 7 additionally provides a quick summary of the conclusions drawn from Fig. 10 (a), 10 (b) and Fig. 11 (a), 11 (b). Note that, the best possible region is R_2 sine both manufacturing process and gauge are sufficiently capable.

https://doi.org/10.26219/heal.aueb.6231

Region	Satisfied measure	Unsatisfied Measure	Improvement suggestions/ decisions
R_1	*P/R, SNR/ DR*	C_p/ C_{pk}	The gage is capable but the process capability can be improved by reducing σ_P
R_2	*P/R, SNR/ DR,* C_p/ C_{pk}	None	Both the manufacturing process and the gauge are capable
R_3	*P/T,* C_p/ C_{pk}	None	The manufacturing process is capable, but there is room for improvement for *SNR/ DR* by reducing σ_g
R_4	*P/T,* C_p/ C_{pk}	*SNR/ DR*	The manufacturing process is capable, and actions to improve the accuracy of the gauge should be taken
R_5	*SNR/ DR*	C_p/ C_{pk}	Both σ_P and σ_g should be reduced in order to improve the PTR and C_p/ C_{pk}
R_6	None	C_p/ C_{pk}	Both σ_P and σ_g should be reduced in order to improve the *PTR, SNR*(*DR*), and the process capability
R_7	*SNR/ DR,* C_p/ C_{pk}	None	The manufacturing process is capable, and it is advised to lower the σ_g to increase the gauge's accuracy
R_8	C_p/ C_{pk}	None	The manufacturing process is capable, and it is advised to lower the σ_g to increase the gauge's accuracy
R_9	C_p/ C_{pk}	*SNR/ DR*	The manufacturing process is capable, but the gauge is inadequate and must be change
R_{10}	*SNR/ DR*	P/T, C_p/ C_{pk}	To improve the *PTR*, process capability, and maintain *SNR* (*DR*) to the same level, both σ_P and σ_g should be decreased.
R_{11}	None	*P/T,* C_{pk}/ C_{pk}	The gauge must be change
R_{12}	None	*P/T, SNR/ DR,* C_p/ C_{pk}	The gauge and the manufacturing process are incapable, and it must be update
R_{13}	C_p/ C_{pk}	*P/T*	The manufacturing process is capable, but the measurement system should be updated to improve the gage capability
R_{14}	C_p/ C_{pk}	*P/T, SNR/ DR*	The manufacturing process is capable, but the measurement system should be updated to improve the gage capability

Table 7: The regions obtained from relating *P/T, SNR/ DR* and C_p/ C_{pk} capability indices.

As we can confirm from Table 7, the ideal area for both the gauge and the manufacturing process is **R_2** because every criterion either measurement (*P/T, SNR, DR*) or process (C_p, C_{pk}) is satisfied. On the contrary, the most inappropriate region is **R_{12}** as it doesn't meet any of the aforementioned requirements. In the other areas

https://doi.org/10.26219/heal.aueb.6231

smaller or bigger modifications in the process or in the measurement system are suggested in order to approve the overall procedure.

This is a first potential procedure for evaluating the capabilities of a measurement system and a manufacturing process utilizing GR&R designed experiments with four performance measures, including PTR, SNR, DR, and process capability indices C_p/ C_{pk}. The company is recommended to investigate all the possible scenarios and take all the aforementioned quantities into consideration in order to make the optimal decisions.

2.2 Relating Correlation ρ with other Quality Metrics

2.2.1 Correlation in product measurements

Another measure that is utilized to evaluate how well two parties can communicate through dimensional data is the correlation in repeated measurements that we derive as a function of σ_P and σ_g. Relating *P/T* ratio, process capability index Cp, and correlation ρ develops a relationship between the three quality measures, and it offers solutions for eventual issues that may arise when employing multiple quality measurements.

Let's assume a company that produces circuit boards. Electronic manufactures will need the supplier to produce 25-30 boards. The supplier will then measure the electrical properties of these boards and ship them, along with the measurement data, to the electronics manufacturer. The electronics manufacturer will measure the boards again, generating a second set of readings. We have seen circumstances where there are statistically insignificant discrepancies in the mean and variance of the measures between a manufacturer and a supplier, yet at the same time the two sets of measurements have very low correlation. This lack of correlation in the repeat measurements is a drawback since sets the two parties unable to agree on the condition of the circuit boards (Majeske, Karl D., and Richard W. Andrews (2002)).

For a particular part, we assume that Y_1 and Y_2 stand for the supplier and manufacturer measurements, respectively. Denote that Y_1 and Y_2 are normally distributed and satisfy eq. (1.1): $y = X + \varepsilon$ from Chapter 1, where X is the actual true value, ε is the gauge error and y represents the observed/ measured value. Consequently, each variable has mean $\mu_{Y_1} = \mu_{Y_2} = \boldsymbol{\mu}$ and variance $\sigma_{Y_1}^2 = \sigma_{Y_2}^2 = \boldsymbol{\sigma_P^2} + \boldsymbol{\sigma_g^2}$. Then the correlation is given by:

 https://doi.org/10.26219/heal.aueb.6231

$$\rho(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sigma_{Y_1}\sigma_{Y_2}} \qquad (2.8)$$

and can be expressed in terms of gauge variance σ_g and product variance σ_P. Initially, note that:

- $E[Y] = E\big[E[Y|X]\big] = E[X]$
- $Var[Y] = \sigma_P^2 + \sigma_g^2$
- $Cov(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1]E[Y_2]$

Therefore,

$$E[Y_1 Y_2] = E[E[Y_1 Y_2|X] = E[E[Y_1|X]E[Y_2|X]]$$

$$= E[XX] = E[X^2]$$

and now

$$Cov(Y_1, Y_2) = E[X^2] - (E[X])^2 = \sigma_P^2$$

As a result the correlation ρ can be expressed as,

$$\rho(Y_1, Y_2) = \frac{\sigma_P^2}{\sqrt{\sigma_P^2 + \sigma_g^2}\sqrt{\sigma_P^2 + \sigma_g^2}} = \frac{\sigma_P^2}{\sigma_P^2 + \sigma_g^2} = \frac{1}{1 + \frac{\sigma_g^2}{\sigma_P^2}} \qquad (2.9)$$

Equation (2.9) displays the association between the correlation in repeated measurements, and the ratio of gauge variability to part variability. When the ratio ρ is equal to zero, i.e. when we have zero measurement error variance, then the correlation is one, which means that the repeat measurements are perfectly correlated. Moreover, when gauge variance equals to product variance, i.e. $\sigma_g^2 = \sigma_P^2$, the correlation is 0.5.

Remark that the correlation is always greater than zero and approaches one in value the when the gauge variance is smaller compared to the part-to-part variance. At the same time, a smaller correlation results from a bigger gauge variance in respect to part-to-part variance. Stated differently, a high correlation indicates inefficient performance for the production process and/ or a capable and reliable measurement process, whereas a low correlation is a proof of a good production process and/ or an alarm for the measurement process. Undoubtedly, both variances are interesting, and it's crucial to maintain them both as small as possible. Hence, this investigation implies that correlation cannot supply all the answers on its own and we need more quality metrics to reach a conclusion.

 https://doi.org/10.26219/heal.aueb.6231

2.2.2 Relation of Correlation ρ with Precision-to-Tolerance ratio *P/T*

Suppose now, that a manufacturer has decided that in order for a measurement system to be deemed suitable for measuring parts, it must have a precision-to-tolerance ratio of $P/T \leq \delta_0$. Furthermore, the manufacturer demands that the correlation between his measurements and those of the supplier be at least $\rho \geq \rho_0$. As a result, the supplier for a given δ_0 can obtain the following inequality to meet the P/T criterion:

$$P/T \leq \delta_0$$

$$\frac{6\sigma_g}{TOL} \leq \delta_0$$

$$\sigma_g \leq \frac{\delta_0 TOL}{6} \qquad (2.10)$$

As we can observe, for a fixed tolerance, the precision-to-tolerance ratio depends only on gauge variability.

In addition, from eq. (2.9) and for a given ρ_0, the manufacturer and the supplier can obtain this inequality to satisfy the correlation criterion:

$$\rho \geq \rho_0$$

$$\frac{1}{1 + \frac{\sigma_g^2}{\sigma_P^2}} \geq \rho_0$$

$$\sigma_P \geq \sigma_g \sqrt{\frac{\rho_0}{1 - \rho_0}} \qquad (2.11)$$

Here, the correlation criterion depends on both gauge and product variability (Majeske, Karl D., and Richard W. Andrews (2002)).

Figure 12 is a graphical representation in the two-dimensional space (σ_g, σ_P) of precision-to-tolerance ratio *P/T* and correlation criterion incorporating the inequalities (2.10) and (2.11). The two lines which correspond to the two criteria separate the whole area into four distinct regions. Gauges whose σ_g is on or to the left of the vertical blue line would pass the *P/T* criterion while gauges on or above the red line would meet the correlation criterion.

https://doi.org/10.26219/heal.aueb.6231

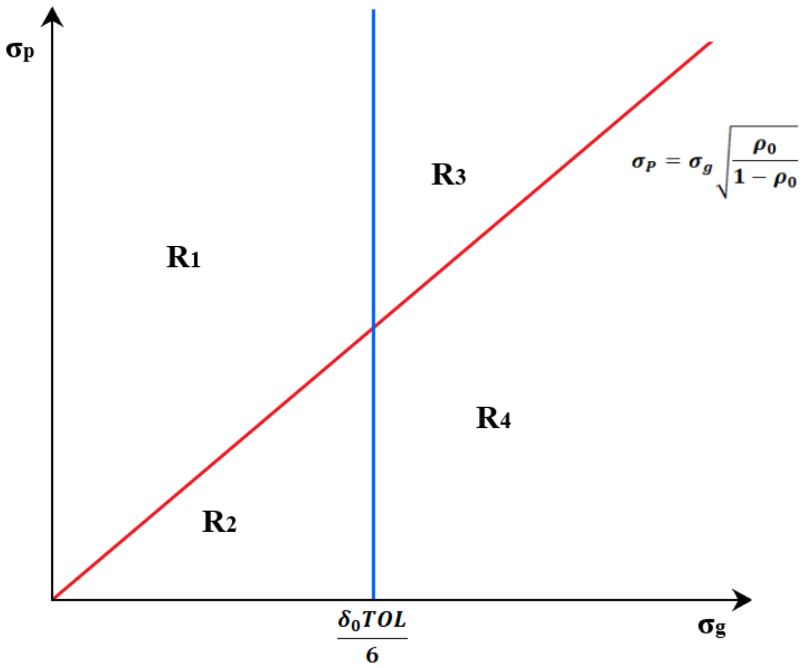

Figure 5: Comparison of *P/T* and correlation ρ gauge approval criteria.

More precisely, the following remarks are helpful:

- In region R_1 the gauge satisfies both *P/T* and correlation requirements and the manufactures should approve it.
- In region R_2 the gauge does not meet correlation, since it is bellow the red line but it passes the *P/T* criterion.
- In region R_3 the gauge passes the correlation criterion, but it does not meet *P/T*, as it is at the right side of the blue line, i.e. it is not accurate enough to measure parts.
- In region R_4 the gauge does not meet any of the required criteria, since it is located under the red line and in the right side of the blue line, so the manufactures should reject it.

Additionally, it is crucial to predefine the following quantities to construct the diagram in Fig. 12:

i. TOL : the design tolerance range for the quality characteristics
ii. δ_0 : upper bound/ the maximum acceptance value for P/T ratio to approve a gauge
iii. ρ_0 : the lower bound/ minimum acceptance value for the correlation criterion

By substituting the above three quantities in eq. (2.10) and (2.11) the analyst can create the analogous lines in the (σ_g, σ_P) space and make inferences.

https://doi.org/10.26219/heal.aueb.6231

2.2.3 Relation of Correlation ρ with Precision-to-Tolerance ratio *P/T* and Process Capability C_p

In this subunit we will investigate the relationship between correlation ρ, precision-to-tolerance ratio *P/T* and process capability C_p. As we have already prove, all these quality metrics can be expressed in terms of gauge variability σ_g and part variability σ_P as follows:

- $\rho = \frac{1}{1+\frac{\sigma_g^2}{\sigma_P^2}}$ (2.12) correlation in repeat measurements
- $P/T = \frac{6\sigma_g}{TOL}$ (2.13) precision-to-tolerance ratio
- $C_p = \frac{TOL}{6\sigma_g}$ (2.14) process capability

All the above quantities for a predetermined tolerance range, TOL, are defined by gauge and process variance σ_g and σ_P. By solving eq. (2.13) with respect to σ_g and eq. (2.14) with respect to σ_P we have the corresponding equations:

$$\sigma_g = \frac{\left(\frac{P}{T}\right) TOL}{6} \qquad (2.15)$$

$$\sigma_P = \frac{TOL}{6C_P} \qquad (2.16)$$

By substituting them in eq. (2.12), we obtain the relationship:

$$\rho = \frac{1}{1 + (\frac{P}{T} C_p)^2} \quad (2.17)$$

Manufactures aiming for consistency and reliability in their quality measures should figure out what acceptable values for *P/T* and C_P are and then use eq. (2.17) to solve for ρ. Equivalently as above, by plotting eq. (2.10), (2.16) and (2.11) in space (σ_g, σ_P), we graphically illustrate in Fig. 13 the relationship between those three quantities.

