*Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)*
*Edited by* Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P485-cd

# Safety of machinery - Proposal for a comparative method for statistical tools with examples

Heinrich Mödden

*German Machine Tool Builders' Association (VDW), Lyoner Straße 18, 60528 Frankfurt am Main, Germany.*
*E-mail: h.moedden@vdw.de*

In the Research Institute of the German Machine Tool Builders' Association (VDW), a joint discourse on statistics in machine safety took place 2022. The aim of the discourse was to define a practical basis through which a standardized application of the most important basic statistical methods and a uniform presentation of the results would be made possible.

"Uniform" because at any given time there was a whole range of research projects in the VDW Research Institute using a variety of statistical methods to evaluate their results. Crucial to the discourse was that a broad base of scientific expertise in the VDW environment was available from the outset. Likewise, results from these projects could be used as examples that represented tangible problems for the company representatives of the participating companies.

In the discourse on statistics the need for action in machine safety was identified because statistical evaluations are an essential basis for argumentation in this field. However, those responsible for machine safety practitioners in the member companies of the VDW are faced with the challenge of evaluating complex statistical issues in the context of their daily work.

With a statistical toolbox it is intended to define a practical basis by which a standardized application of the most important basic statistical methods and a uniform presentation of the results are made possible, to create a recognition value for the industry representatives in particular.

*Keywords*: Safety of machinery, descriptive statistics, inference statistics

## 1. Background

Statistical evaluations are an essential basis for argumentation in the field of machine safety. However, statistics in the mathematical sense is a broad field, and the methods used are usually not intuitive, including confusion due to terminology. In particular, practitioners responsible for machine safety in VDW member companies are faced with the challenge of evaluating such complex issues in terms of their daily work. The remedy was a joint discourse on statistics with a statistical a statistical toolbox.

The aim of the joint discourse on statistics in machine safety was to define a practical basis through which a standardised application of the most important basic statistical methods and a uniform presentation of the results would be possible. "Joint" because at any given time there are a whole series of research projects at the VDW Research Institute that draw on a wide variety of statistical methods for their evaluation of results. A decisive factor for the discourse was that a broad base of scientific expertise was available in the VDW environment from the very beginning. Likewise, results from these projects could be used as examples that addressed tangible problems for the company representatives involved.

The need for action was identified in three directions at the VDW Research Institute:

1. the results from the projects in the field of machine safety flow into standardisation or form the basis for discussions with stakeholders, e.g. occupational health and safety. For this, it must be ensured that the project results are reliably secured and resilient.

2. In many projects, test data are evaluated using statistical methods. As a rule, different software tools are used for this, in which methods may be called differently and have to be parameterised differently. There is thus a lack of certainty as to which methods are to be used how and within what limits - i.e. whether and which methods can be used in a basic way and from what limits advanced methods are justified.

3. The representatives from industry in the VDW working group on safety technology are for the most part not trained in the interpretation of statistics. In this respect, they may find it difficult to assess project results or to check them for plausibility.

In an ad-hoc working group, under the technical leadership of the external expert Dr.-Ing. Matthias Voigt from the Technical University of Dresden / Faculty of Mechanical Engineering, the statistical terms and concepts primarily relevant to machine tools were compiled in a 13-page statistical toolbox, [1].

Fig. 1 shows an example: using the histogram representation, terms such as mode, median and mean value of the sample are clearly contrasted. The Statistical Toolbox also contains recommendations on which standard methods should first be used to prepare and analyse any sample in order to justify the use of higher or more complex methods, if necessary, on this basis.

