Some notes about the application of statistics

Dipl.- Ing. (FH) Klaus Hoffmann, Derikumer Weg 26, Neuss, 15.12.2020

Introduction

The roots of statistics go back to 5000 years into the past. The first censuses were already executed in ancient Egypt more than 2000 years ago. Today, the statistics are more or less omnipresent. [1]

> *„Statistics that's a subject that you could, and should, use on [a] daily basis." Statistics enables the analyst to analyse trends. „It's predicting the future." […] „In summary, instead of our students learning about the techniques of calculus, I think it would be far more significant if all of them knew what two standard deviations from the mean means." [2]*

The intended audience of this paper are engineers, practitioners, writers who are not primarily statisticians and naturally interested readers. For this reason, this paper is based on a rather heuristic approach to make the explanations easy to understand. The science of statistics deals with the collection, analysis, interpretation and presentation of data. Statistics is based on the so- called induction, i.e. on the abstracting conclusion from an observed phenomenon to gain knowledge. [3] A statistic is a numerical characteristic of the sample; a statistic estimates the corresponding population parameter and draws conclusions regarding the population [4, 5] The ways to achieve all these mentioned statistical aspects has to be seen under the influence of the economic efficiency of the statistical analysis in combination with the technological progress which can be reached. This script, will show the basic principle briefly, to realise a statistical test as one example of the application of a statistical method as part of the mathematical statistics. [6] Inferential statistics, also called statistical inference or inductive statistics, deals with the estimation of population parameters based on a sample. Statistical inference uses probabilities to determine how confident the results can be that our conclusions are correct. Effective interpretation of data is based on good procedures for producing data with thoughtful examination. The understanding of the data set must come from the user. If the user can thoroughly grasp the basics of statistical evaluation, the person can be more confident of the decisions based on the statistical method. Statistical testing is a part of a much larger concept. This concept begins with a set of assumptions upon the theory, the model of distribution. [4, 5, 7, 8] In most cases the model of distribution is based on a hypothetical distribution, which is used to approximate the empirical distribution. [3] This theory, if it has validity, will lead to predictions; which is called hypotheses. This is the role of statistics in this case, to test the hypotheses of theories to determine if they should be admitted into the accepted body of knowledge or not. This process is called " hypothesis testing." The beginnings of the hypothesis testing go back to pioneers like Francis Ysidro Edgeworth (1885) (1845-1926) and Karl Pearson (1900) (1857-1936). Both started to formulate the fundamentals of statistical testing more systematically. [3] The classical test theory based for the most part on the works by two mathematicians J. Neyman (1894-1981) und E. S. Pearson (1895–1980). In parallel with, an „original version" of the significance test by R. A. Fisher (1925) (1890-1962). [7] For the assessment of the test results the statistical significance will be used. The concept of statistical significance was first published in 1692 and 1710 by John Arbuthnot (1857-1936), therefore he is credited with "… the first use of significance tests …"[6] The examination of the statistical significance in relationship with the method of test statistic, the method of p-value for the testing hypotheses and the different uncertainties and errors will be discussed in this paper as an example for a better understanding and use of the statistical significance. [4, 5] Because a wrong interpretation the handling of the statistical significance etc. and the application of statistic in general can cause incorrect decisions with drastic consequences. [9, 10]

The continuous probability density function

Continuous metric random numerical variables have many applications to describe technical processes. The graph of a symmetric continuous probability distribution is mostly represented by a curve shaped similar to a bell. The probability is represented by the area below this curve. The curve is called the probability density function. In general, the integral calculus is needed to analyse the area underneath the curve. The course of the curve approaches the x- axis asymptotically, whereby the area below the entire curve is 1. [4, 7, 8] When using a continuous probability distribution to model probability, the distribution used is selected to fit the particular situation in the best possible way. In the case of continuous random variables, the Gauß- distribution [1, 4, 7, 8], which was developed by Robert Adrain (1808) (1775 – 1843) and Johann Carl Friedrich Gauß (1809) (1777 - 1855) [11], and especially the Student's t- distribution are applicable for a lot of issues. [4, 7, 8] The second mentioned distribution was developed by William Sealy Gosset (1908) (1876-1937) under the pseudonym Student. Each particular distribution is associated with a hypothesis testing. The graph below (Figure 1) shows the standard normal distribution, as an example, with the area between x= 1 and x= 2 shaded to represent the probability that the value of the random variable x is in the interval between one and two. [4]

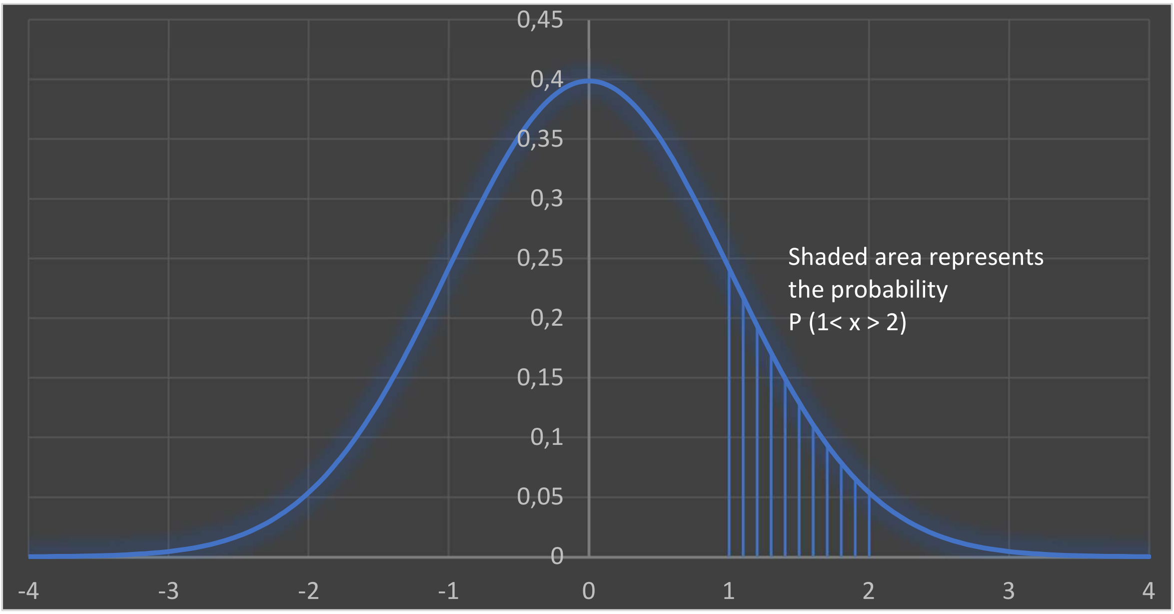


Figure 1 [4]