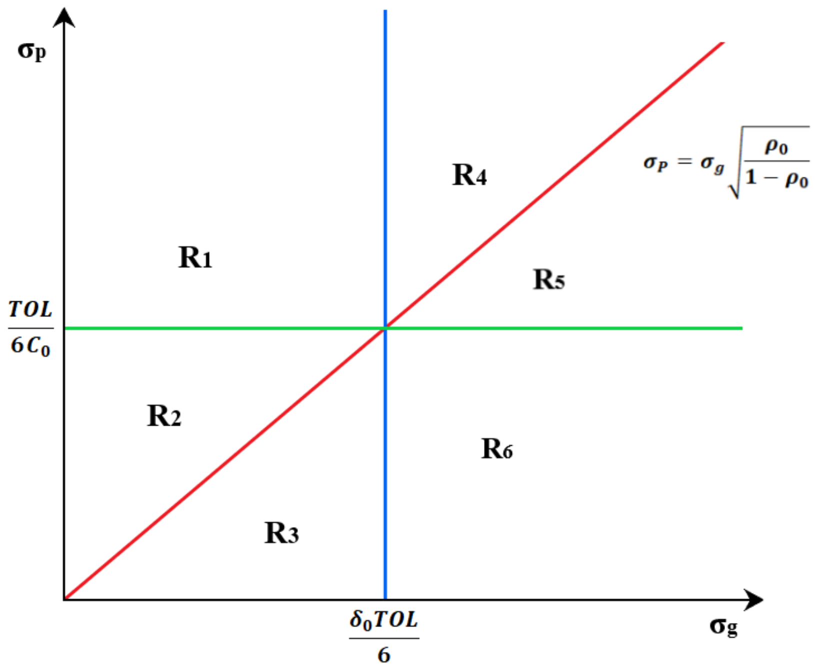

Figure 6: Comparison of *P/T*, correlation ρ, and process capability index C_P approval criteria.

Now the area is divided in six regions, since we added the green line that corresponds to C_p criterion, and indicates an upper bound for the process. So according to this index, the production process is capable if σ_P takes values smaller or equal to $\frac{TOL}{6C_P}$, i.e. if we are under the green line. Consequently, for the graph in Fig. 13 based on Majeske, Karl D., and Richard W. Andrews (2002) we can make the following inferences:

- In region R_1 the measurement system is considered acceptable, because it meets *P/T* and correlation requirements. However, C_p criterion rejects the production process. There is a remarkable advantage of being in this region, since the manufacturer has the ability to detect potential improvements for the production process and reduce the part variability σ_P.
- All the three requirements are satisfied in region R_2, so the manufacturer and supplier can approve both the measurement system and the process.
- In region R_3, despite the fact that the process and gauge meet the capability standards and precision-to-tolerance ratio, respectively, the correlation restriction is not met.
- In region R_4 the measurement system meets the correlation criterion, but it does not satisfy the P/T and C_p indices. The manufacturer is advised to reduce both gauge and process variance.
- In region R_5 none of the three criteria are satisfied. The manufacturer should reject the process and make a continuous effort to decrease both the measurement error and product variation.

https://doi.org/10.26219/heal.aueb.6231

- In region R_6, even though the correlation and measurement requirements are unsatisfactory, the process is approved for production. Hence, the manufacturer should immediately work on measurement system improvement.

As we can confirm from the comments above, the ideal area for both the manufacturer and supplier is **R_2** because the measurement system and the production process are approved since they meet all the requirements. On the other hand, **R_5** represents the worst possible scenario seeing that neither the correlation criterion, nor the two quality metrics C_p and *P/T*, are satisfied. The other four regions indicate modifications and updates in production process or in measurement system to guarantee reliable and consistent results for both the manufacturer and supplier.

This is another potential approach for a company to analyze the overall manufacturing process. Once the constraining lines are plotted, the estimates of σ_P and σ_g will signal which of the three quality criteria have been met. The manufacturer should take into account all three quality measures *P/T*, C_p and ρ in order to approve or reject the production process and/ or the measurement system.

https://doi.org/10.26219/heal.aueb.6231

CHAPTER 3: Relations between the Actual AC_p and the Observed OC_p process capability indices

As we have already discussed, gauge errors usually distort process capabilities, yielding two different values, i.e. the observed process capability index (OC_p) and the actual one (AC_p). Measured data tends to be less reliable due to gauge uncertainty. Consequently, the estimated process capability indices will differ, leading to an underestimating of the actuation potentials. Therefore, Gauge R&R studies are traditionally conducted according QS9000 standards to ensure the measurement system's accuracy. The primary objective of this chapter is to describe the relationship between OC_p and AC_pthrough signal-to-noise ratio (SNR) and consequently to jointly evaluate the gauge and the two capability indices.

3.1 Critical value of process capability C_p index

We have already define in 2.1 unit of the 2nd chapter the C_p index as $C_p = \frac{USL-LSL}{6\sigma_P}$ under the assumption of normality. The numerator gives the range between the upper and lower specification limits (tolerance) predetermined by product designers, whereas, the range of the actual process variation is given in the denominator. This quality metric uses the product's standard deviation (σ_P) to scale the tolerance in order to approve or reject a manufacturing process.

We can perform a statistical test for σ_P in the following form:

$$H_0 : \sigma_P \geq \Sigma$$

$$H_1 : \sigma_P < \Sigma$$

with Σ being some reference above which the manufacturing process is rejected. However, this test is equivalent to nullifying the process incapability through C_p index, given that if $C_p \leq c$, where c is the critical value, the process is considered as incapable. Then the hypothesis test is expressed as:

$$\boldsymbol{H_0} : C_p \leq c \text{ (the process is capable)}$$

$$\boldsymbol{H_1} : C_p > c \text{ (the process in not capable)}$$

By using the estimator $\hat{\sigma}_P = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$ as an estimator for the process standard deviation σ_P, and $\hat{C}_p = \frac{USL-LSL}{6\hat{\sigma}_P}$ for the capability index C_p, we can prove that the ration between the actual and the estimated indices follow the χ^2 distribution, i.e.

 https://doi.org/10.26219/heal.aueb.6231

$$(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 \sim \chi^2_{n-1}$$

and the critical value c_0 of the above hypothesis is given by

$$c_0 = c\sqrt{\frac{n-1}{\chi^2_{\alpha,n-1}}} \qquad (3.1)$$

where $a = P(\hat{C}_p < c_0 | C_p = c)$ and $\chi^2_{\alpha,n-1}$ is the value of χ^2 distribution that correspond to the upper α quantile and has $n-1$ degrees of freedom (for further details check Appendix 1). Under the assumption of the ideal gauge, α is typically preferred to be as small as possible.

Accordingly, a practitioner can determine the estimated index's value and subsequently test it for approval or rejection. For instance, if we want to assess the capability of a process at $c = 1.67$ (i.e. $H_0 : C_p \leq 1.67 \quad VS \quad H_1 : C_p > 1.67$) with $a = 5\%$ confidence level and for $n = 100$ samples, then by applying the formula in eq. (3.1) we take a critical value of $c_0 = 1.498$. Hence, the process is approved as capable only if the calculated C_p is higher thanc_0 (Dalalah, Doraid (2023), Dalalah, Doraid, and Dania Bani Hani (2016)).

3.2 The actual AC_p and the observed OC_p process capability

As measured errors are unavoidable, gauge analysis is vital to be conducted before any process improvements are made. The process and measuring tools are typically identified as the primary sources of variability. As we previously touched upon, the major components of gauge variation, $\sigma^2_{Repeatability}$ and $\sigma^2_{Reproducibility}$, are estimated via Gauge R&R analysis. In particular, **repeatability** assesses the consistency of the measurement system when all factors except the repetition are held constant, whereas **reproducibility** evaluates the degree of the instrument's uniformity among different operators. So, the gauge standard uncertainty is expressed as

$\sigma^2_g = \sigma^2_{Repeatability} + \sigma^2_{Reproducibility}$ and the total variability as

$\sigma^2_{Total} = \sigma^2_P + \sigma^2_{Repeatability} + \sigma^2_{Reproducibility} = \sigma^2_P + \sigma^2_g$.

The actual process capability AC_p is defined as:

$$\boldsymbol{AC_p = \frac{USL - LSL}{6\sigma_P}} \qquad (3.2)$$

and relates the range of the specification limits to the Six-Sigma range of the process.

https://doi.org/10.26219/heal.aueb.6231

By considering the gauge uncertainty, the corresponding observed process capability OC_p can be written as:

$$\boldsymbol{OC_p = \frac{USL - LSL}{6\sigma_{Total}} = \frac{USL - LSL}{6\sqrt{\sigma_P^2 + \sigma_g^2}}} \qquad (3.3)$$

By solving eq. (3.3) with respect to σ_P, we obtain the following expression for part variability:

$$\sigma_P = \sqrt{\left(\frac{USL - LSL}{6OC_p}\right)^2 - \sigma_g^2} \qquad (3.4)$$

Then, we substitute it in the AC_p eq. (3.4) and we get:

$$AC_p = \frac{1}{\sqrt{\left(\frac{1}{OC_p}\right)^2 - \left(\frac{6\sigma_g}{USL - LSL}\right)^2}} \qquad (3.5)$$

(See in Appendix 2 the proof of eq. (3.5)) (Dalalah, Doraid (2023), Dalalah, Doraid, and Dania Bani Hani (2016)).

From this formula we can draw some noticeable conclusions. When $\sigma_g \to 0$, the OC_p converges to AC_p. On the contrary, as σ_g, the OC_p is getting lower than AC_p. In other words, the more accurate and precise the gauge is, the closer the values of OC_p and AC_p will be. Otherwise, the observed process capability index OC_p will be always smaller than the actual AC_p.

Let's assume now a quality attribute of the random variable $X = (X_1, X_2, \cdots, X_n)$ where $X \sim N(\mu, \sigma_P^2)$, then the estimator of the actual process capability AC_p is:

$$\widehat{AC}_p = \frac{USL - LSL}{6\hat{\sigma}_P}$$

where $\hat{\sigma}_P = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$. But in practice, the sample observations include also the gauge error. So, it is more realistic to assume that the random sample as $Y = (Y_1, Y_2, \cdots, Y_n)$, where $Y \sim N(\mu, \sigma_P^2 + \sigma_g^2)$. Consequently, the natural estimator for the observed process capability OC_p can be expressed as:

$$\widehat{OC}_p = \frac{USL - LSL}{6\hat{\sigma}_{Total}}$$

where $\hat{\sigma}_{Total} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$. As an aftermath, these estimated indices, $\widehat{AC}_p$ and $\widehat{OC}_p$, will follow different distributions. Figure 14 bellow illustrates graphically this fact.

https://doi.org/10.26219/heal.aueb.6231

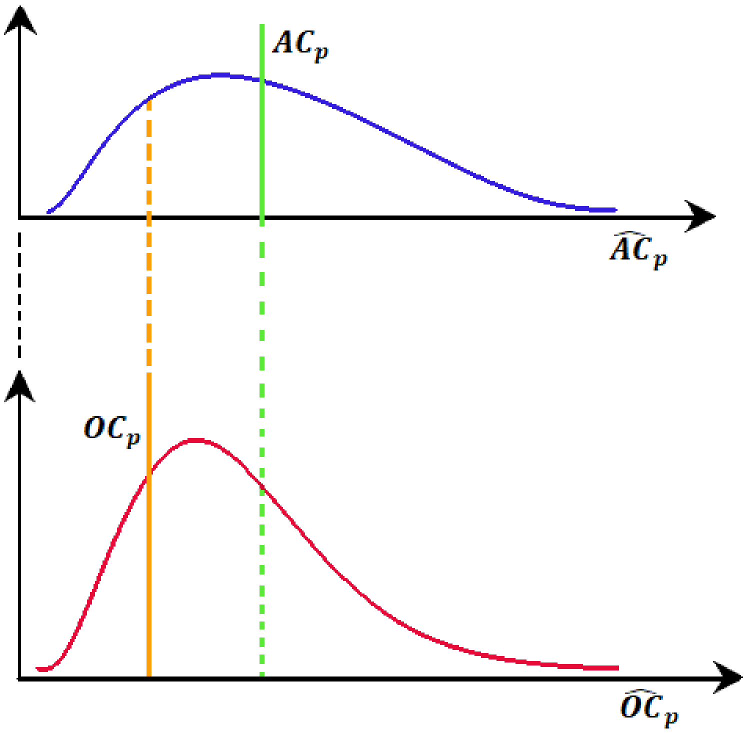

Figure 7: Distributions of estimated process capability indices $\widehat{AC}_p$ and $\widehat{OC}_p$.