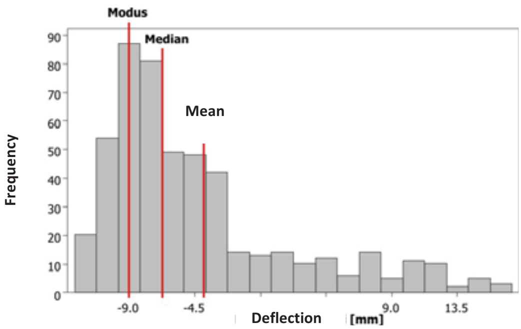


Fig. 1. Statistical location measures (general example)

### 1.1 Companies and research centres involved

VDW Research Institute, Frankfurt am Main
Technical University of Dresden, Faculty of Mechanical Engineering
Heller Machine Factory, Nürtingen
Schuler Presses, Göppingen
Trumpf, Ditzingen,
Technical University Berlin, Institute for Machine Tools,
University of Stuttgart
Chemnitz University of Technology

### 1.2 Further objectives of the tool-box

The statistical toolbox also contains recommendations on which standard methods should first be used to prepare and analyse any sample, in order to justify the use of higher or more complex methods, if necessary, only on this basis.

The statistical evaluation of test series and samples is usually carried out in a software tool. There are many of these, and their use is at the discretion of the institutions working on the project and depends, among other things, on availability (software licences) and the previous knowledge of the respective processor. Accordingly, no specifications can be made for a specific evaluation software. It is therefore all the more important to check the comparability of the different tools. For this reason, a test data set was finally developed with which these different software tools can be validated and compared. The statistical toolbox is available to members and cooperating research institutions on request. available on request.

### 1.2 Planing is all-important for collecting test data

The formulation of the global objective of the study to be conducted was most likely carried out in the application phase of a project. Here, the statistical aspect of the formulation of the objective should be briefly addressed. To this end, addressing the following questions before collecting the data is helpful and should help to define as much information and data as possible before the study.
- What is the goal of the study?
o to solve a problem
o to show possibilities
o To answer a question (e.g. Which of two alternatives is better?).
- What information is needed to achieve the aims of the study?
- Is additional information needed that cannot be obtained within the study?
- How will the results of the study be used?
Once the formulation of the study's objective has been completed, it is necessary to determine how many studies will be conducted. Answers to the following questions are helpful in this process:
- What is to be measured?
- How is it to be measured?
- How will the sample be selected?
- What other data sources are available?

Furthermore, the questions that influence the possible sample size are important:
- How long will the survey take?
- How large should or can the sample be?

In addition to the question of whether existing data material can be used, questions about quality assurance, for example, are interesting:
- Who collects the data?
- How is the quality of the collected data ensured?
Finally, before collecting data, it is necessary to define the overall scope of documentation and how it is to be documented.

### 1.3 Collection of data

The statistical data to be collected should be used to make a general statement about the population. Since it is usually impossible to test the entire population, the primary goal of data collection is to ensure the representativeness of the collected data. To this end, it is necessary to clarify in advance:
- Is there a defined population?
- Does the sample to be collected allow statements to be made about the population?
- Is the experimental design adapted to the representativeness of the samples?
- Does the experimental design influence the representativeness of the sample?

## 2. Data preparation and data analysis

An important element of many engineering investigations is the performance of the measurement system or experimental instrument used to provide the measurement results, see the example of "radar sensor" below. In any problem involving measurements, the observed variability is composed, on the one hand, of the experimental object being measured (variability of the object), and, on the other hand, of errors in the measurement (variability of the measurement equipment). Every measurement result is therefore falsified by errors [2]. The different types of errors should not be part of this paper. They can be found in [3], for example.

Before data analysis, the data from the test series must be checked for real gross measurement errors or errors within the data. Proceed with extreme caution! The aim is only to eliminate data points that are not part of the population to be analysed due to measurement and processing errors.

From the basic population (e.g. all lathes in the world of a certain type), a representative sample (see Figure 2) was determined during the collection of data, which consists of individual objects, the so-called feature carriers (e.g. three-jaw chucks). These have the respective characteristic (e.g. the tightening torque in [Nm]). These characteristics can be univariate (one-dimensional) or multivariate (combinations of several individual characteristics).

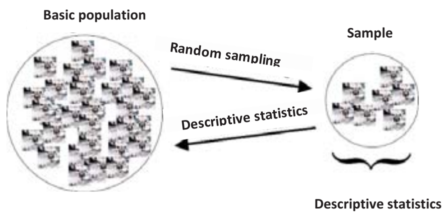


Fig. 2. Statistical structure of the study

Descriptive statistics are used to analyse the data of the sample. For this purpose, similar characteristic values can be summarised in a first step and characteristics can be represented by frequencies. The absolute frequency is the number of observed values that are identical to a certain characteristic. This sum of all absolute frequencies results in the sample size n. The relative frequency results from the quotient of the absolute frequency and the sample size. These frequencies can be represented comprehensibly in tables and diagrams. However, it is important to present all information of the data belonging to the population, otherwise the statement of the data may be falsified. In particular, the specification of the sample size is indispensable for the assessment of the statistical significance.