In the case of an infinite large size of numbers of samples, the normal distribution respectively the Gauß- distribution represents the central limit theorem [1, 4, 8] developed by Pierre-Simon (Marquis de) Laplace (1778) (1749 -1827) [11] When the number of samples is equal to or less than 30 [1, 4, 5] the central limit theorem is represented by the Student's t- distribution. The law of large numbers says, if you take more samples from any population (formular 1), then the mean of the sampling distribution (formular 2), tends to get

$$\mu = \frac{1}{N}\sum_{k=1}^{N} x_i$$

Formular 1

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$$

Formular 2

closer to the true population mean, the population standard deviation (σ) (Formular 3) (Whereby σ represents also the inflexion point of the function of Gauß- distribution on the x- axis [1, 12]) and sample standard deviation (s) (Formular 4), consequently closer to reality. [4]

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

Formular 3

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

Formular 4

Both results of formular 3 (σ) and formular 4 (s) represent the variability of the random variable (x) around the expected value (μ, $\bar{x}$). [12; 25] The law of large numbers is a natural law [13] and fundamental for the inductive statistics. [8] To obey this law, it is essential to take the following conditions into consideration:

Any different sample has to be independent of each other, they have to be selected randomly [5, 13] to ensure representativeness [7] and they must not correlate with other samples and none of them must have a pivotal influence. [13] The law of large numbers represents formally the convergence of the means. [14] It is necessary to underscore that probability density functions in general are mathematical models which embodies a set of statistical assumptions concerning the generation of sample data and similar data from a larger population, consequential density functions are approximation functions to describe the reality.

Also, valid here is the principle: A mathematical model is only as strong as its underlying assumptions. [4, 7, 8, 13, 15] The use of the central limit theorem goes further because its use makes it possible to describe the mean, the standard deviation and consequently the confidence interval of the theoretical distribution. The implications of an increase or decrease of the number of samples (n) is very important and far-reaching. The number of samples (n), influences~~d~~ the shape of the distribution curve as shown in figure 2.

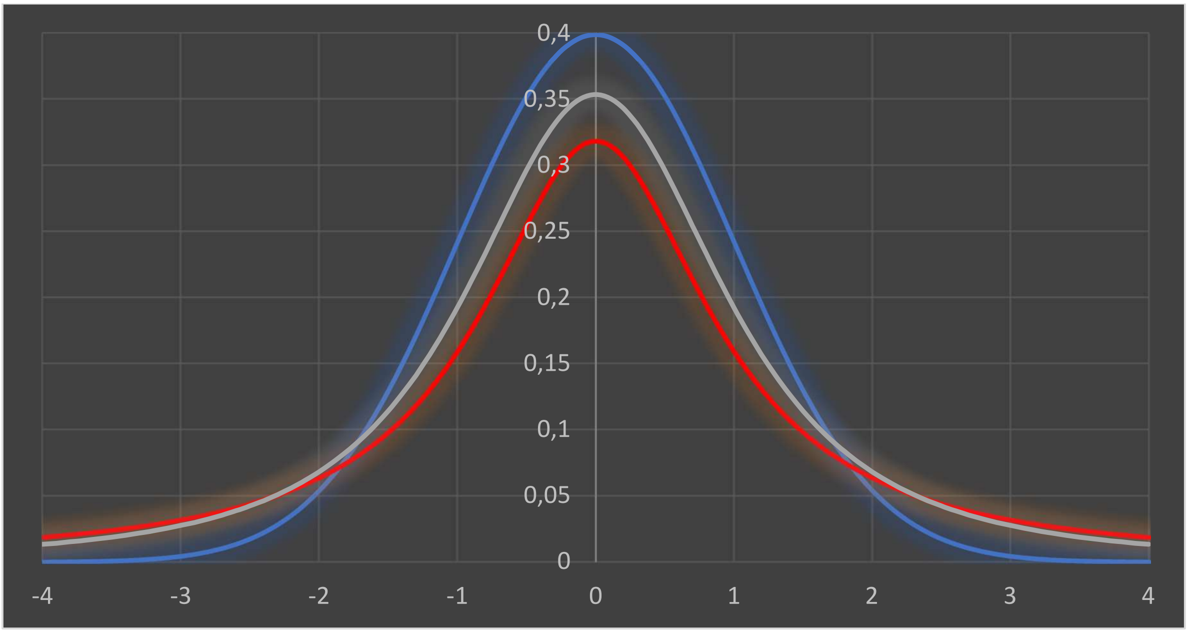


Figure 2

A smaller standard deviation (σ) causes the graph of the distribution to be "higher", this represents an increase of the maximum of the probability around the mean (μ), and "narrower" and covers less of the range of possible values than a distribution with a smaller number of samples. [4, 7] (Figure 3)