Note of that where the value of $\widehat{AC}_p > \widehat{OC}_p$, the variability of $\widehat{AC}_p$ is also greater than the variability of $\widehat{OC}_p$.

3.2.1 The AC_p, OC_p and χ^2 distribution

Let's consider now that we obtain some data from a measurement system with a known signal-to-noise ratio (*SNR*) from which the process capability index $\hat{C}_p$ is estimated. With a certain probability, this computed index may be actual, i.e. no gauge errors exist, or observed, i.e. gauge errors exist. So, the analysts use statistical hypothesis tests to draw conclusions about this estimated value. In particular, as we explained in section 3.1, if $C_p > c$ (H_1), the process satisfies the quality requirements, while if $C_p \leq c$ (H_0), the null hypothesis cannot be rejected. The quantity *c* represents a benchmark/ reference value, such as 1, 1.33, 1.67, 2, …, etc.

Figure 15 illustrates the distributions of both process capability estimators $\widehat{AC}_p$ and $\widehat{OC}_p$. Remark that for a fixed rejection region in the actual distribution, any increase in the area of the observed probability density function (pdf) of the process capability indicates that the $\widehat{OC}_p$ distribution is approaching the $\widehat{AC}_p$ distribution. For both distributions, the critical value c_0 serves the acceptance limit.

https://doi.org/10.26219/heal.aueb.6231

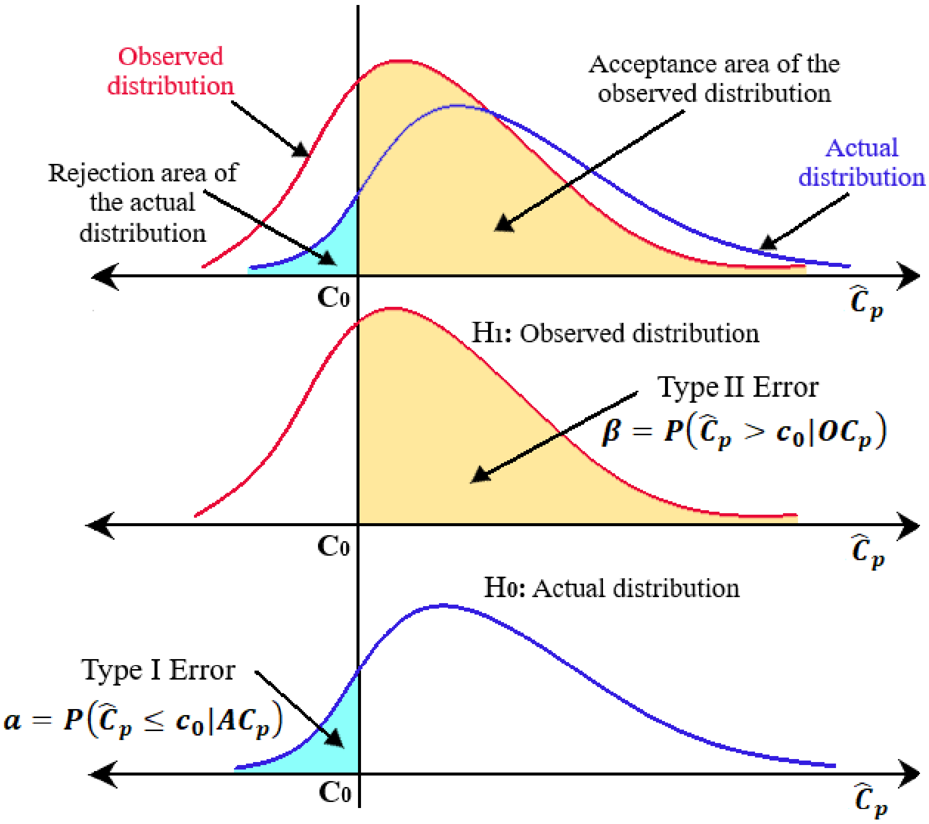

Figure 8: The distribution of the actual AC_p and observed OC_p process capability indices.

Since the actual distribution represents the in-control condition of the process, which exhibits non-negative range with higher variance, the null hypothesis H_0 will be to the right of the alternative one H_1 (Dalalah, Doraid (2023), Dalalah, Doraid, and Dania Bani Hani (2016)).

In earlier studies, the probability density function (pdf) of the capability index with the presence of gauge measurement errors has been well established by using PTR as the sole parameter and R as the reference value (i.e. a parameter of the distribution), which equals the index value given by the common formula $C_p = \frac{USL-LSL}{6\sigma}$. Therefore, we have:

$$f_{\hat{C}_p^g}(x) = 2\frac{\left(\sqrt{\frac{n-1}{2}}R/\sqrt{1+PTR^2R^2}\right)^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)}x^{-n}exp\left(\frac{-(n-1)R^2(2x^2)^{-1}}{1+PTR^2R^2}\right) \quad (3.6)$$

Similarly, we can define the distribution of eq. (3.6) as a function of SNR. The derivation of the $\hat{C}_p$ distribution is presented in Appendix 3, and the final formula is:

$$f_{\hat{C}_p^g}(x) = \frac{\left(\sqrt{\frac{n-1}{2}}R/\sqrt{1+2/SNR^2}\right)^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)}x^{-n}exp\left(\frac{-(n-1)R^2(2x^2)^{-1}}{1+2/SNR^2}\right) \quad (3.7)$$

 https://doi.org/10.26219/heal.aueb.6231

where $f_{\hat{C}_p^g}$ is the distribution of the capability index when the gauge errors are taken into account. Figure 16 shows a pdf plot for different values of SNR (Dalalah, Doraid (2023)).

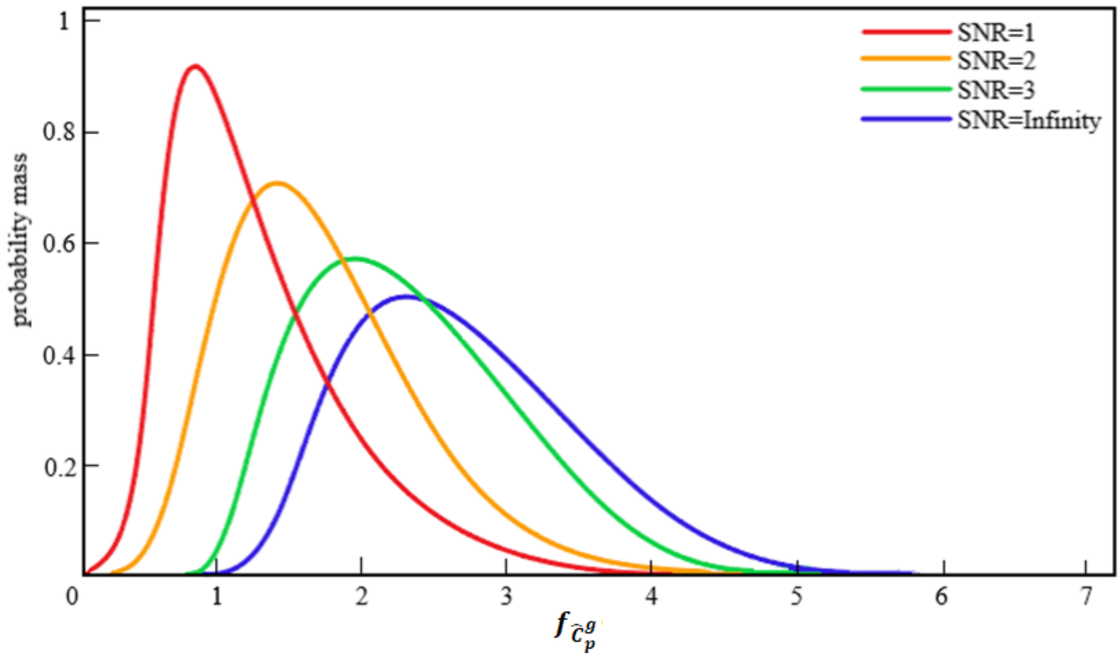

Figure 9: The actual and the observed distributions for different values of SNR for a fixed sample size n and a reference value R.

The blue line corresponds to the actual process capability density plot with $SNR \to \infty$, while the other three represent the observed process capability density plots for different values of SNR. We can observe that by having high values of SNR, the distribution of $\widehat{OC}_p$ gradually approaches the actual distribution. Remark that the variance of the $\widehat{OC}_p$ is always less than the variance of $\widehat{AC}_p$ (the blue line is "wider" than the other three), which leads to underestimation of the actual process estimation. Moreover, as $SNR \to \infty$, the equation (3.6) becomes equivalent to the equation (3.7) as $PTR \to 0$.

Measurement system inaccuracies usually have an impact on the capability index's estimation. Therefore, according to Dalalah, Doraid (2023), the following hypothesis test can be conducted to determine whether the capability index is the actual one or not.

$$H_0: R = AC_p$$

$$H_1: R = OC_p$$

When H_0 is true, the errors do not exist in the measurements, whereas when H_1 is true, the gauge errors are statistically significant and affect the measurements. Note of that both $\widehat{AC}_p$ and $\widehat{OC}_p$ satisfy the equation (3.5). As a consequence, the risk value of a (i.e. Type I error) of the critical point c_0 can be written in the following form:

 https://doi.org/10.26219/heal.aueb.6231

$$a = P(reject\ H_0|H_0\ is\ true) \quad (3.8)$$

which is equivalent to

$$a = P(\hat{C}_p < c_0 | R = AC_p) \quad (3.9)$$

where a is the probability that the estimated $\hat{C}_p$ is reported to include measurement errors, while it does not, or in other words, claiming that $\hat{C}_p$ is erroneous, while it is actual/ true. Relating this to the χ^2 distribution by taking the reciprocal, squaring both sides and multiplying by the degrees of freedom $(n-1)$ will result in:

$$a = P\left((n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 > (n-1)\left(\frac{AC_p}{c_0}\right)^2 |R = AC_p\right) \quad (3.10)$$

We have already proved in Appendix 1 that $(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2$ follows the χ^2_{n-1} distribution, so we can conclude to this formula:

$$a = P\left(\chi^2_{n-1} > (n-1)\left(\frac{AC_p}{c_0}\right)^2\right) \quad (3.11)$$

So, for a given certain actual value (AC_p), the critical point c_0 is calculated as:

$$c_0 = AC_p\sqrt{\frac{n-1}{\chi^2_{\bar{a},n-1}}} \quad (3.12)$$

where $\chi^2_{\bar{a},n-1}$ is the value of χ^2 distribution that corresponds to the upper percentile of α with $n-1$ degrees of freedom (the upper dash in α indicates the upper percentile of the χ^2 distribution).

In the same way we can define the risk value of β (i.e. Type II error) of the critical point c_0 as:

$$\beta = P(accept\ H_0|H_1\ is\ true) \quad (3.13)$$

which is equivalent to

$$\beta = P(\hat{C}_p > c_0 | R = OC_p) \quad (3.14)$$

where β is the probability that the estimated $\hat{C}_p$ is reported with no measurement errors, while in fact it does, or state id differently, claiming that $\hat{C}_p$ is the true/ actual, while it is not. We relate in the similar way as above the formula in eq. (3.8) to the χ^2 distribution and we get the expression:

 https://doi.org/10.26219/heal.aueb.6231

$$\beta = P\left((n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 < (n-1)\left(\frac{OC_p}{c_0}\right)^2 |R = OC_p\right) \quad (3.15)$$

Since $(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2$ follows the χ^2_{n-1} distribution, so we can equivalently conclude to the formula:

$$\beta = P\left(\chi^2_{n-1} < (n-1)\left(\frac{OC_p}{c_0}\right)^2\right) \quad (3.16)$$

As a result, for a given certain actual value (AC_p), the critical point c_0 is computed as:

$$c_0 = OC_p\sqrt{\frac{n-1}{\chi^2_{\underline{\beta},n-1}}} \quad (3.17)$$

where $\chi^2_{\underline{\beta},n-1}$ is the value of χ^2 distribution that corresponds to the lower percentile of β with $n-1$ degrees of freedom (the lower dash in β indicates the lower percentile of the χ^2 distribution). Figure 17 illustrates graphically α and β regions on the same χ^2 distribution (in a similar way to Fig. 15), which corresponds to Type I and Type II errors respectively. The two percentiles $\chi^2_{\bar{a},n-1}$and $\chi^2_{\underline{\beta},n-1}$ can be found both either from the χ^2 distribution or from the actual and observed distributions. We have therefore located the critical ratios/ values: $(n-1)\left(\frac{AC_p}{c_0}\right)^2$ and $(n-1)\left(\frac{OC_p}{c_0}\right)^2$, on the same χ^2 distribution. It is also important to highlight that the distance between these critical ratios/ values is nullified when OC_p approaches AC_p, which is attainable when $\alpha + \beta \to 1$ and it happens under the scenario of a perfect gauge.

