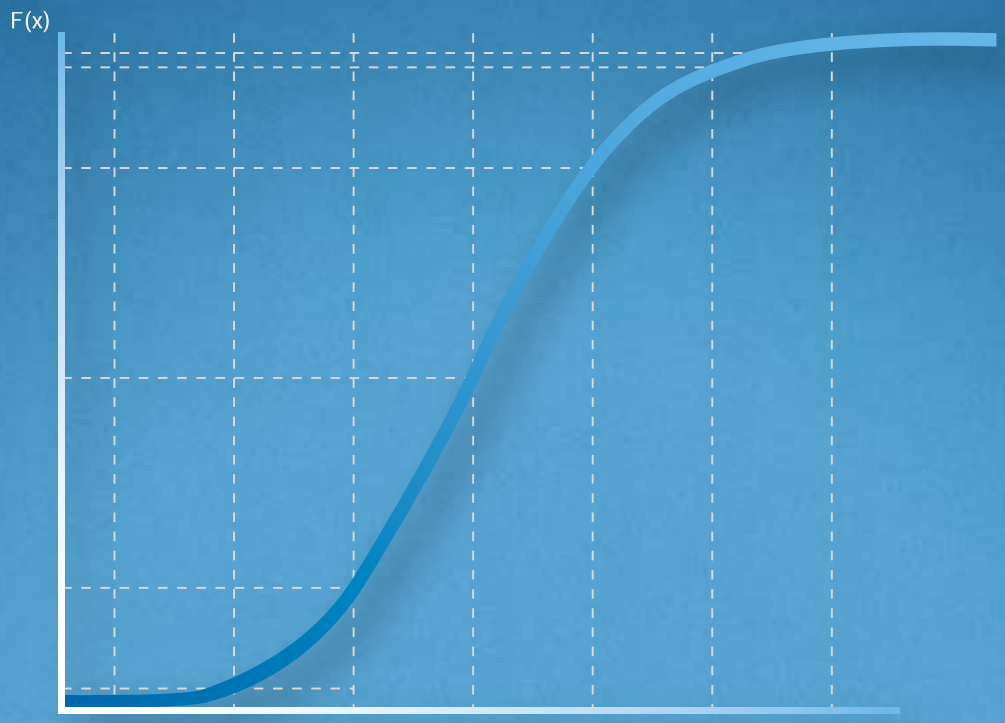



2020-04-06 - SOCOS

Quality Management in the Bosch Group | Technical Statistics

# 1. Basic Concepts of Technical Statistics
# **Continuous Characteristics**

BOSCH
Invented for life

# Quality Management in the Bosch Group
# Technical Statistics

## Booklet 1 – Basic Principles of Technical Statistics:
## Continuous Characteristics

**Edition 01.2016**

2020-04-06 - SOCOS

**Edition 01.2016**

## Preface

Before the computer was invented and comfortable statistical programs were developed, the graphical display and evaluations in the present work had to be done by hand using special forms and statistical tables.

In the interest of deep understanding, it is still sensible today to reproduce the respective procedures in detail, for example when studying statistics or completing training. For this reason, the procedures are described in such a way that allows them to be evaluated by hand without the use of a calculator.

# Table of contents

# 1 Introduction

Mathematical statistics, or statistics for short, has its origin in censuses, which serve to determine the state (Latin. "status") of a country and describe macroeconomic characteristics. According to the Duden dictionary, it is the “science of numerically recording, studying and evaluating mass occurrences”.

This definition includes two essential aspects of statistics: recording, structuring and portraying statistical data are the goals of descriptive statistics, while evaluating (analysis, interpretation) the data is the task of inductive statistics. In the media we find many different examples for the respective use of both subfields.

Some examples for the use of descriptive statistics are illustrations of

- variations over time of foreign exchanges rates or stock indices (original value charts)
- seat allocations in parliaments (pie charts)
- portions of various auto brands as part of the total number of newly registered vehicles in Germany in one year (histograms)

or data on the per capita use of dairy products in the EU countries in one year (means).

The following examples of procedures used in inductive statistics still have an unmistakable relationship to censuses:

- forecasting voting results on voting day on the basis of representative surveys,
- projecting viewer figures of television programs on the basis of viewer ratings from chosen test viewers,
- estimating the number of visitors at major (public) events,
- estimating the population of a certain animal species (total population) in an area of a known size,
- analyzing the effects of advertising campaigns on purchasing behavior in a market using the behavior of chosen test customers.

In all of these situations mentioned above, a statement on a larger collective (population) is derived from knowledge of a limited portion of individuals (sample).

This process makes use of the fact that, for many mass phenomena, the result of an individual observation (annual number of lightning strikes per square kilometer in area xy) is indeed random (and thus cannot be predicted for certain), but it can still be expressed in mathematical terms.

However, inductive statistics often also strives to determine future behavior from a momentary condition (trend), or in other words, to predict the future in a manner of speaking.

To achieve this, statistics works with mathematical models (distribution functions) that describe the properties of so-called random variables.

Misunderstandings based on the use of statistical methods are almost always the result of disregarding the relationship of the models and the assumptions associated with the methods.

For the beginner understanding statistical statements and methods is made more difficult by the many problematic aspects listed on the following page.