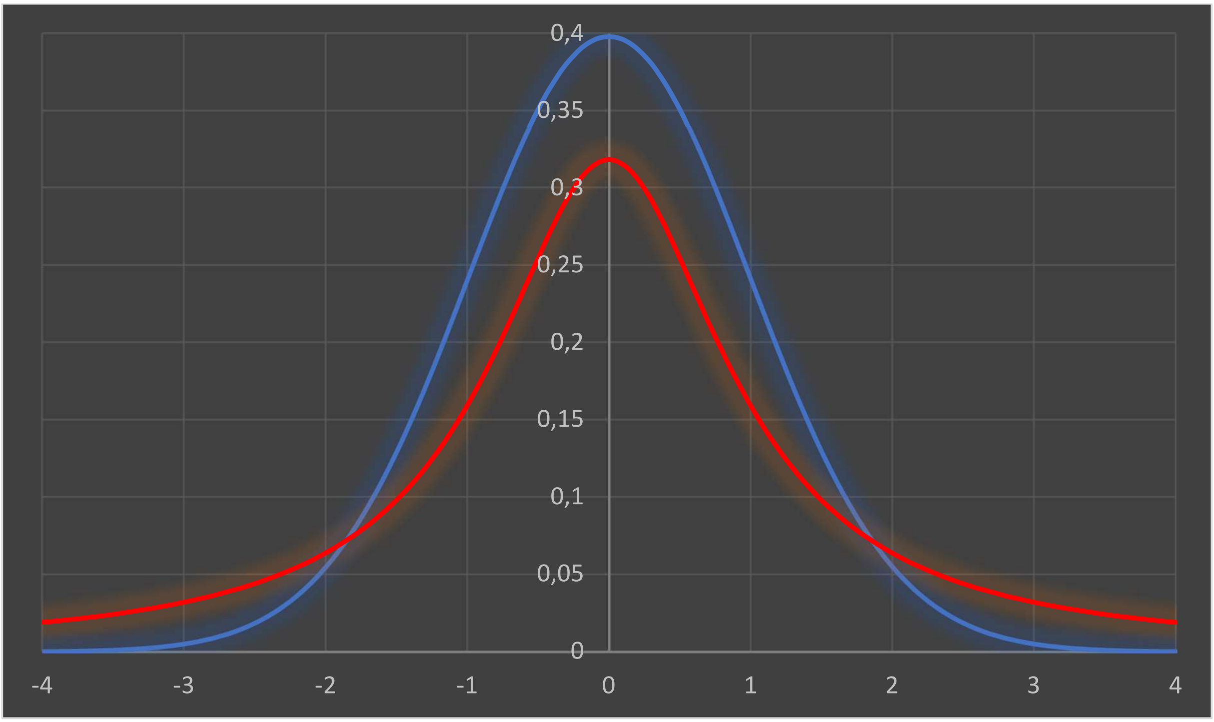


Figure 3

The confidence interval

The concept of the confidence interval based on an idea of Jerzy Neyman, it was presented to the Royal Statistical Society 1934 for the first time. [3] This concept is a part of the inferential statistics; it makes it feasible to estimate population parameters based on samples. It is a concept of estimation, with a known level of probability respectively accuracy. The determination of the confidence interval is based on the sample mean $\bar{x}$ to estimate the population mean (μ) and the standard variation (s) to estimate the population standard deviation (σ). The influence of the sample size and the knowledge of the distribution are also essential in these analyses. [8] In contrast to the mentioned point estimations, the confidence interval is another type of estimating, it is an interval estimation. [1, 4, 5, 7, 8, 12] In other words, the confidence interval contains the most plausible values of the unknown parameter of interest. [7] The concept of confidence interval is based on the same fundamentals as the hypothesis tests. Among both mentioned concepts of inferential statistics exist a relationship respectively a duality. [3, 8] It follows that confidence intervals can also be interpreted as the interval on the abscissa, within the set of date caused that the corresponding null hypothesis cannot be rejected. [3] This interval estimate for the unknown population parameters depends on:

- the desired confidence level,
- information that is known about the distribution (for example, standard distribution receptively Gauß- distribution or t- distribution),
- the sample and its size.

Concerning to the approximate normal distribution with a known population standard deviation (σ) and t- distribution, it is mostly common to define a confidence interval of 95% of the samples which will be within a confidence coefficient of 1,96 ($Z\alpha$) (normal distribution) standard deviations (1,96 σ) of the population mean μ. The figure below shows an example, with a mean of 68 and the mentioned confidence interval of 95 %.

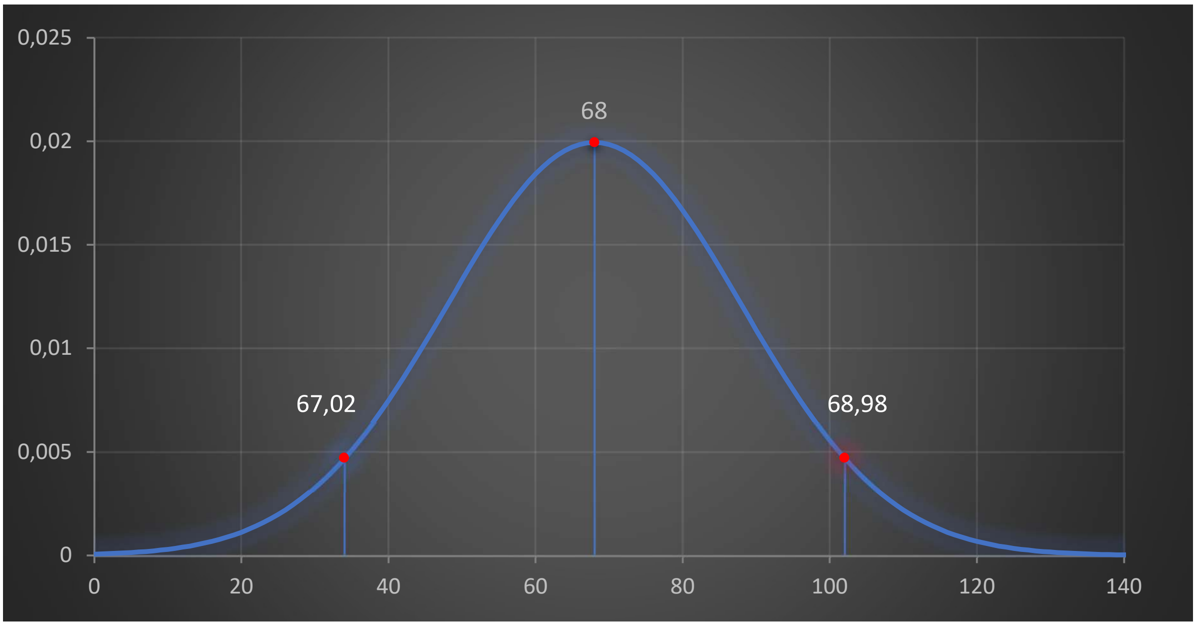


Figure 4 [4]