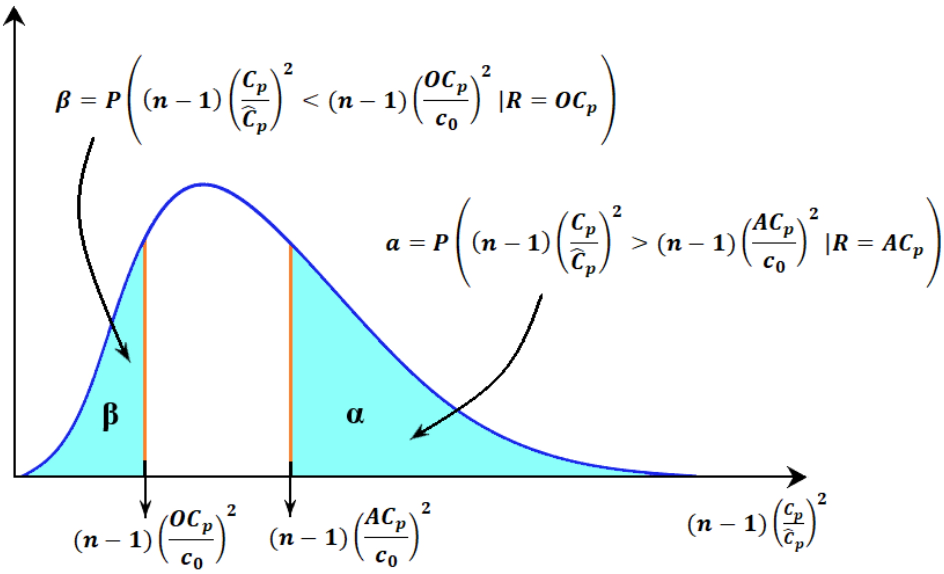

Figure 10: The risk values of α and β in the χ^2 distribution.

https://doi.org/10.26219/heal.aueb.6231

Generally, gauge error is unavoidable and therefore the concurrent study of the observed and actual capability indices AC_p and OC_p is very important, since it provides a more objective and realistic perspective on the process. Signal-to-noise ratio is employed as a mean to jointly evaluate the capacity and gauge indices. Also, the likelihood that a process capability has been measured with an erroneous gauge can now be ascertained by computing the distribution of the process capability as a function of SNR (Dalalah, Doraid (2023)).

 https://doi.org/10.26219/heal.aueb.6231

CHAPTER 4: Computational Part

Measurement System Analysis and Gauge R&R studies have been evolved rapidly in the context of experimental applications and nowadays are extremely famous especially in the industrial sector. Since measurement error is ubiquitous, the validity and accuracy of the data is the primary target. As a result, every company has as a priority the securing of a reliable measuring equipment.

In this chapter, we will apply the biggest part of the aforementioned theory by setting-up a hypothetical numerical experiment, in order to study further this field from a practical aspect. By quantifying the repeatability and reproducibility of the measurement system, we can ensure that the observed variations in compression force are truly reflective of the spring quality, rather than an artifact of measurement system errors. For this purpose we will use the **R** statistical software to create simulated data and conduct the appropriate analysis.

4.1 Experimental set-up

Suppose that we are statistical analysts in an automotive industry and we work in the department of manufacturing coil springs for suspensions. In general, the equipment consists of the measuring gauge and various operators/ appraisers. The objective is to evaluate the dependability of the in question measurement system used to determine the compression force of these coil springs by identifying the proportion of the observed variation in the compression force measurements that can be attributed to the measurement device versus to the operators measuring the force. In particular:

- **Part/ Product:** Coil springs used in automotive suspension systems ($p = 15$ parts)
- **Measurement:** Compression force (in Newton scale, N) ($o = 3$ operators)
- **Sample:**
  - We collect randomly 15 coil springs from the production line
  - Each spring is measured 2 times by 3 operators using the same force gauge ($n = 2$ repetitions)
- **Population assumption:** The compression force values are normally distributed with a mean of 100 N and a standard deviation of 4 N.

As a result, for each of the 15 springs, there are 6 measurements (2 per operator, with 3 operators). Consequently, the total number of readings is:

https://doi.org/10.26219/heal.aueb.6231

$$\mathbf{15}\ \boldsymbol{springs} \times \mathbf{3}\ \boldsymbol{operators} \times \mathbf{2}\ \boldsymbol{repeatitions} = \mathbf{90}\ \boldsymbol{measurements}$$

There will be three primary components to the variance in the recorded compression force:

- **Repeatability:** This component refers to the variance caused when the same operator measures the same spring multiple times under the same conditions.
- **Reproducibility:** This component refers to the variance caused when different operators measure the same spring under the same conditions.
- **Part-to-Part:** This component refers to the actual/ true variation in compression force between the 15 springs.

As we already know, the first two components represent the total gauge variability. By adding the third one we take as a result the total process variability which takes into account both the fluctuations between the coil springs and the variation from the measurement system.

We will construct a standard model for a gauge R&R experiment (eq. (4.1)), which has been described extensively in the 1.3.6 subunit of the 1st chapter.

$$\boldsymbol{y_{ijk}} = \boldsymbol{\mu} + \boldsymbol{P_i} + \boldsymbol{O_j} + (\boldsymbol{PO})_{\boldsymbol{ij}} + \boldsymbol{\varepsilon_{ijk}} \qquad (4.1)$$

$$\text{where} \quad \begin{cases} i = 1,2,\dots,15\ parts \\ j = 1,2,3\ operators \\ k = 1,2\ repeatitions \end{cases}$$

and the model parameters P_i, O_j, $(PO)_{ij}$ and ε_{ijk} are all independent random variables. More specifically:

- The first component $\boldsymbol{\mu}$ represents the general mean measurement (constant)
- The second component $\boldsymbol{P_i}$ corresponds to the effect of the i-th coil spring (part)
- The third component $\boldsymbol{O_j}$ corresponds to effect from the j-th operator/ appraiser
- The fourth component $(\boldsymbol{PO})_{\boldsymbol{ij}}$ stands for the interaction effect between the i-th part and the j-th operator
- The fifth component $\boldsymbol{\varepsilon_{ijk}}$ is the random error (measurement error or repeatability) from the i-th spring measured from the j-th operator in the k-th trial.
- The response/ depended variable $\boldsymbol{y_{ijk}}$ stands for the observed compression force (observed measurement) from the k-th trial on the i-th spring by the j-th operator.

The primary tools that will be utilized for performing our statistical analysis will be ANOVA tables and graphs and base on them we are going to establish our conclusions. Our analysis will be conducted at significance level $\boldsymbol{a} = \mathbf{5}\%$.

https://doi.org/10.26219/heal.aueb.6231

4.2Statistical Analysis

4.2.1 Data Visualization

Some principal descriptive statistics from our simulated data that represent the compression force on the coil springs are the following:

- Min value: 90.662
- Max value: 110.786
- Range: 20.124
- Mean: 100.717
- Standard Deviation: 4.719

We begin our statistical analysis by plotting a histogram (Fig. 18) with our data setted in a standard normal scale. On the same plot we add the curves of the standard normal distribution and the empirical density in order to compare the results. At first glance we cannot claim with absolute certainty that our measurements appear to be normally distributed. By performing Shapiro-Wilk and Lilliefors (Kolmogorov-Smirnov) tests, we receive p-values equal to 0.1147 and 0.1521 respectively. Hence, normality is approved.

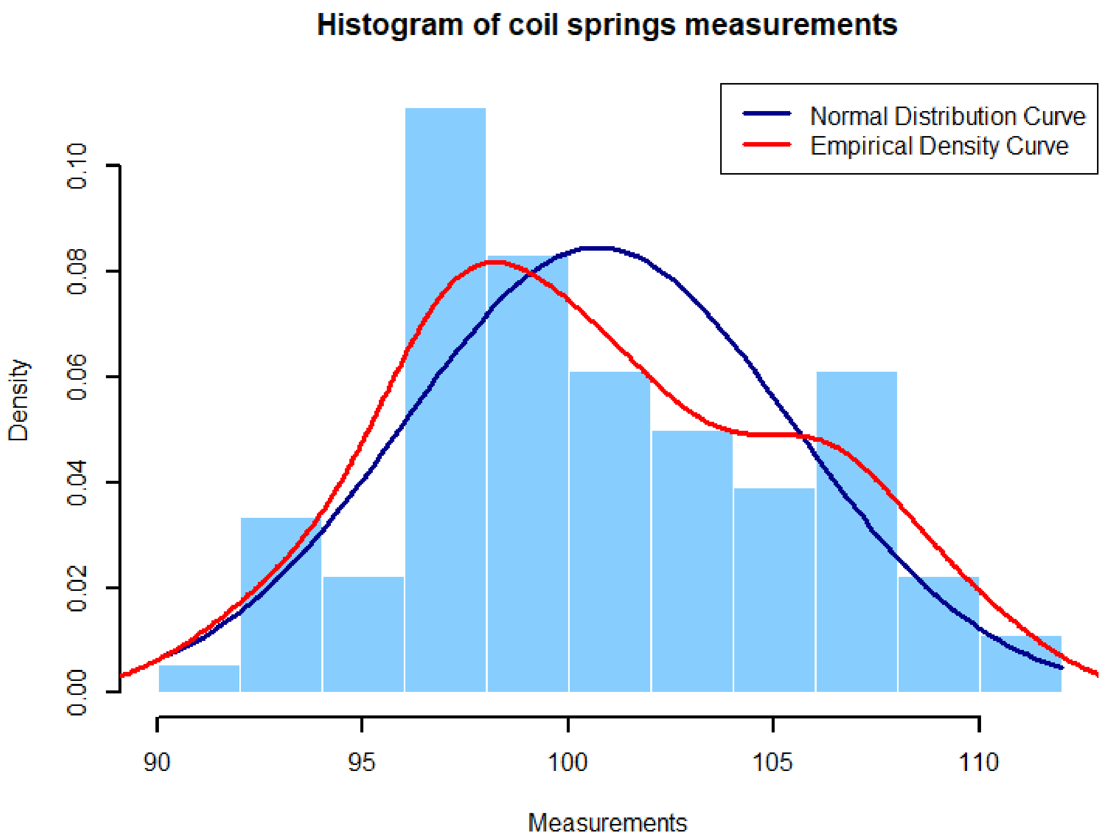

Figure 18: Histogram of coil springs measurements with the empirical density and normal distribution curves.

https://doi.org/10.26219/heal.aueb.6231

Another useful graph is a box plot grouped by operator and part. It provides us the opportunity to make a simple comparison of variation across appraisers or springs. In Fig. 19 the blue box plots correspond to the first appraiser, the purple to the second and the green to the third. In the x-axis we have the spring measurements grouped by the operators and in the y-axis we have the readings. Note of that each box plot is determined by two values since each spring is measured two times. We can observe slightly large spreads at every operator's group within the spring measurements. This is a first indication of a part-to-part variability existence.