## 2.1 Confidence interval

The measures presented in the last section can be determined exactly for the population. With the samples of the population that are usually only available to us, the statistical measures are only point estimates and the measures themselves do not contain any information about the quality of the estimate. With the help of an interval estimate we can estimate how good the precision of the estimate is. The result of this interval estimation is a confidence interval, which is calculated with probability

$$\gamma=1-\alpha$$

contains the true value of the estimate (e.g. empirical mean). $\alpha$ is the error probability, which is often chosen as 0.05, but also 0.1, 0.01 or 0.001. The lower the confidence level is chosen, the larger is the confidence interval. The size of the confidence interval is a measure of the accuracy and certainty of the underlying point estimate. Figure 3 shows an example of a two-sided confidence interval with its left (kl) and right (kr) limits. The calculation of the confidence interval limits depends on the one hand on the theoretical sample distribution of the statistical measure itself. In the literature [4], the calculation rules are presented for various measures. Furthermore, the confidence interval limits are strongly dependent on the number of measurements in the sample.

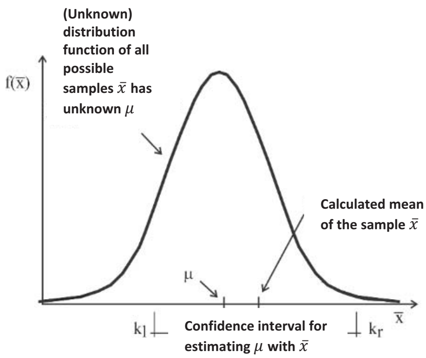


Fig. 3. Confidence interval (CI)

## 2.2 Performance of the radar sensor convincingly demonstrated

One example of the applicability of the statistical toolbox was a research project of the VDW Working Group "Grinding Technology", radar sensor for grinding wheel monitoring, see [5].

The research project deals with the measurement of grinding wheel wear on profiled grinding wheels using radar sensors. The measuring system shown in Fig. 4 was set up at the Technical University of Braunschweig. This should make it possible to measure profiled grinding wheels with fine profile surfaces over a large measuring range of several millimetres during the grinding process. In combination with a rotation angle measurement, a high-resolution 3D model of the grinding wheel is created and the actual profile is calculated.

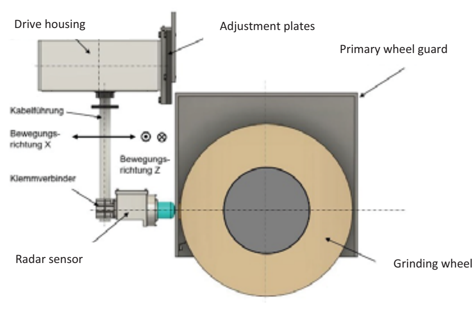


Fig. 4. Design principle of radar sensor (TU Braunschweig)

The work on programming the measurement and control programme has been completed. It enables the recording and evaluation of the radar signal as well as the path-controlled movement of the radar sensor along the grinding wheel. Measurements in the machine have shown that the scattering width of the radar signal increases as soon as the grinding wheel starts to rotate. This is probably due to vibrations in the process. The point measurements also show that the phase evaluation of the radar signal can improve the measurement results by an order of magnitude. The repeatability of the measuring system for the point measurements, including all disturbance variables, is +/-5 µm (frequency evaluation) and +/- 0.8 µm (phase evaluation). In order to determine the profile wear, point measurements were carried out on the relevant profile areas. This makes it possible to measure the progressive grinding wheel wear during the grinding wheel engagement. the grinding wheel engagement.