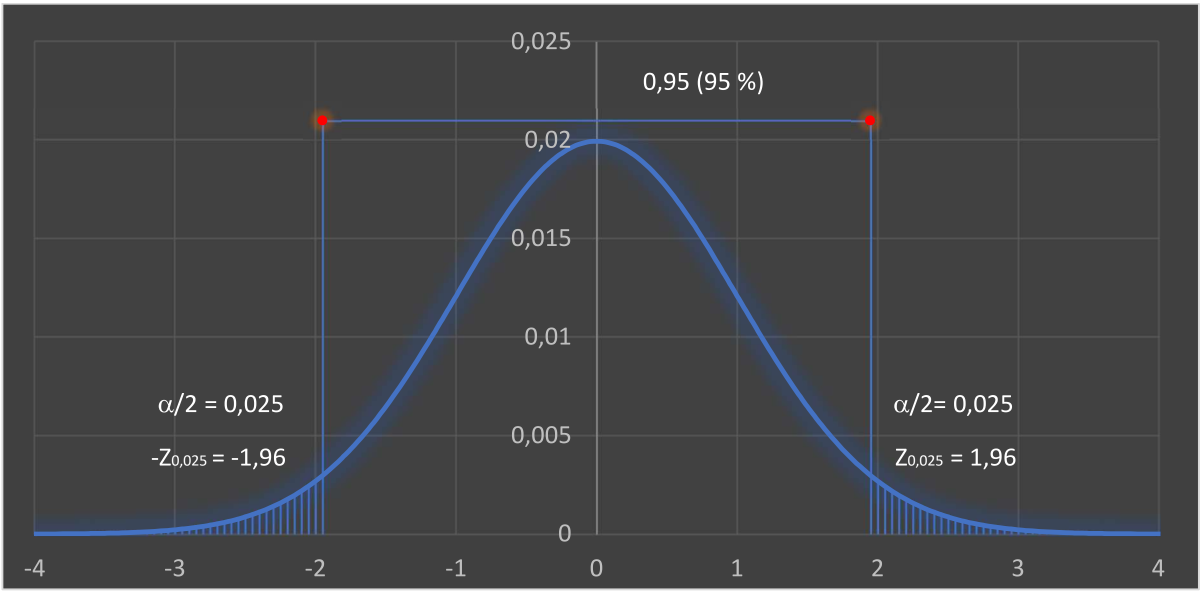


Figure 5[4]

This confidence interval implies two possibilities:

Either the interval contains the true mean μ, or our sample produced an $\bar{x}$ that is not within the interval of the true mean μ. The second possibility happens for only 5% (α– error) of the samples, because they are outside of the interval. For the confidence interval for a mean (μ) the formula is:

$$\mu = \bar{x} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

Formular 5

$Z\alpha$ is determined by the level of confidence desired by the analyst, and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Formular 6

is the standard deviation of the sampling distribution respectively standard error for means given by the central limit theorem. The Formular 6 describes the reduction of the population standard deviation, caused by the increase of samples. This reduction causes a more precise estimation of the mean of the population (μ) with the help of the mean of the random sample ($\bar{x}$). The error bound for a mean is a simplification of the description as shown below

$$EBM = Z_\alpha \frac{\sigma}{\sqrt{n}}$$

Formular 7

The confidence interval estimation can be written:

$$\mu = \bar{x} \pm EBM$$

Formular 8

The confidence level (CL) in the case of a one – side interval estimation is defined as

$$1 = \mathrm{CL} + \alpha$$

Formular 9

respectively in the case of a two – side interval estimation as follows

$$1 = \mathrm{CL} + \frac{\alpha}{2} + \frac{\alpha}{2}$$

Formular 10

$Z\alpha$ is the number of standard deviations where x lies from the mean with a certain probability. A value of $Z\alpha$ = 1,96 represents a confidence level of 95% respectively a confidence coefficient of 0,95. The most convention in economics, technical- and also most social sciences sets confidence levels at either 90 % ($Z\alpha$ = 1,64), 95% ($Z\alpha$ = 1,96) certainty is considered as probable respectively significant [7], 99% ($Z\alpha$ = 2,58) security is considered as significant respectively very significant [7] and 99,9 % ($Z\alpha$= 3.29) security is considered as highly significant. [1, 4, 5, 7, 8, 12]