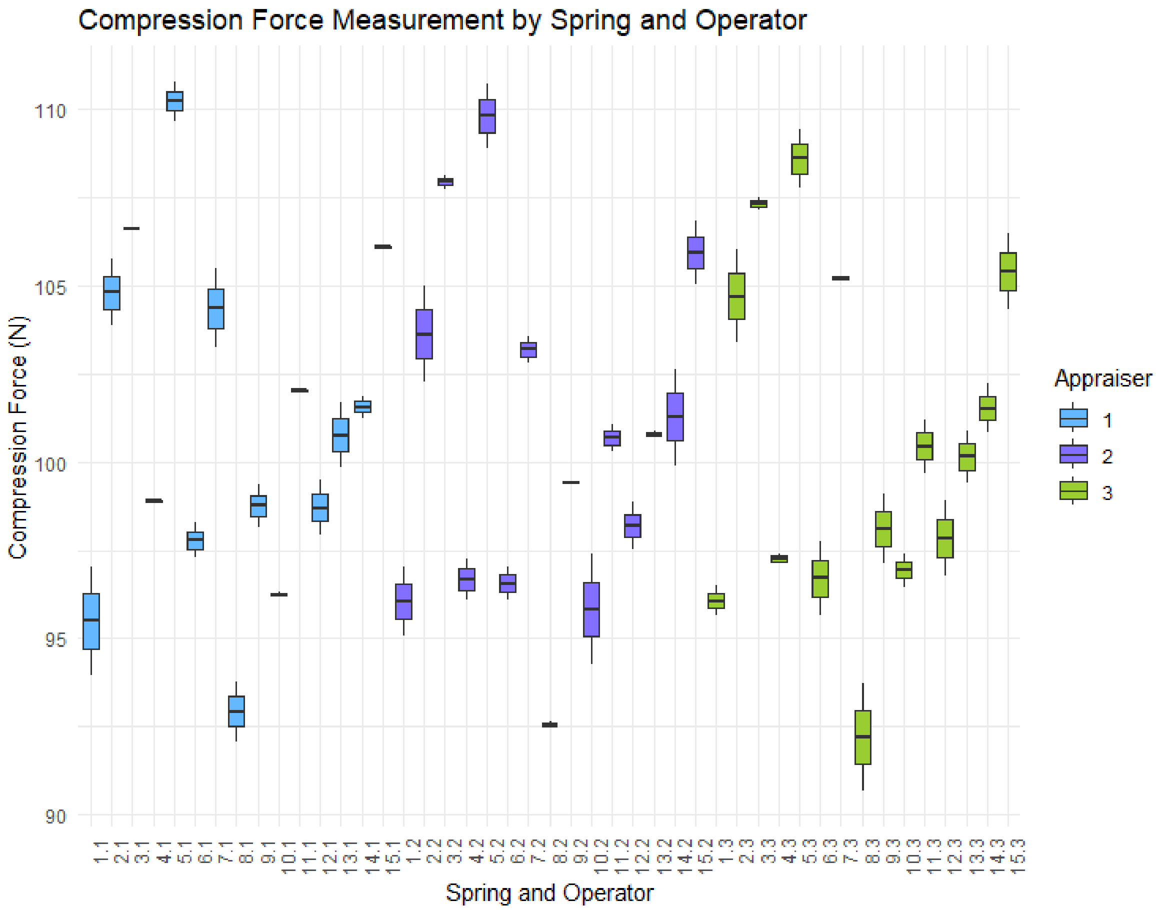

Figure 19: Box plot. grouped by operator and part.

Furthermore the box plot in Fig. 20 illustrates the measures we obtained by each appraiser. We can claim visually that Reproducibility is satisfactory because all three box plots seem to be at the same level and their means are pretty close. Thus, their outputs are rather similar.

https://doi.org/10.26219/heal.aueb.6231

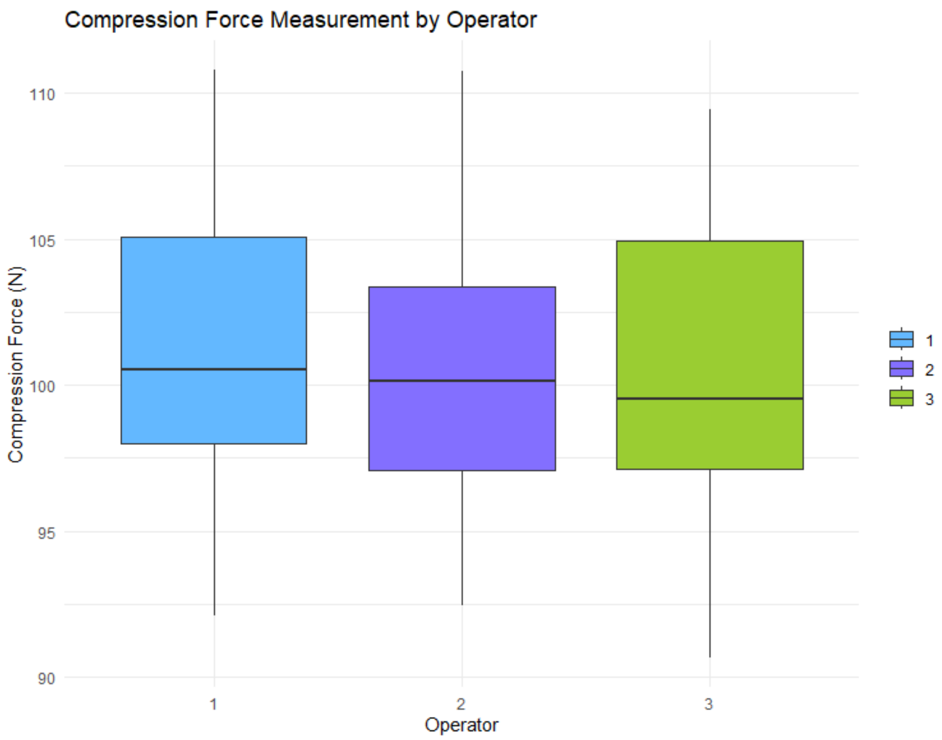

Figure 20: Spring measurements box plots grouped by operator.

4.2.2 ANOVA results

At this point we are ready to present the results from our statistical analysis. The Gauge R&R has been computed by conducting the Two-Way ANOVA method (Table 8), from which we took the following results:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Part	14	1897.265	135.519	104.927	0.000
Appraiser	2	4.076	2.038	1.578	0.218
Part:Appraiser	28	22.214	0.793	0.614	0.914
Residuals	45	58.120	1.292		

Table 8: Two-Way ANOVA table

Repeatability can be obtained directly from the ANOVA table as the residuals mean square $MS_E = \sigma^2 = 1.292$. The rest of the components of the total variability can be computed through the following formulas that have been presented in 1.3.6 subunit:

- Part-to-part variability: $\hat{\sigma}_P^2 = \frac{MS_P - MS_{P\times O}}{on} = \frac{135.519-0.793}{3\times 2} = 22.454$
- Operator/ Appraiser variability: $\hat{\sigma}_O^2 = \frac{MS_O - MS_{P\times O}}{pn} = \frac{2.038-0.793}{15\times 2} = 0.041$
- Interaction effect variability: $\hat{\sigma}_{PO}^2 = \frac{MS_{P\times O} - MS_E}{n} = \frac{0.793-1.292}{2} = -0.25$

Since a squared value cannot be negative, we assume that the interaction effect variability $\hat{\sigma}^2_{PO}$ is equal to 0. Hence,

- Reproducibility: $\sigma^2_{Reproducibilty} = \sigma^2_O + \sigma^2_{PO} = 0.041 + 0 = 0.041$
- Gauge variability: $\sigma^2_{Gauge} = \sigma^2 + \sigma^2_O + \sigma^2_{PO} = 1.292 + 0.041 + 0 = 1.333$
- Total variability: $\sigma^2_{Total} = \hat{\sigma}^2_P + \sigma^2_{Gauge} = 22.454 + 1.333 = 23.787$

Table 9 below is a useful guideline that will help us to our statistical inferences.

$\boldsymbol{GR}\&R < 10\%$	Acceptable
$\boldsymbol{GR}\&R\ between\ 10 - 30\%$	Marginal or Acceptable for some applications
$\boldsymbol{GR}\&R > 30\%$	Unacceptable
$\boldsymbol{ndc} \geq \mathbf{5}$	Acceptable
$\boldsymbol{ndc} < 5$	Unacceptable

Table 9: AIAG MSA manual acceptance guidelines

For the validation of measurement system accuracy it is important to determine the proportional contribution of the Gauge R&R to the overall variability, which ideally according to Table 8, must be lower than 10%. Also, the number of distinct categories ndc (integer number) is another value to be computed for the same purpose. These quantities are computed as follows:

$$\%\ Gauge\ R\&R\ Variability = \frac{\sigma_{Gauge}}{\sigma_{Total}} \times 100\% = 23.7\%$$

$$ndc = \frac{\sigma_P}{\sigma_{Gauge}} \times 1.41 = 6$$

From the first metric we conclude that the measurement system can be considered as marginal acceptable and the company is consulted to make improvement actions. However, the second index shows that the gauge is capable to differentiate between 6 different levels of the springs being measured, which is a satisfactory indication of its precision and resolution.

4.2.2 Further analysis and graph illustration via "`SixSigma`" package

In **R** is available the "`SixSigma`" package that is appropriate for Gauge R&R studies since it provides graphs and numeric computations collected in the function `ss.rr()`. By running this command for our experiment we obtain a text output that contains the complete model (with the interaction effect between part and appraiser), the reduced model (without the interaction effect between part and appraiser), an

https://doi.org/10.26219/heal.aueb.6231

ANOVA table and numerical details about the sources of variance. Additionally we have the corresponding plots. In the following paragraphs, we outline and explain these results.

Interpretation and commentary of the graphs:

The bar plot in Fig. 21 is a practical graph that illustrates the contribution of each component to the total variance. With a quick first glance we are able to figure out that the main source of variability comes from the coil springs (parts). Gauge variance components both satisfy the limits of approval with Repeatability to be between 10-30% and Reproducibility to be lower than 10%. The fact that Repeatability is higher than Reproducibility implies that there is potential for the equipment to be more accurate and precise. On the other hand, the low value of Reproducibility verifies that all three appraisers offer readings pretty equivalent and comparable.

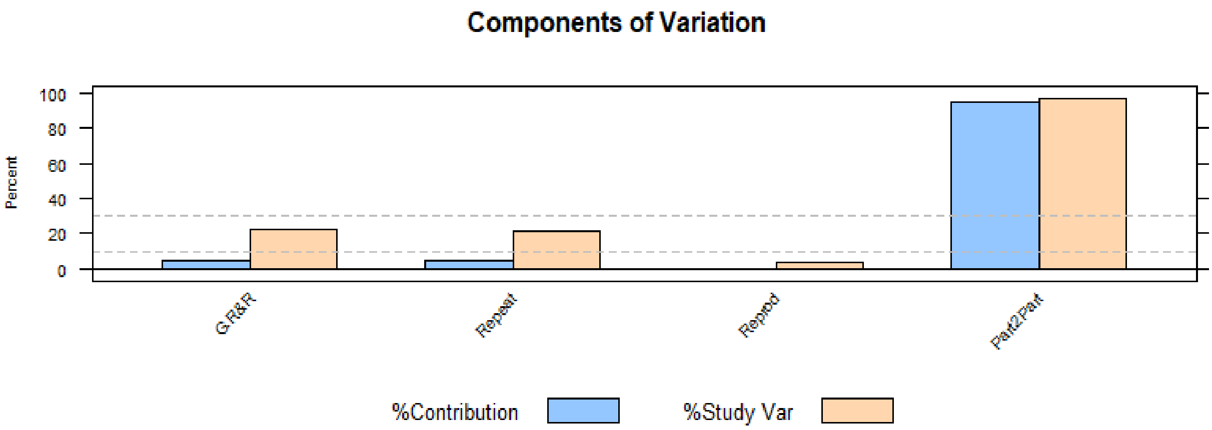

Figure 11: Bar plot of components variation at % percentage scale.

In Figure 22 we have the R control chart grouped for every appraiser. It shows the magnitude of measurement error that is directly related to the gauge capability and it consists an appropriate tool to evaluate the repeatability. We have the central line (dashed black line in the middle) and the control limits (red horizontal lines) that are adapted for R&R studies and they are computed as follows:

$$Central\ line: \bar{R}$$

$$Upper\ Limit: \bar{R} \times D_4$$

$$Lower\ Limit: \bar{R} \times D_3$$

where D_3 and D_4 are the Shewhart's constants for sample size $n = 2$ that corresponds to 2 observations/ measurements per appraiser, and $\bar{R}$ is the overall average range. For our data $\bar{R}$=1.32 and for $n = 2$ we have $D_3 = 0$ and $D_4 = 3.267$. Hence, the R-chart control limits in our experiment are:

$$Central\ line: \bar{R} = 1.32$$

$$Upper\ Limit: \bar{R} \times D_4 = 4.312$$

$$Lower\ Limit: \bar{R} \times D_3 = 0$$

At first glance we see that all points fall inside the control limits, even though some of them are borderline. Both similarities and differences exit to the pattern of the lines at each of the three portions that correspond to each appraiser. This fact indicates a mediocre repeatability, which is also verified by the second set of bar plots in Fig. 21 too.