### 2.3 General reflection on the indication of measurement accuracies

The specification of measurement accuracies (or unavoidable inaccuracies) is relevant for all VDW research projects. There has been a recurring discussion about this for some time, because measurement results are fundamentally subject to error. The scatter (variance) from the variability of the experimental objects to be measured and the errors of the measurement equipment add up to the scatter of the results:

$$\sigma^2_{results} = \sigma^2_{object\ variability} + \sigma^2_{measure\ equipment}$$

Of interest, however, is the object variability. It is immediately apparent that the measuring equipment should be as well adjusted and calibrated as possible in order to obtain the best possible results for the object to be measured. object to be measured.

If you want to indicate the achievable measurement accuracy (or the measurement inaccuracy that cannot be avoided), it is indispensable to indicate the confidence intervals on which the set of measurements is based. In the simplest case, one orients oneself to the range of +/- one standard deviation (of the sample) around the mean value (of the sample), which is approx. 67% confidence (for the estimation of the mean). In quality assurance, however, it is quite common to apply even the range of +/- three standard deviations (the so-called “6-sigma” range). This corresponds to a confidence (here: getting good parts) of 99.8% (i.e. only 0.2% faulty parts).

### 2.4 Influence of sample size on measurement accuracy

From a sample size of $n$ = 36 the mean $\bar{x}$ and the standard deviation of the sample $s$ are derived in Figure 5 to:

$$\bar{x} = \frac{\sum x_i}{n} \rightarrow \bar{x} = 421.9998\ [mm]$$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \rightarrow s = 3.9[\mu m]$$

With this, one can estimate the population mean $\mu$ (of very many measured values) with the arithmetic mean of the sample $\bar{x}$ and the standard error standard error SE: $\left(s/\sqrt{n}\right)$.

And this as an estimate in a two-sided confidence interval (CI) with an error probability of $\propto$ (e.g. $\propto = 5\%$):

$$\mu \in \{x_{lower\ limit}, x_{upper\ limit}\}$$

$$\mu \in \left\{\bar{x} - z_{\frac{\propto}{2}} \cdot \left(s/\sqrt{n}\right), \bar{x} + z_{1-\frac{\propto}{2}} \cdot \left(s/\sqrt{n}\right)\right\}$$

The measurement accuracy in this case is for a sample size of $n$ =36 for $\bar{x} = 421{,}9998\ [mm]$ with an error probability of $\propto = 31{,}7\%$:

$z = \pm 1$ for $\varphi(z) = 0.683$ (2-sided CI)

$$x = \bar{x} \pm \Delta x = \bar{x} \pm z\frac{s}{\sqrt{n}} = \bar{x}\ \pm 3.9\ [\mu m]$$

And based on a Student-t-distribution with $df = n - 1$ =35:

$t = \pm 1.039$ for $\varphi(t, df) = 0.683$ (2-sided CI)

$$x = \bar{x} \pm \Delta x = \bar{x} \pm t\frac{s}{\sqrt{n}} = \bar{x}\ \pm 4.05\ [\mu m]$$

A minimum sample size of $n$ =3 would lead to $df = 2$:

$t = \pm 1.392$ for $\varphi(t, df) = 0.683$ (2-sided CI)

$$x = \bar{x} \pm \Delta x = \bar{x} \pm t\frac{s}{\sqrt{n}} = \bar{x}\ \pm 5.43\ [\mu m]$$

Consequently, higher measurement accuracies can only be achieved by increasing the sample size.

This conclusion is even more relevant, if the tolerable error probability is being reduced, e.g. to $\propto = 5\%$ (vs.: $\propto = 1\%$):

$z = \pm 1.96(2.58)$ for $\varphi(z) = 0.95(0.99)$ 2-sided CI

$$x = \bar{x} \pm \Delta x = \bar{x} \pm z\frac{s}{\sqrt{n}} = \bar{x}\ \pm 7.64(10.06)[\mu m]$$

And based on a Student-t-distribution with $df$ =35:

$t = \pm 2.03(2.724)$ for $\varphi(t, df) = 0.95(0.99)$ 2-sided CI

$$x = \bar{x} \pm \Delta x = \bar{x} \pm t\frac{s}{\sqrt{n}} = \bar{x}\ \pm 7.92(10.62)[\mu m]$$

Quote from [2]: "There has been an ongoing discussion about the error in measurement for some time. The mathematical treatment goes back to C. F. Gauss, which in turn is based on probability theory... Measurement results are fundamentally subject to error. The question is with what probability of occurrence an error at a certain level is still tolerable... Measurements should be as accurate as necessary and not as accurate as possible!"