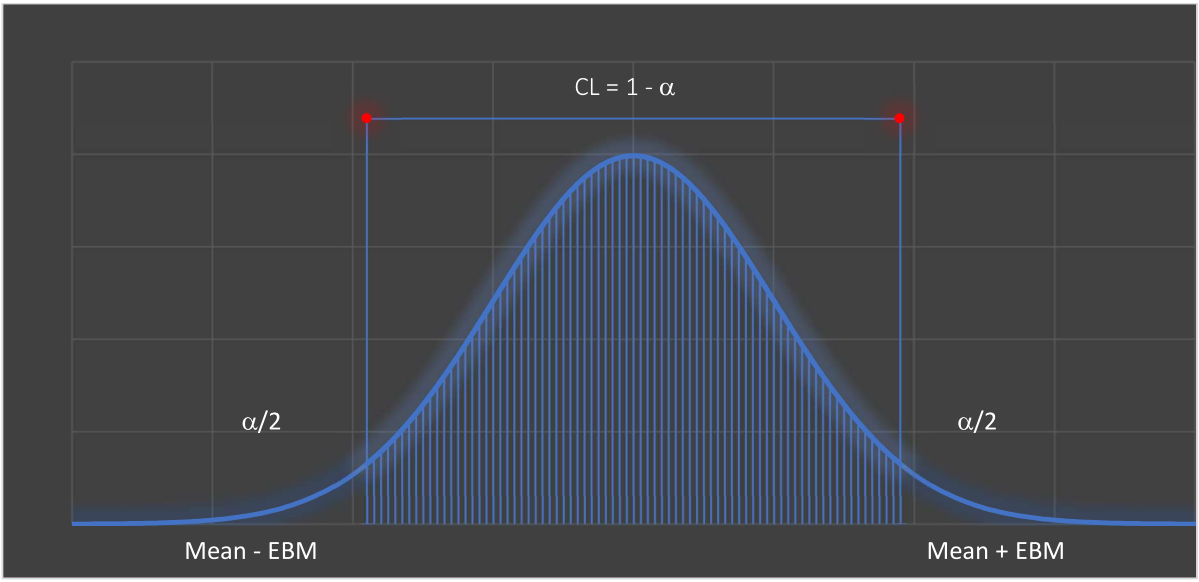


Figure 6 [4]

The confidence level and the width of the interval is a compromise. In general, a narrow confidence level will be aspired, because a wide interval provides only little information. In contrast a high level of confidence within the interval will be requested, these are the inverse tendencies. The level of confidence must be pre- set and not subject to revision as a result of the calculations!

There is absolutely nothing to guarantee that the results of the calculation will describe the reality correctly!

Furthermore, if the true mean falls outside of the interval, analyst will never know this. The analyst must always remember that we will never know the true mean. Statistics allows the analyst, with a given level of probability (confidence), to determine that the true mean is within the calculated range. This is what often called the "level of ignorance admitted". [3, 4] Confidence intervals generate, primarily, information about the accuracy of estimates. While hypothesis tests check the estimation and its compatibility with the observed sample data. [7]

Hypothesis tests

All statistical hypothesis tests and all statistical estimates - as the confidence interval - are derived via statistical models. More generally, statistical models are part of the foundation of statistical inference. [4, 16] These statistical inferences are often realised by e.g. the z- tests respectively Gauß- tests or the t-tests. To make decisions about claims by the means of statistic is called "hypothesis testing." [4] Hypothesis tests based on statistical significance are another way of expressing confidence intervals. [8, 17] A hypothesis test specifies the result of a study in respect of the hypothesis at a pre- specified level of significance. The decisions based on the using of a prechosen measure of deviation from that hypothesis the so-called "test statistic" or the "p- Value." [4] For more detailed information on this complex of themes, the author refers to bibliography. [3, 4, 5, 7, 8, 18]

The statistical significance

Statistical significance indicates the probability with which a result is produced only by chance. [19] Statisticians name differences among measurands or variables as significant, when the likelihood come about by chance is slight. If something is significant it can be statistically inferred - e.g. with hypothesis tests - that a difference actually statistically existent. Nevertheless, such a difference is not necessarily coercively really existing. Even those differences which are statistically significant can be random. [8, 20] The term significance expresses that the decision will be made objectively based on the rules of the mathematical statistics and the fact that the decision will be made under the consideration of a certain significance level α. [6] This significance level α will also be named α- error, the type I error, error of the first kind, probability of error, statistical uncertainty, measure of risk or safety threshold. [12]

- Statistical significance is a determination that the results in respect of the examined data sets are not explainable by chance alone.

- Statistical significance is the likelihood that a relationship between two or more variables is caused by something other than chance.

- Statistical significance is used to provide evidence concerning the plausibility of the null hypothesis (null hypothesis: It is a statement of no difference between a sample mean or proportion and a population mean or proportion.), in the using of statistical tests. [4] The definition of α determines the probability whether the test statistic is located in the region of rejection, under the assumption that the H_0 is true. [3] This relationship shows the importance to perform distribution tests concerning the distribution underneath H_0 and that sampling was taken representatively. [18]

Compressed into one phrase: Statistical hypothesis testing is used to determine whether the result of a data set is statistically significant. [4] Ronald Aylmer Fisher mentioned (1925), that the original intention for statistical significance was simply a tool to indicate when a result warrants further scrutiny. Statistical significance was never meant to imply scientific importance. [21]