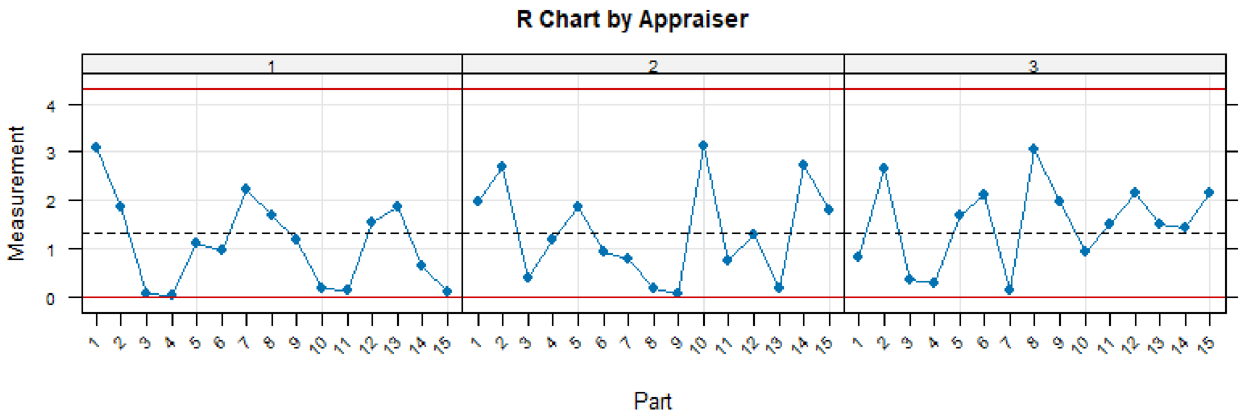

Figure 12: Range chart for each of the 3 appraisers.

In the measurement by part plot in Fig. 23, the horizontal axis stands for the i-th coil spring (where $i = 1, \cdots, 15$) and the vertical axis shows the measured values for each part. We have 6 points for each part that correspond the total number of measures for each spring ($3\ apprasers \times 2\ trials$). Remark that the points do not deviate significantly from one another. In the scenario where the measures would be obtained from perfect appraisers under ideal conditions, all the points would match for every part. Moreover, the lines are connecting the mean value of every one of the 6 measures for every coil spring.

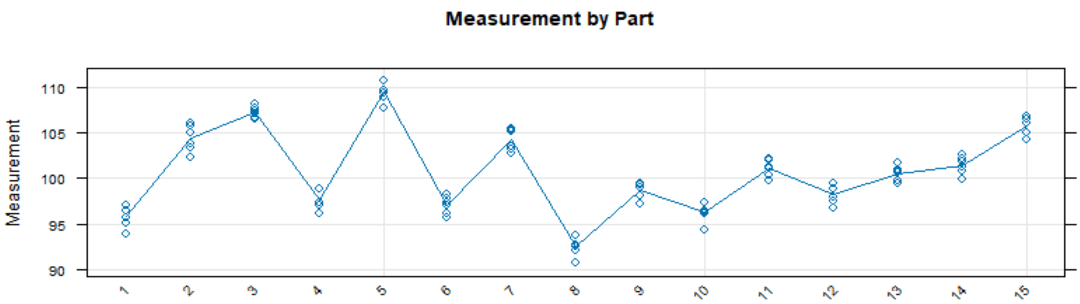

Figure 13: Measures for each coil spring linked by their mean value.

https://doi.org/10.26219/heal.aueb.6231

The plot in Fig. 24 has equivalent interpretation with the one above, with the difference that in this case the x-axis stands for the three appraisers. The blue line links the mean value from each appraiser's measurements. In the context of an ideal measurement system this line would be exactly parallel to the horizontal axis, since all the measures derived from the three appraisers would be identical. However, our experiment does not deviate a lot from this scenario, because this line is almost horizontal.

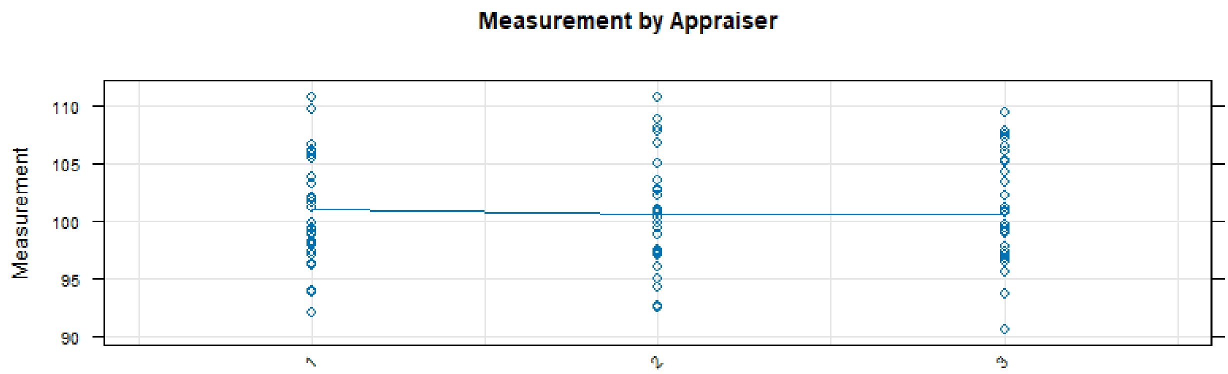

Figure 14: Measures obtained by every appraiser linked by their mean value.

Fig. 25 is a line plot that illustrates the interaction effect between the coil springs and the operators, but we can make also further inferences about the whole production process. The x-axis represent the parts and the y-axis the measurements. The dot points in the edges of the connecting lines are the average measurement for every part obtained from each operator across the two trials. The blue, orange and the green line stands for the first, second and third operator respectively. We observe that both the lines and the dot points are fairly close for every part with small deviations observed in parts 4 and 11. In other words, if the lines were highly separated, it would suggest significant differences between the operators (bad reproducibility). This suggests that some variation between appraisers exists but is not highly significant for most of the products. Also, we can see clear variation in the measurements from part to part, as indicated by the up-and-down pattern of the lines. The plot's colored lines are almost entirely parallel, implying that there is little interaction between the operators and the parts.

https://doi.org/10.26219/heal.aueb.6231

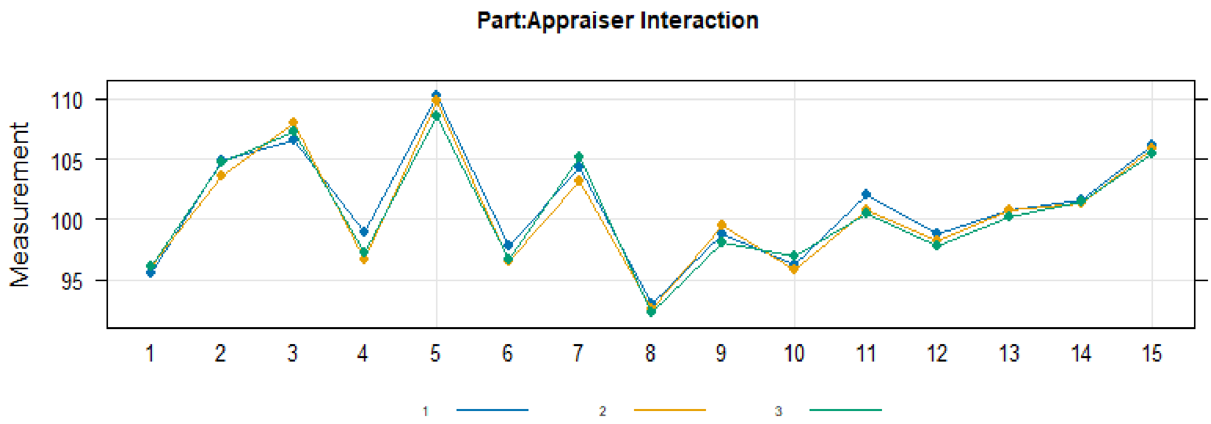

Figure 15: Line plot for the interaction effect between the coil spring and the appraiser.

Interpretation and commentary of the numerical outputs:

In Table 10 we see the contribution of each variance component to the total variance of the production process. In particular, the second column **%Contrib** provides the percentage contribution of each variance component. It is clear that the major variability comes from the fluctuations between the parts (95.19%) and not from the measurement system. Only 4.81% of the variation is due to the gauge itself and by splitting it into the usual components, we notice that Repeatability (4.68%) plays greater role than Reproducibility (0.13%).

	VarComp	%Contrib
Total Gage R&R	1.132	4.81
Repeatability	1.100	4.68
Reproducibility	0.031	0.13
Appraiser	0.031	0.13
Part-To-Part	22.403	95.19
Total Variation	23.535	100.00

Table 10: Contribution of the variance components.

In Table 11 the first column **StdDev**, refers to the standard deviation, the second, **StudyVar**, to the amount of variation each factor contributes in terms of standard deviation, and the third, **%StudyVar**, stands for the percentage of total variation due to each source, relative to the total variability. Notice that 21.93% of the total variance

 https://doi.org/10.26219/heal.aueb.6231

is attributed to the measurement system, which is not less than 10%, but it still an acceptance value. Apparently, we draw equivalent conclusions as above from Repeatability and Reproducibility.

	StdDev	StudyVar	%StudyVar
Total Gage R&R	1.064	6.383	21.93
Repeatability	1.049	6.294	21.62
Reproducibility	0.177	1.061	3.64
Appraiser	0.177	1.061	3.64
Part-To-Part	4.733	28.399	97.57
Total Variation	4.851	29.108	100.00

Table 11: Study Variance of each component.

4.3Process capability analysis results

As we explained in Chapter 2, C_p and C_{pk}, are the two principal metrics that are utilized in process performance evaluation. They provide a more comprehensive view of the process's potential to meet both internal and customer quality requirements. $\boldsymbol{C_p}$ measures the potential capability of a process assuming it is centered between the specification limits, while $\boldsymbol{C_{pk}}$ the actual capability, taking into account whether the process is centered or not. In this unit we discuss and show some numerical results from the process capability analysis which we will tie to gauge performance.

The upper and the lower natural tolerance limits, UNTL and LNTL, of the total production process represent the range within which most of the process measurements will lie and they are calculated as:

$$\boldsymbol{UNTL = \mu + 3\sigma}$$

$$\boldsymbol{LNTL = \mu - 3\sigma}$$

where μ is the overall process mean and σ is the overall process standard deviation. In our MSA experiment $\mu = 100.717$ and $\sigma = 4.719$. As a result, the natural tolerance limits for the coil spring measurements are:

https://doi.org/10.26219/heal.aueb.6231

$$UNTL = 114.873$$

$$LNTL = 86.561$$

None of our data exceed these values, which mean that we do not have outliers and all the measurements are included in the $\pm 3\sigma$ spread. Also, the process control limits in the context of this MSA experiment can be computed as:

$$Upper\ Limit: \bar{\bar{x}} + A_2 \times \bar{R}$$

$$Lower\ Limit: \bar{\bar{x}} - A_2 \times \bar{R}$$

where A_2 is the Shewhart's constant for sample size $n = 2$ that corresponds to 2 observations/ measurements per appraiser, $\bar{\bar{x}}$ is the overall mean and $\bar{R}$ is the overall average range. For our data $\bar{R}$=1.32 and $\bar{\bar{x}} = 100.717$, and for $n = 2$ we have $A_2 = 1.88$. Thus,

$$Upper\ Limit: 103.199$$

$$Lower\ Limit: 98.235$$

To continue our analysis we assume that the specification limits (SLs) that the automotive company has predefined for the compression force (in Newton scale) on the coil springs are:

$$USL = 110$$

$$LSL = 90$$

The NTLs (natural process variation) are significantly wider than the SLs (specification limits). This suggests that the process is not capable of consistently producing parts that meet the specifications. The capability indices in Table 12 confirm also the process low performance, since 0.706 and 0.656 are very low for both C_p and C_{pk} respectively.