### 2.5 Exemplary results of radar sensor

Fig. 5 shows the measurement results of the TU Braunschweig for a straight grinding wheel diameter of approx. 422 mm for three cases:

1. stationary grinding wheel,
2. rotating grinding wheel,
3. rotating grinding wheel with cooling lubricant

The sample size with n>30 is well within the range of the assumption of a normal distribution, so that z-values can be used for the specification of confidence intervals. Here z=1 was used for a two-sided confidence interval of 68.27 percent.

The following value ranges result for the three cases

$$\bar{x} \pm \Delta x = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

and given in [per cent] $\pm \frac{x}{\Delta x}$ :

1) 421.9998 mm$\pm$372 nm → $\pm$0.000088 [per cent],

2) 422.0003 mm$\pm$3.33 µm → $\pm$0.0079 [per cent],

3) 421.9995 mm$\pm$1.5 µm → $\pm$0.0036 [per cent].

The absolute accuracies demonstrate the suitability of the suitability of the radar sensor for the measurement object used (straight grinding wheel). grinding wheel), the relative accuracies are impressive.

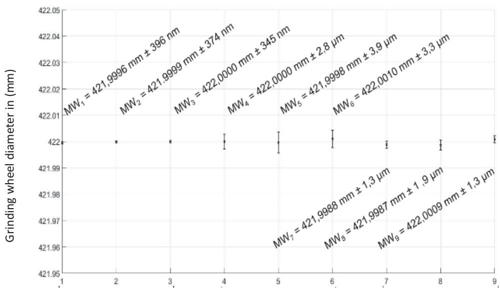


Fig. 5. Statistical measurement results

## 3. t-Test

The t-test is one of the most frequently used hypothesis tests in statistics. The one-sample t-test uses the mean value $\bar{x}$ of a sample to test whether the mean value of the population µ is different from a given value. The two-sample t-test uses the mean value of two independent samples to test how the mean values of two populations relate to each other. In principle, it is assumed for both t-tests that the respective population is normally distributed and the sample size is sufficiently large so that the central limit theorem is fulfilled. The variance equality of the two samples is also assumed for different sample sizes in the classic two-sample t-test. Other tests for evaluating the difference in means, where these boundary conditions do not have to be met, are described in the literature (e.g. [2,4]).

### 3.1 Graphical data analysis

A good overview of the frequency distribution of a sample is obtained with a histogram (Figure 6). The data are first divided into classes. Each of these classes is then represented in the histogram by rectangles that lie directly next to each other.

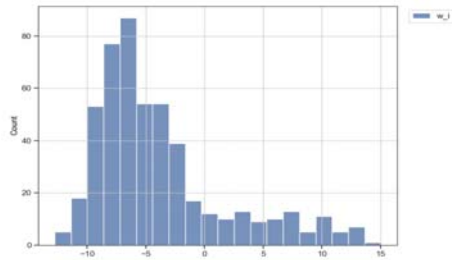

Fig. 6. Example of a histogram

The number of classes should be chosen carefully because an unfavourable number of classes can distort the frequency distribution.

### 3.2 Box-Plot

In Figure 7 the box-plot of the test data in 4.1 and Figure 8 below is show. These plots allow the representation of a lot of information of a sample [2]. For example, the left and right sides of the box show the 1st and 3rd quartiles. The T-shaped whiskers represent the min and max values of the sample. In Figure 7, the double arrow indicates the position of the arithmetic means. In this case, there is a shift of the mean to higher values, however, the spreading of the values is strongly overlapping. Box plots are a very efficient way to compare samples and the statistical measures just mentioned.