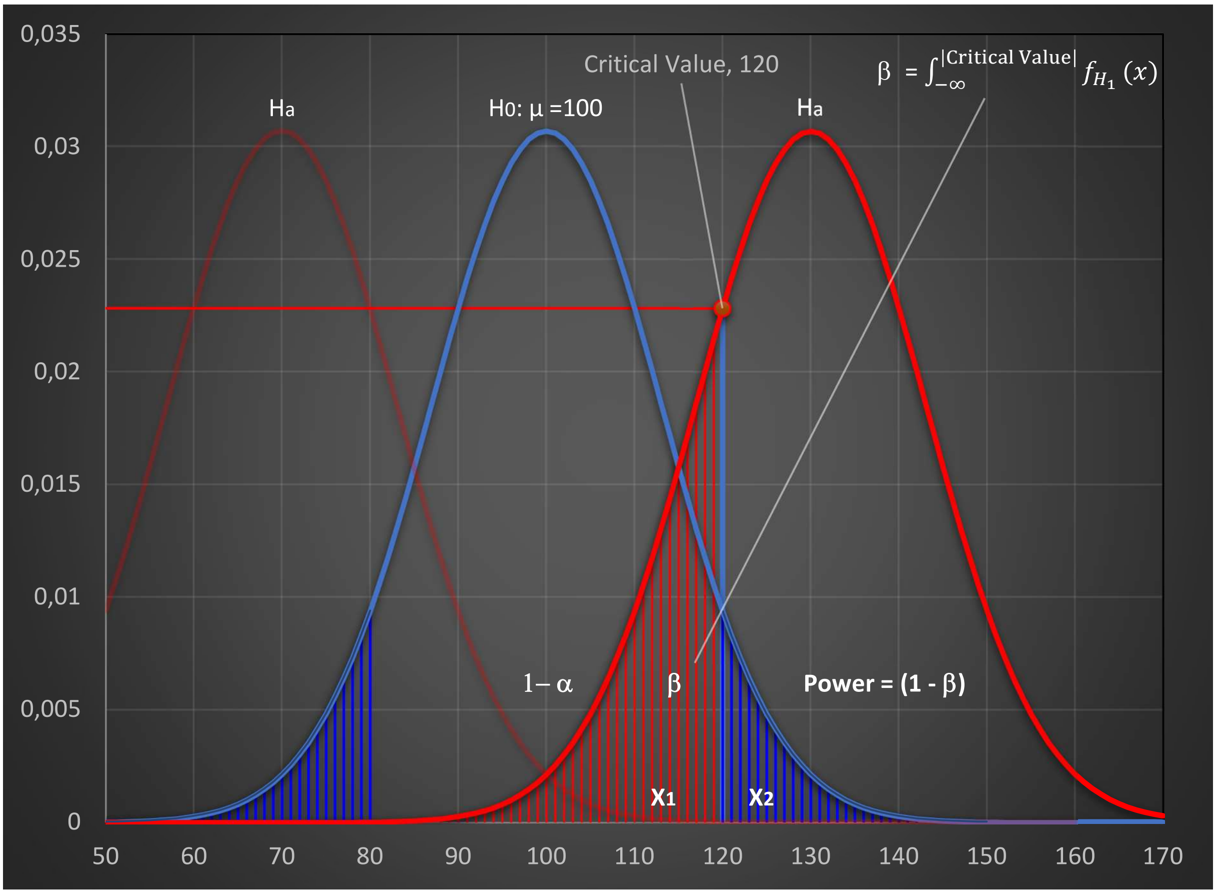
Outcomes and the type I, type II and type III errors

When the analyst performs a hypothesis test, there are four possible outcomes depending on the actual truth (or falseness) of the null hypothesis H_0 and the decision to reject or not. [4] Incorrect decisions are generally possible, because statistical tests based on samples and not on the real population, therefore systematic and random errors can occur. [8] There are two types of errors, the already mentioned the type I -error (α- error) that the decision H_1 is true, although H_0 is true. The probability of this type of error is controlled by the level of significance α. The type II error the β- error is the decision that H_0 is true, although H_1 is true. The four possible decision options are shown in Tab.1

	H_0 (null hypothesis) H_0 is true Condition of the population μ equates H_0	**H_1 (alternative hypothesis)** H_1 is true Condition of the population μ equates H_1
Decision for H_0	Right decision (1- α)- confidence interval	False decision β– error, false negative (Type II error)
Decision for H_1; H_0 will be rejected	False decision α– error, false positive (Type I error, significance level α)	Right decision (1- β)- power

[3, 8, 12] Tab. 1

Each of the errors occur with a particular probability. Generally, the error of first type is selected as the more serious error. An analyst cannot reject the null hypothesis within the probability of (1-α), the confidence interval. The determination of the significance level α determines also the probability of the error of second type (β error) [4, 7, 12] this definition of α makes the determination of β possible [3]. A reduction of α causes the enlargement of β thus both are inversely linked. The α error should be reduced as much as possible so that the correct hypothesis is not rejected. If the distance between the parameter value (mean value in this example) of the sample and the true value of the mean of the population are large, a small β error value will be produced. The probability $(1 - \beta)$ the so-called power respectively power of the test, quantifies the likelihood that a test will yield the correct result of a true alternative hypothesis being accepted. [4] On the other hand, when the distance between the mean values respectively the parameter values becomes closer, the β error value increases. Following on the interdependence of both errors, both cannot be reduced simultaneously. Based on the law of large numbers, an increase of the sample size reduces the dispersion and this causes an increase of the sample size makes it more likely to produce a significant result. [3, 7, 8, 12, 18] The α error can also be interpreted as the risk to accept an effect which does not exist. And the β error can -next to the above mentioned- also be interpreted as the risk to overlook an effect which does exist. [18]



[4,22] Figure 7

Figure 7 visualise an example of a two- sided hypothesis test. The distribution in the centre is marked H_0 and represents the distribution for the null hypotheses H_0: $\mu = 100$. There are, in fact, an infinite number of distributions from which the data could have been drawn. These infinite numbers of distributions underneath Ha are unknown. Therefore, the probability for the β- error depends also on the distributions below Ha [7] To test a hypothesis, the analyst takes a sample from the population and determines if it could have come from the hypothesized distribution with an acceptable level of significance. If the sample mean marked as X_1 is in the tail of the distribution of H_0, the analyst concludes that the probability that it could have come from the H_0 distribution is less than alpha. The analyst consequently state, "the null hypothesis cannot be accepted with (α) level of significance". The truth may be that this X_1 did come from the H_0 distribution. If this is so then we have falsely rejected a true null hypothesis and have made a Type I error. Statistics have to provide an estimate about what we know. It is also seen in Figure 7 that the sample mean could really come from the Ha distribution, but within the boundary set by the α- level. Such a case is marked as X_2. There is a probability that X_2 actually came from Ha but shows up in the range of H_0 between the two tails. This probability is the β- error. The fundamental problem is that it is only possible to set the α- error because there are an infinite number of alternative distributions. As a result, the statistician places the burden of proof on the alternative hypothesis. The example above for a test of a mean, represents the same logic which is applicable to tests of hypotheses for all statistical parameters which the analyst wants to test. [4] It is essential to point out, that decisions are always made on the basis of H_0 and its properties. [3]. Some statisticians are adopting a third type of error, Type III error, which is where the null hypothesis was correctly rejected but for the wrong reason. The problem is that there may indeed be some relationships between the variables, but these influences are not considered or stated in the hypothesis. There is no error in rejecting the null hypothesis, but the error lies in accepting an incorrect alternative hypothesis. The economist Howard Raiffa named this error Type 0 and described this error as getting the correct answer to an incorrect question. [23]

The statistical power and effect sizes

The different difficulties especially concerning the interpretation of significance and the description of the errors I and II are not only described to sensitise the reader. Next to other methods like the tests concerning the distribution, the methods of descriptive statistics, it is fundamental to analyse samples properly. In that respect the Shapiro- Wilk test and the Kolmogorov- Smirnov test or the Q-Q- plot should also be mentioned as further examples. [18] These methods were used to check normality of the data, which is often an assumption of statistical tests. To handle the interpretation of significance and the β-error the effect sizes and the statistical power are suitable and powerful statistical tools.