Capability Indices	Values	Confidence Intervals
C_p	0.706	[0.603,0.810]
C_{pk}	0.656	[0.537,0.774]

Table 8: Capability indices with confidence intervals

Figure 26 is a grouped representation of the overall process capability analysis results. It provides the histogram of the measurements with the specification limits (dashed white vertical lines), a check of data normality, and the corresponding process capability indices.

https://doi.org/10.26219/heal.aueb.6231

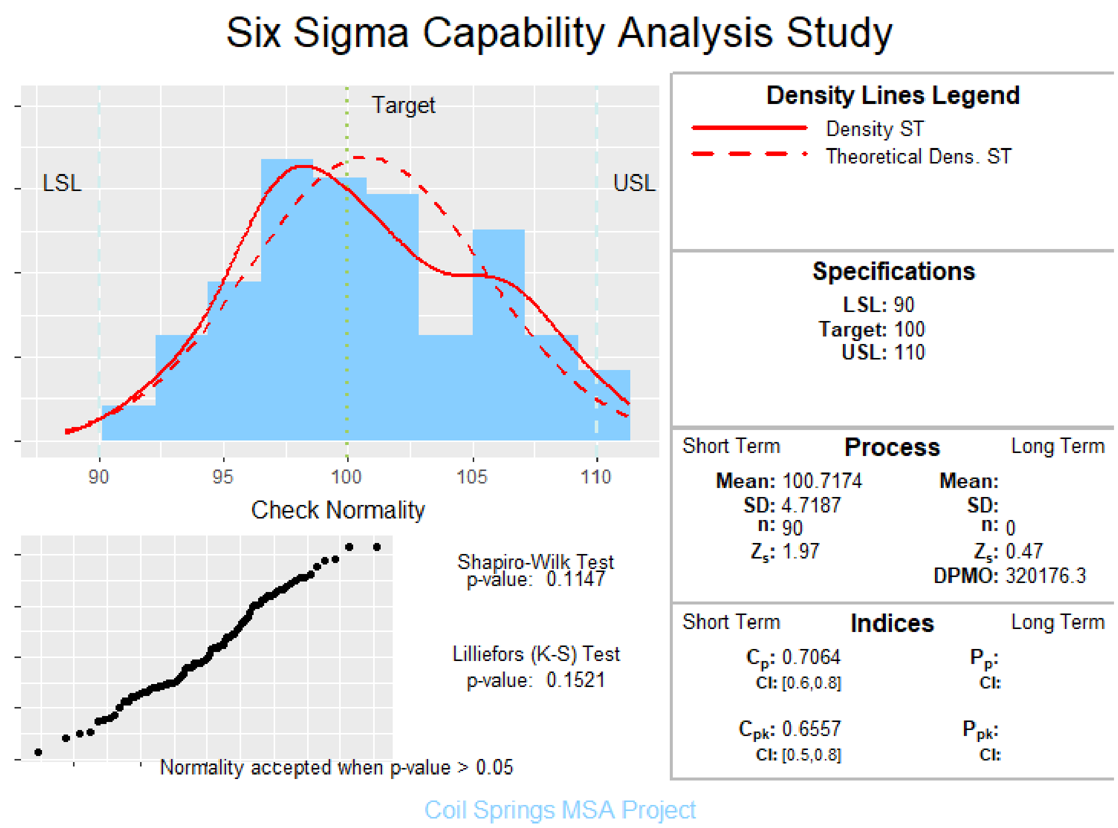

Figure 16: Process capability analysis results of the whole measurement system.

Additionally, Signal-to-Noise ratio $SNR = 6.27$, a satisfactory value higher that 5 that indicates a reliable measurement system capable to distinguish between the parts.

Fig. 27 demonstrates the process capability analysis quality attributes that we obtained from the 1st appraiser, Fig. 28 stands equivalently for the 2nd and Fig. 29 for the 3rd . The explanation is equivalent to Fig. 29. Also, Table 13 provides information about the three capability indices separately.

Appraisers	C_p	$CI(C_p)$	C_{pk}	$CI(C_{pk})$
Appraiser A	0.709	[0.528,0.890]	0.637	[0.434,0.840]
Appraiser B	0.688	[0.511,0.863]	0.648	[0.443,0.854]
Appraiser C	0.701	[0.522,0.880]	0.662	[0.454,0.870]

Table 13: Capability indices with confidence intervals by appraiser

https://doi.org/10.26219/heal.aueb.6231

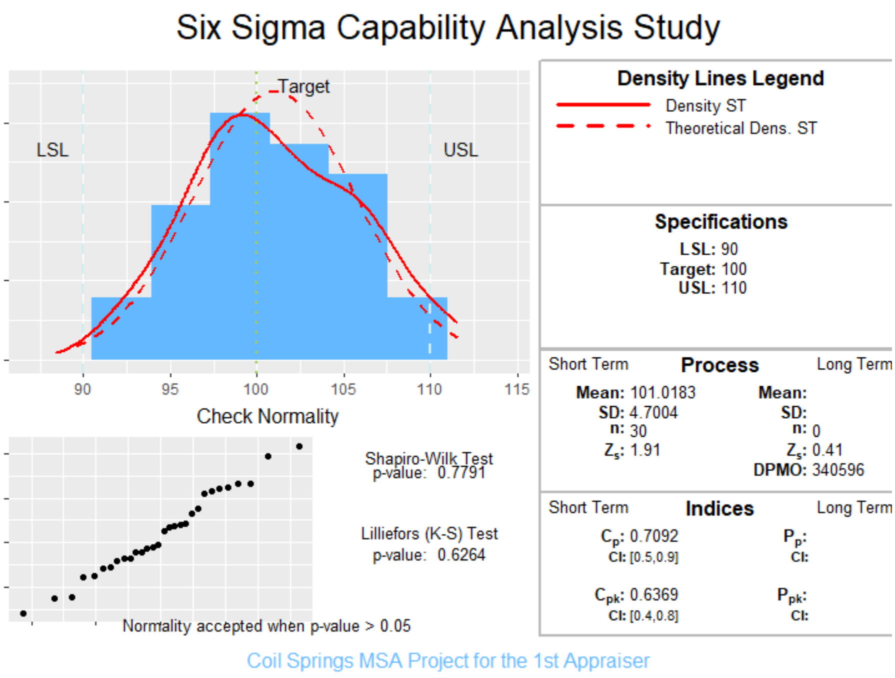

Figure 27: Process capability analysis results from the 1st appraiser.

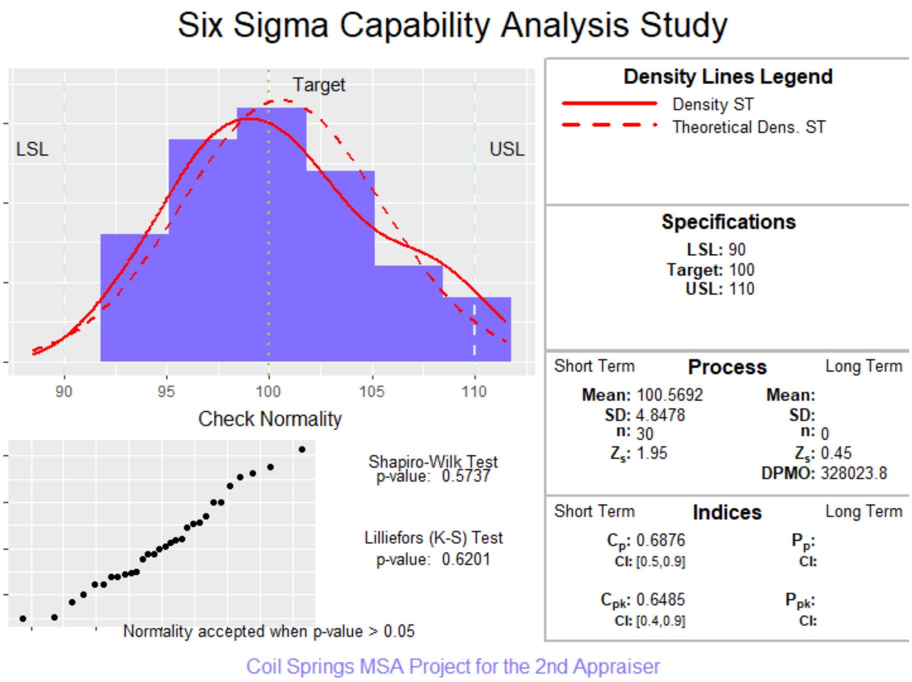

Figure 28: Process capability analysis results from the 2nd appraiser.

https://doi.org/10.26219/heal.aueb.6231

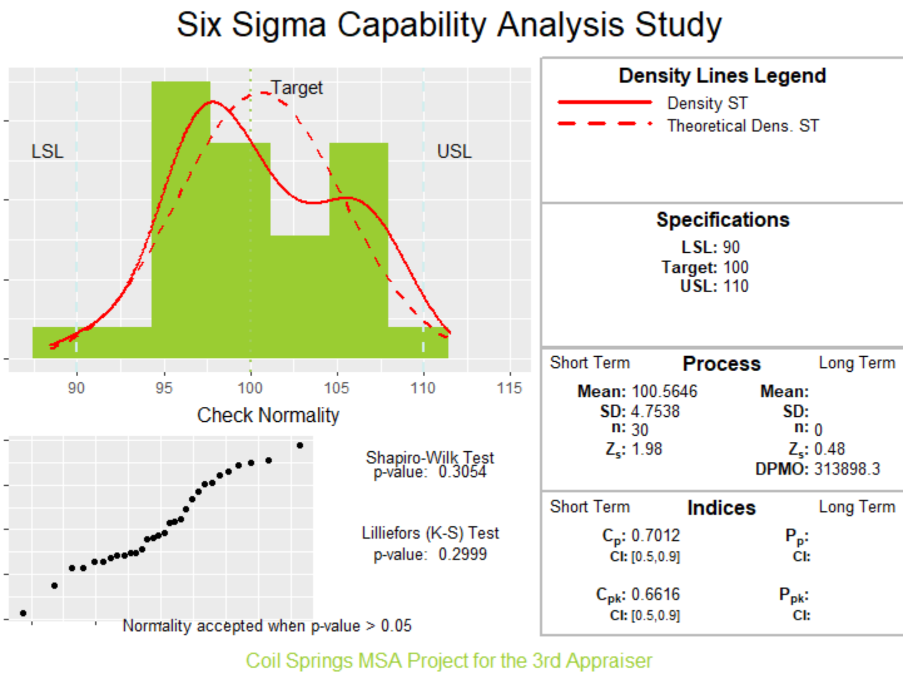

Figure 29: Process capability analysis results from the 3rd appraiser.

4.4Statistical Analysis Conclusions

In this chapter we applied a big part of the theory in a numerical experiment in order to study the performance of a theoretical measurement system with normalized data. In the context of this investigation we analyzed the gauge principal variability components, Repeatability and Reproducibility, and the discrepancy between the products, Part-to-Part variability. Additionally, we dedicated the last unit to assess the process capability analysis and draw further conclusions for our experiment.

According to the MSA indices we can approve the measurement system and claim that it provides consistent and reliable measurements. It seems that there does not exist interaction effect between the gauge and the products, i.e. the gauge is measuring equipment. However, we can make actions to improve Repeatability and make more accurate and precise each appraiser in order to optimize the performance of the whole process.

On the other hand the capability analysis indices proved that the process is poor, since the variability between the parts is really high. This fact is verified by the two-way ANOVA results (Table 15) and by Tables 17 and 18 that show the variability contribution of each component to the total process variability, where the part-to-part has the higher value compared to the others.

https://doi.org/10.26219/heal.aueb.6231

To sum up, even though the bad process performance, it is apparently that the measurement system works properly and is capable to detect the shifts and discriminate between the good and the bad parts. Hence, we consult the manufacturing department of the automotive company to initiate to initiate changes that will instantly regulate and enhance the product's manufacturing.

https://doi.org/10.26219/heal.aueb.6231

CHAPTER 5: Conclusions on Measurement System Analysis (MSA) and Gauge R&R Studies and Further Research

Measurement System Analysis (MSA) is an essential part of any quality improvement program since it assesses the performance of the measurement system used to collect data. This study is conducting at the Measure phase of a Design of Experiment (DOX) procedure because it consists a preliminary action before any further analysis. For any manufacturing company, a reliable and precise measurement system is necessary because the information obtained from it serves as the basis for choices concerning process performance, quality control, and ongoing development. The absence of this tool may lead the analysts to erroneous decision-making and consequently to incorrect recommendations and suboptimal conclusions about the process. Since the presence of variability is ubiquitous and inevitable, the practitioners utilize MSA as a manner to prevent the major error from being added by the measurement system and through this they validate that the actual performance of a process or product is not obscured. Stated differently, it is a way to investigate with clarity the potentials of the equipment, because without a solid measurement system, efforts to improve process performance may be misguided.

The major part of this thesis is dedicated to Gauge R&R studies that are included in the broader framework of MSA, and they have popular applications at a global industrial level. In Chapter 1 we introduce definitions like Repeatability and Reproducibility which are the two key sources of variance that are responsible for measurement system inaccuracy. By decomposing and evaluating these components, we are able to determine both the gauge and part-to-part variability. For these purpose, we employ several indices including P/T, SNR and DR ratios, $\%R\&R$ and ndc criteria, and we base our judgments according to the values that have been predefined by the AIAG as reference points for capability assessment. Additionally, a detailed theoretical part about ANOVA tables is provided, as it consists the integral part of this research, because they are used to compare the parts with the operators, and also check the interaction effect between them.