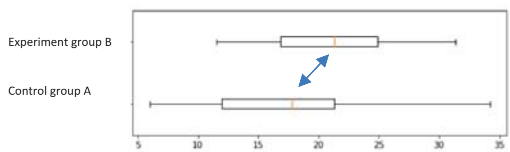


Fig. 7. Box-Plot of data sets A vs. B in 4.1 below

## 4. Interpreting the data

Finally, the results of the statistical analysis must be interpreted and the necessary conclusions about the objective discussed. In addition to the per se abstract statistical findings, the expert competence in the individual working groups of the VDW Research Institute can be consulted to assess the practical significance of the statistical results.

### 4.1 Test data set and statistical results

Due to the fact that different software tools are used in the various VDW projects, a test data set is provided here. For this data set, results with different software tools for a 2-sided t-test are listed below, as illustrated in Figure 8, which can be used to validate the respective tool to be applied.

Example from VDW study on safety of workpiece clamping:

**Control group A**: measured tightening torques in [Nm]:

{6, 6.5, 7.5, 7.8, 8.1, 10.5, 11, 11.3, 11.9, 12, 12.1, 12.4, 13, 14, 14.1, 16, 17, 17.5, 17.8, 18, 18.1, 18.5, 19.9, 20, 20.1, 20.8, 21, 21.3, 22.5, 24, 25, 25.5, 26, 29, 32, 33, 34.2}

**Experiment group B**: hypothetically obtained tightening torques in [Nm]:

{11.53, 11.68, 12.49, 13.2, 13.66, 14.58, 15.04, 15.87, 16.03, 17.41, 18.7, 16.82, 17.18, 18.6, 18.62, 19.67, 21, 21.32, 19.57, 21.8, 22.02, 22.37, 23.74, 23.82, 24.67, 24.91, 25.29, 25.44, 25.62, 26.31, 24.6, 27.79, 27.13, 27.48, 24.33, 30.2, 31.38}

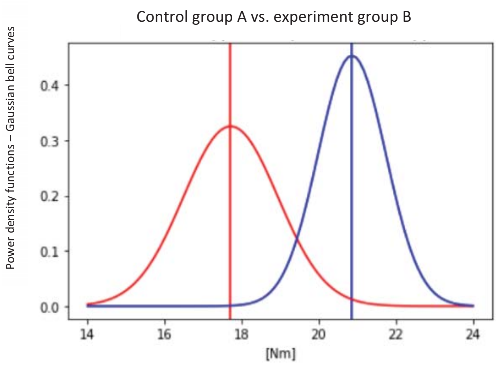


Fig. 8. Comparison of both empirical density functions

Table 1. Comparison of the test statistics and p-values

| | t-statistic | p-value |
|---|---|---|
| **Excel** | -1,99714 | 0,04114 |
| **Mathematica** | -2.01054 | 0.04112 |
| **Python (Scipy)** | -2.08355 | 0.04075 |
| **Minitab** | -2.08355 | 0.04114 |
| **MATLAB** | -2.08360 | 0.04110 |
| **SPSS** | -2.08400 | 0.04100 |
| **R** | -2.08360 | 0.04075 |

The calculated p-values in Table 1 for the 2-sided t-test in Figure 8 are almost identical for all seven software tools, p = approx. 0.04. This proves the recognisability between different software tools, although they often have completely different syntax. Thus, the researchers have the freedom to choose a preferred software tool, and to verify each other.

Here, the exemplary finding is that the result of the experimental group compared to the control group is statistically significant, i.e. the null hypothesis can be rejected if a threshold value of p=0.05 has been set beforehand.
If a threshold of p=0.01 had been set, as is usual in machine safety, the result would not be statistically significant, and the null hypothesis would hold (i.e. no significant effect in the experimental group).