Statistical power

Statistical power, or the power of a hypothesis test is the probability that the test correctly rejects the null hypothesis. That is, the probability of a true positive result. It is only useful when the null hypothesis is not true "… statistical power is the probability that a test will correctly reject a false null hypothesis. Statistical power has relevance only when the null is false." [18, 24] The higher the statistical power for a given experiment, the lower the probability of making a Type II error, which is the higher the probability of detecting an effect when there is an effect. In fact, the power is precisely the inverse of the probability of a Type II error. Power analyses are normally run before a study is conducted. A prospective or a priori power analysis can be used to estimate any one of the four power parameters but is most often used to estimate required sample sizes. [24]

The power is defined as follows:

$$\text{power} = 1-\beta$$

[4, 8, 18] Formular 11

- Low Statistical Power: Large risk of committing Type II errors.

- High Statistical Power: Small risk of committing Type II errors.

It is common to design experiments with a statistical power of 80% or more. This means a 20% (β= 20%) probability of encountering a Type II error. [8, 18, 25] This is different to the 5% (α= 5%) likelihood of encountering a Type I error for the standard value for the significance level.
Statistical power is one piece of a set of four power parameters respectively statistical tools which were used to analyse and influence the power. The hereinafter mentioned influencing factors are the most important points from the statistical point of view:

- The Sample Size. The number of observations in the sample.
  The increase of sample size causes an increase of the power.

- The Significance. The significance level used in the statistical test.
  The increase of the significance level (α- level) causes an increase of the power.

- The Standard deviation. The number that is equal to the square root of the variance and measures how far data values are from their mean.
  The increase of the standard deviation causes a decrease of the power.

- The Hypothesis for a one or a two-tailed test.
  The application of the one-tailed test causes an increase of the power.

- The Effect Size (which will be discussed later). The quantified magnitude of a result present in the population. The increase of the effect sizes causes an increase of the power. [4, 18, 24]

The calculation of the β-Error respectively the power is possible, under the assumption that the distribution of the alternative hypothesis is normally distributed. Generally, the distribution below Ha (Figure 7) is often unknown and therefore, it is common practice to use specialised software. [28]
Beside the described use of statistical power, this approach offers also the analyst a way to identify the most suitable test among competing tests. [8]

Effect sizes

"Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude - not just, does a treatment affect people, but how much does it affect them." Gene V. Glass [3]
Effect sizes are the most important outcome of empirical studies. Analysts are often reminded to report effect sizes, because they are useful for three reasons:

- First, they allow researchers to present the magnitude of the reported effects in a standardized metric which can be understood regardless of the scale that was used to measure the dependent variable. Such standardized effect sizes allow researchers to communicate the practical significance of their results (what are the practical consequences of the findings for reality), instead of only reporting the statistical significance.

- Second, effect sizes allow researchers to draw meta- analytic conclusions by comparing standardized effect sizes across studies.

- Third, effect sizes from previous studies can be used when planning a new study. [25] Among other measures, "Cohens' d" is the most common. It is used in particular in the comparison between groups, like in the application of t- tests respectively z- tests. In contrast to the significance, the effect strength actually says something about the practical significance of results. A value of d = 0,1 [19] up to 0,2 [25] indicates a small effect, an effect strength from d= 0,3 [19] up to 0,5 [25] indicates a mean effect and from about d= 0,5 [19] up to 0,8 [25]- and higher - the effect sizes indicates a strong effect. Research results that are relevant to practice should have the greatest possible impact strengths in basic research, a medium or even small effects can be of interest. [19, 25]

General conclusion

Good statistical practice, is an essential component of good scientific practice, emphasized principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context and complete reporting and proper logical and quantitative understanding of what data summaries mean. [26] As Gelman and Stern famously observed, the difference between "significant" and "not significant" is not itself statistically significant. Moving beyond "statistical significance" opens researchers to the real significance of statistics, which is "the science of learning from data and of measuring, controlling, and communicating uncertainty". It is possible to summarize the recommendations in two sentences (ATOM):

- "Accept uncertainty. Be thoughtful, open and modest."

"Modelling assumptions should be sufficiently documented that independent parties can critique and replicate the work." These include relevant prior evidence, plausibility of mechanism, study design and data quality, and the real- world costs and benefits that determine what effects are scientifically important. The scientific context of studies matters should guide the interpretation. "A core problem, is that both scientists and the public confound statistics with reality. But statistical inference is a thought experiment, describing the predictive performance of models about reality. Necessarily these models are simplified relative to the complexities of actual study conduct and of the reality being studied. Statistical results must eventually mislead when they are used and communicated as if they represent this complex reality, rather than a model for it. A better use of statistics is not a problem of statistical methods. It is a problem of interpretation and communication of results." [21] „Our trust in science, like science itself, should be based on evidence, and that means that scientists have to become better communicators. They have to explain to us not just what they know but how they know it, and it means that we have to become better listeners. " [27]

This paper is based on a rather heuristic descriptions of frequently applied approaches which bases on the statistical test of the means and the use of the confidence intervals.

The text should show critical points and also should show solutions. [8] The relevance of statistics with all its applications will increase in the future. The increased application not only based on economic reasons, [29] the increase of the available data volume and the increase of the performances of statistical software will make statistics and its application more important than in the past.

> *In respect of technical systems, the use of artificial intelligence (AI), which based to a large extent on the application of statistical tools will become more important in the near future and will influence many areas of everyday life. A wide range of applications face technical as well as social challenges. [10]*

Therefore, a deeper understanding of the application of statistics is and will be increasingly important.