Meanwhile, capability indices are frequently used in conjunction with Gauge R&R studies to evaluate a process's reliability to generate results within the designated tolerance limits. The most common used are C_p and C_{pk}, that quantify the potential and the actual process capability respectively. As we remarked in the 2nd Chapter, high values of these metrics signify a robust performance, while low values imply that there may be issues with process variability or with the measuring instrument. By combining the gauge and capability indices, the analyst can form a more comprehensive and integrated form of the total production performance.

 https://doi.org/10.26219/heal.aueb.6231

In the context of further expansion, Chapter 3 focuses on the study of Actual Process Capability (AC_p) and the Observed Process Capability (OC_p) indices, consists a useful approach for assessing the true performance of the production in relation to its specified limits. On the one hand AC_p is a theoretical estimate of a process's potential performance under ideal operating conditions. On the contrary, OC_p offers a more accurate and realistic perspective, as it accounts all the potential causes that are responsible for variability existence at any level of the process. To put it more simple, AC_p shows the process's potentials under the assumptions that it is centered and controlled, while OC_p focuses on the real-world behavior. Apparently, a large discrepancy between these two quantities indicates that the process does not work properly and further investigation is necessary to detect and eliminate the odd error. Hence, both indices are helpful and widely used in MSA.

Lastly, in Chapter 4 we conducted a Gauge R&R experiment under hypothetical conditions in order to present the practical application of the aforementioned theory. The results obtained by the statistical analysis were satisfied for the measurement system, even though the process had a poor performance because of the high variability between the products. In any case, it is vital for any DOX experiment to have a reliable measuring device, since we can rely on it and fix the whole process.

Actually, any process improvement effort needs MSA and in particular Gauge R&R studies to guarantee the dependability and exactness of the data obtained by any measurement system. By ensuring that the measurement system itself is not compromising the integrity of process data, manufacturers and quality professionals make better, data-driven decisions. This culminates in an overall optimal process control management, and consequently to higher-quality products that meet client's requirements.

Furthermore, MSA and Gauge R&R studies can expand through research the applicability of their tools. For instance, many real-world processes produce non-normal data, and traditional Gauge R&R methods assume normality. Hence, it would be useful to develop this theory in order to handle skewed or multi-modal distributions and cover better the needs of the modern industry. Bayesian statistical methods are appropriate to explore the uncertainty of a measurement system, as using the prior information, they offer the opportunity of probabilistic modeling.

Also, traditional Gauge R&R studies usually are limited to manage one-dimensional characteristics in a production process. On the flip side, the contemporary measuring systems are more advanced and they are capable to make multiple measures simultaneously. So, instead of conducting separate Gauge R&R analysis for each product attribute, it would be instantly practical to develop multivariate techniques and analyze the relationships between the different variable and their contribution to measurement system variability. A recommended way to deal with high-dimensional data is via Principal Component Analysis (PCA) and Multivariate Analysis of

https://doi.org/10.26219/heal.aueb.6231

Variance (MANOVA). PCA points out the main sources of variability across all measurements and then by reducing the dimensionality, allows the gauge to focus only on critical metrics while maintaining the majority of the variability information. In the same framework, MANOVA helps us to understand how multiple response variables change concurrently across the parts and appraisers.

The extension of Gauge R&R studies in more complex environments is a very important step for the global industry. It is crucial to ensure that measuring equipment remains consistent and dependable even in the most demanding production settings.

In conclusion, Gauge R&R studies are a big part of MSA that has been developing rapidly in recent years due to its multi-level practical applications. This extremely useful tool consist a principal foundation of Statistical Process Control and Design of Experiment. Undoubtedly, the untapped opportunities for further research in this field are remarkable and we expect many innovative ideas in the immediate future.

https://doi.org/10.26219/heal.aueb.6231

https://doi.org/10.26219/heal.aueb.6231

APPENDIX

Appendix 1

According to Doraid Dalalah (2023), under the assumption that $X \sim N(\mu, \sigma^2)$ the estimated value of the standard deviation σ can be calculated by:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

The estimated value of the process capability index C_p can be expressed as:

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}}$$

We divide the index C_p by its estimator $\hat{C}_p$ and we get:

$$\frac{C_p}{\hat{C}_p} = \frac{\hat{\sigma}}{\sigma}$$

Next step is to square the above ratio and multiply it by the degrees of freedom:

$$(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2$$

As a result we have that:

$$(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 \sim \chi^2_{n-1}$$

Additionally, the critical value at a risk level of β can be determined in a way that:

$$1 - a = P\left(\hat{C}_p > C_p \middle| C_p = c\right)$$

$$1 - a = P\left(\frac{C_p}{\hat{C}_p} < \frac{C_p}{c_0} | C_p = c\right)$$

$$1 - a = P\left((n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 < (n-1)\left(\frac{C_p}{c_0}\right)^2 | C_p = c\right)$$

We proved above that the quantity $(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 \sim \chi^2_{n-1}$, so the final formula is:

$$1 - a = P\left(\chi^2 < (n-1)\left(\frac{C_p}{c_0}\right)^2 | C_p = c\right)$$

https://doi.org/10.26219/heal.aueb.6231

Therefore, the quantity $(n-1)\left(\frac{C_p}{c_0}\right)^2 = \chi^2_{\bar{a}n-1}$ is found, where $\chi^2_{\bar{a}n-1}$ is the value of χ^2 distribution at the upper α and $n-1$ degrees of freedom. Consequently, the fraction $\frac{c_0}{c}$ after some simple algebraic operations reduces to the following form:

$$\frac{c_0}{c} = \sqrt{\frac{n-1}{\chi^2_{\bar{a}n-1}}}$$

Appendix 2

(Doraid Dalalah (2023))

Given that:

$$\sigma_P = \sqrt{\left(\frac{USL-LSL}{6OC_p}\right)^2 - \sigma_g^2} \quad (A.2.1) \qquad \text{and} \qquad AC_p = \frac{USL-LSL}{6\sigma_P} \quad (A.2.2)$$

the AC_p index can be written as:

$$AC_{\boldsymbol{p}} = \frac{USL-LSL}{6\sqrt{\left(\frac{USL-LSL}{6OC_p}\right)^2 - \sigma_g^2}}$$

$$= \frac{USL-LSL}{6\sqrt{\left(\frac{USL-LSL}{6}\right)^2\left(\frac{1}{OC_p}\right)^2 - \sigma_g^2\frac{\left(\frac{USL-LSL}{6}\right)^2}{\left(\frac{USL-LSL}{6}\right)^2}}}$$

$$= \frac{USL-LSL}{6\sqrt{\left(\frac{USL-LSL}{6}\right)^2\left[\left(\frac{1}{OC_p}\right)^2 - \frac{\sigma_g^2}{\left(\frac{USL-LSL}{6}\right)^2}\right]}}$$

$$= \frac{USL-LSL}{6\left(\frac{USL-LSL}{6}\right)\sqrt{\left(\frac{1}{OC_p}\right)^2 - \left(\frac{6\sigma_g}{USL-LSL}\right)^2}}$$

$$= \frac{1}{\sqrt{\left(\frac{1}{OC_p}\right)^2 - \left(\frac{6\sigma_g}{USL-LSL}\right)^2}}$$

https://doi.org/10.26219/heal.aueb.6231

So, the final form after those mathematical operations is:

$$AC_p = \frac{1}{\sqrt{\left(\frac{1}{OC_p}\right)^2 - \left(\frac{6\sigma_g}{USL-LSL}\right)^2}} \qquad (A.2.3)$$

Appendix 3

This proof has been presented by Barbosa, G. F., G. F. Peres, and J. L. G. Hermosilla in 2014. The C_p distribution as a function of PTR is given by the formula:

$$f_{\hat{C}_p^g}(x) = 2\frac{\left(\sqrt{\frac{n-1}{2}}C_p \Big/ \sqrt{1+PTR^2{C_p}^2}\right)^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)} x^{-n} exp\left(\frac{-(n-1){C_p}^2(2x^2)^{-1}}{1+PTR^2{C_p}^2}\right)$$

$$(A.3.1)$$

where $f_{\hat{C}_p^g}$ is the distribution of the assessed C_p taking into account the presence of gauge errors. In the ideal case of a perfect gauge $PTR = 0$ and the above expression becomes:

$$f_{\hat{C}_p^g}(x) = 2\frac{\left(\sqrt{\frac{n-1}{2}}C_p\right)^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)} x^{-n} exp\left(-(n-1){C_p}^2(2x^2)^{-1}\right)$$

$$(A.3.2)$$

Recall that the term $PTR^2C_p^2$ represents the relationship of:

$$\left(\frac{6\sigma_g}{USL-LSL}\right)^2 \times \left(\frac{USL-LSL}{6\sigma_P}\right)^2$$

which, after the elimination of the term $(USL-LSL)^2$, is equal to:

$$\left(\frac{6\sigma_g}{6\sigma_P}\right)^2 \qquad (A.3.3)$$

Given that SNR is defined as $SNR = \sqrt{2}\frac{\sigma_P}{\sigma_g}$, the variance ratio in eq. $(A.3.3)$ can be also written as $2/SNR^2$ which equals to $PTR^2C_p^2$. By substituting this into eq. $(A.3.1)$, we obtain the following pdf form:

https://doi.org/10.26219/heal.aueb.6231

$$f_{\hat{C}_p^g}(x) = \frac{\left(\sqrt{\frac{n-1}{2}}\,C_p/\sqrt{1+2/SNR^2}\right)^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)} x^{-n} exp\left(\frac{-(n-1){C_p}^2(2x^2)^{-1}}{1+2/SNR^2}\right)$$

$(A.3.4)$

Naturally, in the event of a perfect gauge, i.e. $SNR \to \infty$, the above formula $(A.3.4)$ will reduce to $(A.3.2)$.

https://doi.org/10.26219/heal.aueb.6231

REFERENCES:

Books:

Montgomery, Douglas C. *Introduction to statistical quality control*. John wiley & sons, 2019.

Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery. *Design and analysis of gauge R&R studies: making decisions with confidence intervals in random and mixed ANOVA models*. Society for Industrial and Applied Mathematics, 2005.

Cano, Emilio L., Javier Martinez Moguerza, and Andres Redchuk. *Six sigma with R: Statistical engineering for process improvement*. Vol. 36. Springer Science & Business Media, 2012.

 https://doi.org/10.26219/heal.aueb.6231

Articles:

Al-Refaie, Abbas, and Nour Bata. *"Evaluating measurement and process capabilities by GR&R with four quality measures."* Measurement 43.6 (2010): 842-851.

Automotive Industry Action Group (AIAG), Measurement systems analysis, 3rd edn. (Author, Southfield, MI, 2002)

Barbosa, G. F., G. F. Peres, and J. L. G. Hermosilla. *"R&R (repeatability and reproducibility) gage study applied on gaps' measurements of aircraft assemblies made by a laser technology device."* Production Engineering 8 (2014): 477-489.

Burdick, Richard K., Connie M. Borror, and Douglas C. Montgomery. *"A review of methods for measurement systems capability analysis."* Journal of Quality Technology 35.4 (2003): 342-354.

Dalalah, Doraid. *"Evaluation of gauges in measurement systems."* Production Engineering 17.6 (2023): 929-945.

Dalalah, Doraid, and Ali Diabat. *"Repeatability and reproducibility in med labs: a procedure to measurement system analysis."* IET Science, Measurement & Technology 9.7 (2015): 826

Dalalah, Doraid, and Dania Bani Hani. *"A precision-to-tolerance ratio model for the assessment of measurements uncertainty."* Precision Engineering 44 (2016): 143-151.

Dev, Atul, Pankaj Kumar Jha, and Ved Parkash. *"Effect of automation on GRR in tactile measurement."* Mapan 33 (2018): 449-458.

Kazerouni, Afrooz Moatari. *"Design and analysis of gauge R&R studies: Making decisions based on ANOVA method."* International Journal of Industrial and Manufacturing Engineering 3.4 (2009): 335

Mader, Douglas P., Jack Prins, and Rod E. Lampe. *"The economic EVIPACT of measurement error."* Quality Engineering 11.4 (1999): 563

Majeske, Karl D., and Richard W. Andrews. *"Evaluating measurement systems and manufacturing processes using three quality measures."* Quality Engineering 15.2 (2002): 243-251.

Senvar, Ozlem, and Seniye Umit Oktay Firat. *"An overview of capability evaluation of Measurement Systems and Gauge Repeatability and Reproducibility Studies."* International Journal of Metrology and Quality Engineering 1.2 (2010): 121-127.

Woodall, William H., and Connie M. Borror. *"Some relationships between gage R&R criteria."* Quality and Reliability Engineering International 24.1 (2008): 99-106.

Zanobini, Andrea, et al. *"Repeatability and reproducibility techniques for the analysis of measurement systems."* Measurement 86 (2016): 125-132.

https://doi.org/10.26219/heal.aueb.6231

Zappa, Diego, and Laura Deldossi. *"Misclassification rates, critical values and size of the design in measurement systems capability studies."* Applied Stochastic Models in Business and Industry 25.5 (2009): 601-611.