### 4.2 Proposed improved impact testing of guards

In two ESREL 2022 papers of the author, the interval halving (bisection) method for determining the withstand capability (or impact resistance) of steel sheets and polycarbonate panes by means of material testing on the basis of a standardized projectile impact was explained [6,7]. On 7th of February, 2023, an expert meeting on „Design Provisions of Machine Tool Guards“ took place in Frankfurt, and besides other recommendations, the following proposal was made: “In order to determine the impact resistance, at least three values should be required for passed impact tests (instead of one value so far).” The proposal practically means that the test sequence is repeated twice more after the first result to confirm the determined value. For example, one could take the smallest of the three determined impact resistances to be on the safe side. Here follows the question, how does repeating twice increase the reliability of the result? Answer: The impact resistance determined with only one test sequence necessarily lies close by the modal value of the associated distribution function (assumed here as a Gaussian bell curve). Thus, it has only about 50% reliability (similar to the coin toss), i.e., if one were to repeat the same experiment, it could result in a slightly lower retention capability, which would call the first value into question. Now, if one requires at least three values for the impact resistance by repeating the test sequence twice, it is natural to assume an increase in reliability. However, by specifying three values, there is a limit to the increase in reliability that can be achieved. For explanation, a binomial distribution for the elementary result "test passed" with approximately p=0.5 is assumed. In this case, it is possible that a borderline impact resistance is determined three times in a row (see Moedden et al. 2017, [8]). The probability of this is P(3 times)=1/8. Accordingly, if p>=0.5 is the objective, then P(3 times)<=1/8. The counter probability is 7/8 (or >= 7/8). So, the reliability of the result increases by a doubled repetition by at least 3/8 (i.e. from approx. 4/8 with one shot to approx. 7/8 with three shots).

### 4.3 Confidence interval for Recht-Ipson curves

Luca Landi has repeatedly proposed the Recht-Ipson method in impact testing [9] as a better alternative to the conventional interval bisection method [10,11]. In this method, the value for the impact resistance is determined by means of a ballistic model using the projectile velocities before and after the impact on the basis of a model equation and a regression of the test points. From the deliberations above, it can be concluded that the principle of confidence intervals can also be applied to the impact resistances determined in this way. This is because it only concerns the scatter of the determined limiting speed (i. e. where the regression curve intersects the x-axis). As usual here too, the sample size is used to determine the standard error SE (see above), and from this the confidence intervals are connected with the corresponding t-values.

It should be noted, however, that the sample size refers to the individual value of the impact resistance determined in each case, and not to the many more values that were necessary to determine the individual impact resistances by finding appropriate regression curves.

### 4.4 Simplified statistical methods for phenomenological evidence

Usually, in VDW research projects experimental plans are designed and iteratively adapted, for instance a measurement net in a “parameter landscape” to be investigated, so that the maximum knowledge gain is achieved in the given boundaries with the available project funds. This may mean that a minimum sample size is advisable, for example in the case of the so-called “phenomenological evidence” showing strong effects (i.e. no overlap in the box plots of experimental group vs. control group).
If ambiguities arise, e.g. if two box-plots between the control group and the experimental group actually overlap (see Fig. 7), one can conduct more refined investigations and, for example, increase the sample size to clarify the effect. In this way, so-called "statistical artefacts" could also be avoided, since those can become frustrating, if “result patterns” are being derived which only are based on statistical scatter, but not on evident effects, see [6,7].

## 5. Summary and Outlook

This paper explains basic statistical concepts in a compact way in order to provide a kind of mutual checklist for all partners involved in the VDW research projects.
It is true that statistical methods have been known for a long time and these methods are partly an element of the basic education of an engineering course.
However, the fact that there are numerous literature sources (with partly different nomenclature) and diverse statistical tools (with partly different parameter menus) repeatedly leads to misunderstandings in the interpretation of research results. Even direct comparability between results from two equivalent experimental set-ups is sometimes not readily possible.
This paper aims to improve the situation in a first contribution. Further statistical topics need to considered lateron, if they turn out to be relevant for the machine tool research. There is already a follow-up discussion ongoing on the statistical approach, when it comes to costly material samples in certain research projects. In this case, the possible sample size is a priori limited due the funding frame. Then the question of the minimum necessary sample size needs to be answered.
Furtheron, the topic of "machine learning" is on everyone's lips in the VDW research institute: here, uniform statistical procedures play a similarly important role as in safety technology.