The use and especially the correct and conscientious interpretation of statistical results will influence our future.

List of Abbreviations

		Dimension
CL	Confidence level	-
EBM	Error bound for a population mean	-
H_0	Null hypothesis	-
H_1	Alternative hypothesis	-
N	Number of values in the population	-
s	Standard deviation of sample values	Different measurement units
x	Number of successes in the sample	-
Z	Critical value	-
$Z\alpha$	Critical value, that measures the probability of a type I error, α	-
μ	Population mean	Different measurement units
$\bar{x}$	Sample mean	Different measurement units
n	Sample size	-
σ	Population standard deviation	Different measurement units
α	Probability of a Type I error	-
β	Probability of a Type II error	-
$(1-\beta)$	Power of the test	-
$(1-\alpha)$	Level of confidence	-

Literature

[1] Thomas, Kickhahn: Statistik für Naturwissenschaftler für Dummies; 2017; Wiley-VCH Verlag GmbH

[2] Arthur Benjamin, Professor of Mathematics, on the Harvey Mudd College, TED2009, Teach statistics before calculus!, last access on 01.12.2020 under: https://www.ted.com/talks/arthur_benjamin_teach_statistics_before_calculus/transcript

[3] Herbert Stocker: Methoden der empirischen Wirtschaftsforschung; Universität Innsbruck Sommersemester 2020

[4] A. Holmes, Barbara Illowsky, Susan Dean: Introductory Business Statistics; 2018 Rice University. OpenStax Rice University 6100 Main Street MS-375 Houston, Texas 77005

[5] Heidrun Matthäus; Wolf- Gert Matthäus: Statistik und Excel; Springer Fachmedien Wiesbaden; 2016

[6] Sign test, last access on 01.12.2020 under: https://en.wikipedia.org/wiki/Sign_test#History

[7] Irasianty Frost: Statistische Testverfahren, Signifikanz und p-Werte; Springer Fachmedien Wiesbaden GmbH 2017

[8] L. Fahrmeir, u.a.: Statistik. Der Weg zur Datenanalyse, 8. Auflage; Springer- Verlag, Mai 2016

[9] Christian Honey: Statistik: Eine signifikante Geschichte, last access on 01.12.2020 under: https://www.spektrum.de/news/eine-signifikante-geschichte/1401765(Spektrum.de,7.März 2016)

[10] Michael van den Heuvel: Gute Daten, gute Ideen – dann kommt KI ins Spiel; last access on 01.12.2020 under: https://www.vdi.de/themen/kuenstliche-intelligenz-ki/gute-daten-gute-ideen-dann-kommt-ki-ins-spiel, Veröffentlichungsdatum: 07. Mai 2019

[11] Online Statistics Education: A Multimedia Course of Study (http://onlinestatbook.com/). Project Leader: David M. Lane, Rice University.

[12] Hartmut Schiefer, Felix Schiefer: Statistik für Ingenieure, Springer Fachmedien Wiesbaden GmbH 2018

[13] Terence Tao: Universelle Gesetze; Spektrum der Wissenschaft; Januar; 2014

[14] Gesetz der großen Zahlen, last access on 01.12.2020 under: https://de.wikipedia.org/wiki/Gesetz_der_gro%C3%9Fen_Zahlen

[15] Cox, D. R. (2006), Principles of Statistical Inference, Cambridge University Press.

[16] Adèr, H. J. (2008), "Modelling", in Adèr, H. J.; Mellenbergh, G. J. (eds.), Advising on Research Methods: A consultant's companion, Huizen, The Netherlands: Johannes van Kessel Publishing, pp. 271–304.

[17] John Rice: Mathematical Statistics and Data Analysis (3rd ed.). Belmont, Calif. Thomson/Brooks/Cole 2007

[18] Curt Ronniger: Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden Auflage 1.1 2014, Curt Ronniger

[19] Spektrum der Wissenschaft Kompakt; Statistik Zahlenspiel mit Mehrwert; 13.03.2017; M. Perseke, Auf diese Zahlen kommt es an; Gehirn & Geist, 11/2016

[20] Andrew Gelman und David Weakliem: Haben schöne Eltern mehr Töchter? Spektrum der Wissenschaft, Februar 2010

[21] Ronald L. Wasserstein; Allen L. Schirm & Nicole A. Lazar: Moving to a World Beyond “p < 0.05”; THE AMERICAN STATISTICIAN 2019, VOL. 73, NO. SUP1, 1-19

[22] Dr. Andreas Pfaffel; Statistik Verstehen: Teststärke: Post-hoc Power-Analyse mit G*Power; last access on 01.12.2020 under: https://www.youtube.com/watch?v=XbzLNukV4Og, 13.08.2020

[23] Martyn Shuttleworth, Lyndsay T Wilson (Nov 24, 2008). Type I Error and Type II Error. Retrieved Nov 19, 2020 from Explorable.com: last access on 01.12.2020 under: https://explorable.com/type-i-error III Error

[24] Paul D. Ellis: The Essential Guide to Effect Sizes, Cambridge University Press, 2010

[25] Daniël Lakens: Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs, Front. Psychol., 26 November 2013, last access on 01.12.2020 under: https://doi.org/10.3389/fpsyg.2013.00863

[26] Ronald L. Wasserstein, Nicole A. Lazar: The ASA Statement on p-Values: Context, Process, and Purpose; THE AMERICAN STATISTICIAN 2016, VOL. 70, NO. 2, 129–133

[27] Naomi Oreskes: Why we should trust scientists, TED Salon NY2014, Posted June 2014, last access on 01.12.2020 under: https://www.ted.com/talks/naomi_oreskes_why_we_should_trust_scientists/transcript

[28] G*Power: Statistical Power Analyses for Windows and Mac, last access on 01.12.2020 under: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html

[29] Richard Kitzberger und Ralf-Roman Schmidt: Thermodynamische Untersuchungen in Kraftwerken, VGB PowerTech 8/2011