**Nomenclature**

a.) Statistical measures of the population

N = size of the basic population
μ = mean of the normally distributed population.
σ = standard deviation of the normally distributed population. Its square is the variance, the formula for which is:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

b.) Statistical measures of the sample

n = sample size, also called sample size
$\bar{x}$ = mean value of the sample (is an estimate of μ)

s = standard deviation of the sample or the measurement series. This tells how far individual data points within a sample differ from the sample mean $\bar{x}$. The square is the variance of the sample $s^2$, the formula for this is:

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

$V$ = coefficient of variation
$r$ = Pearson correlation coefficient
$\tilde{r}$ = Rank correlation coefficient according to Spearman
$R^2$= coefficient of determination

c.) Probability of error and quantiles

$\propto$ = level of significance, also probability of error (usually 5% or 1%)
$z_{1-\frac{\propto}{2}}$ =: the quantile (inverse cumulative distribution function) of the standard normal distribution (μ=0, σ=1)
$t_{1-\frac{\propto}{2}}$=: the quantile of the Student t-distribution

**Acknowledgement**

Gratitude shall be expressed to Dr.-Ing. Matthias Voigt from the Technical University of Dresden / Faculty of Mechanical Engineering, for providing a concise summary of the statistical terms and concepts, which seem to be primarily relevant to machine tool practitioners [1].

Thanks also to the experts from the VDW member companies and research institutes listed in 1.1., who contributed actively with their experience to the discourse.

Last but not least many thanks to Dr.-Ing. Alexander Broos, who enabled the funding of the statistical discourse as the head of the VDW research institute.

**References**


1. Matthias Voigt (Technical University of Dresden); *Statistischer Werkzeugkasten für Untersuchungen in der Maschinensicherheit*, *(english: Statistical toolbox for investiga- tions in machine safety)*, VDW-internal doc- ument dated May 2022.
2. Douglas C. Montgomery, George C. Runger, Norma F. Hubele; *Engineering Statistics*; Wiley-Verlag; 5th Edition; 2011; ISBN: 978-0-470-63147-8
3. Dubbel, *Taschenbuch für den Maschinenbau*; Beate Bender, Dietmar Göhlich; Springer-Verlag GmbH; 26. Auflage; 2020; https://doi.org/10.1007/978-3-662-59711-8
4. *Angewandte Statistik: Methodensammlung mit R*; Jürgen Hedderich, Lothar Sachs; Springer Spektrum; 2020; 17 Auflage; ISBN: 978-3-662-62293-3
5. Abschlussbericht zum AiF-Projekt 21110 *„Optimierung eines Radarsensormesssystems zur Schleifscheiben-verschleißmessung an profilierten Schleifscheiben“*, TU Braunschweig Max Albergt
6. Heinrich Mödden, Nils Bergström: *Design of Impact Tests for Polycarbonate Sheets and their Deterioration by Cooling Lubricants – Part 1: Models and Limitations of Measurement* (ESREL 2022 Paper No. 114)
7. Nils Bergström; Mitchel Polte, Heinrich Mödden: *Analysis of the effect of cutting fluids on the impact resistance of polycarbonate sheets by means of a hypothesis test,* (ESREL 2022 Paper No. 115)
8. Uhlmann, E., Meister, F., Mödden, H. (2017). *Probabilities in Safety of Machinery – Hidden Random Effects for the Dimensioning of Fixed and Moveable Guards.* 15th Internat. Probabilistic Workshop 2017 in Dresden (Germany)
9. Borvik T, Hopperstad OS, Berstad T, Langseth M, 2002a - *Perforation of 12mm thick steel plates by 20mm diameter projectiles with flat, hemispherical and conical noses, Part 1: Experimental study.* International Journal of Impact Engineering,27 (1) 19-35.
10. Landi, L., Amici, D. (2016). *Steel sheets impact simulation for safety guards design: problems and perspectives.* In: Proceedings of the ASME 2016 International Mechanical Engineering Congress and Exposition (IMECE), November 11 - 17,2016, Phoenix, Arizona, USA, IMECE20 16-65 181, doi: 10.1115/IMECE2016-65181.
11. Landi, L., Uhlmann, E., Hod, R., Thorn, S., Gigliotti, G., and Stecconi, A. (December 14, 2021). *"Evaluation of Testing Uncertainties for the Impact Resistance of Machine Guards."* ASME. ASME J Risk Uncertainty Part B. June