1. Conceptual difficulties

In everyday speech, the term “probable” is often replaced with other terms like “impossible”, “maybe”, “likely”, “pretty certain”, or “dead sure” which, in our experience, are supposed to represent a way to trust the correctness of a statement. However, depending on the person using such a term, their mood (euphoric, depressed) and the respective situation, each term might have a completely different meaning.

In contrast, statistics uses a mathematically defined “probability”, a number between zero (impossible event) and one (certain event), as a means to determine the expected “occurrence” or “non-occurrence” of an event. This said it's becoming quite obvious that there's no simple explanation for the term probability outside its "statistical realm".



2. Logical difficulties

There is a danger that the user of inductive statistics gets the impression of objective certainty when it does not actually exist. This misunderstanding is mirrored in terms like “unknown”, “random” (random variable), “probable” (“probability”) that culminate in the term “certain” (statement of certainty) in everyday speech.

It should be obvious that in reality is not possible to create a bridge between the conditions “unknown” and “certain”.

3. Transferability

The examples considered in statistical textbooks show that it is difficult to avoid applying known conditions to natural phenomena and related measured values or examples from game theory in order to illustrate random (chaotic?) behavior:

- number of lightning strikes per square kilometer of the earth’s surface and per year,
- annual rainfall per square meter,
- movement of gas molecules (Brownian motion)
- radioactive decay,
- chances of winning games of chance (dice, roulette, lottery).

Compared to such examples, phenomena studied in industrial practice hardly appear to be compatible with the terms “random” or even “chaotic”.

Despite these fundamental problems, statistical methods have firmly established themselves in industrial practice.

The present work provides the “introduction” into the series on “Quality Management in the Bosch Group, Technical Statistics”, which covers a number of special topics.

# 2 Terms

## 2.1 Characteristic

The subject of statistical considerations and calculations in industrial practice are usually continuously changing, measurable characteristics and discrete, countable characteristics of observable units. In accordance with these subgroups, the first two books in the Bosch

series “Quality Management in the Bosch Group” have the subtitles “Continuous Characteristics” (book no. 1) and “Discrete Characteristics” (book no. 2).

The continuous characteristics considered in the present Book 1 are measurable or observable properties (length, weight, temperature) of objects or results (lifetime, bursting pressure).

Although physical dimensions are always given in the form of a measured value (e.g. 48) and a unit (e.g. mm), the units of measurement are of lesser importance for statistical observations. Therefore, we can focus purely on the figures given in the examples.

Statistical analyses of the properties of continuous characteristics are used in many industrial areas, e.g. when

- studying the capacity of measuring devices and machines,
- evaluating production processes,
- applying statistical process control (SPC),
- interpreting trial data.

Inductive statistics procedures are of special interest with regard to risk analyses and, of course, where one has to work with relatively small sample sizes due to economic reasons, e.g. when conducting complex (expensive) quality inspections (destructive testing, service life tests).



## 2.2 Population

The term “population” refers to a limited or unlimited number of observable units that are to be considered concurrent within the framework of an existing statistical problem. Such observable units can, for example, come from "observations” or results from “tests” conducted under the same conditions.

Examples of finite populations are the number of

- students in a school,
- eligible voters within a state,
- television viewers who watched the final match of the last Football World Cup,
- parts in a delivery of goods,
- products manufactured within one shift at factory XY.

Examples of (theoretically) infinite populations are the number of

- points observed when rolling dice,
- results determined when repeatedly measuring against a standard of length,
- parts that a machine will create, under the assumption that it will retain its current condition for ever.

Above all, the previous examples show that a population does not always have to be real; it can also be fictional. Moreover, one can recognize that a statistical problem can sometimes be focused at a prognosis (prediction) of future results.

## 2.3 Sample



On the other hand, a sample is a real and therefore finite number of “things” or events. Examples of this are the set of

- vehicles that passed through the Engelberg Tunnel (near Leonberg) on 5.1.2015,
- results observed when rolling one die 10 times,
- results obtained when conducting 25 measurements against a standard of length,
- 50 parts made while testing a machine’s capacity.

By the way, the German term for “sample” (“Stichprobe”) originates from the practice of “piercing” grain sacks and cotton bales during quality inspection. A sample consists of one or several units that were “drawn” from a real or fictitious population according to the random principle. The number of these elements is called the sample size. The properties of the sample are supposed to represent the population. Random sampling presupposes that each element of the population is given the same chance (same probability) to be picked for the sample. In general, it is rarely possible to apply the random principle in a nearly ideal manner (flipping a coin, roulette, drawing the lottery numbers). The idea is especially problematic with regard to fictitious populations; “drawing” the sample is only possible in a figurative sense.

## 2.4 Random variable

Statistics gets around this problem by introducing the terms “random experiment” and “random variable”. “Random experiment” designates a process that can be repeated as often as desired and whose (individual) results are not predictable (e.g. rolling a die). The “random variable” represents the possible results of a random experiment (e.g. the numbers 1, 2, ..., 6). From the mathematic perspective, it is a function that can be correlated to a real number (e.g. pips on rolled dice) by a “random experiment”. The “units” or “elements”, that are “observed” as the results of a random experiment (“drawn” from the population as a sample), are the so-called “realizations” of these random variables.

If one compares these definitions with the explanations from 2.2 and 2.3, one can recognize that the terms “population” and “sample” from everyday speech and common intuition have been replaced by the mathematical values “random variable” and “realization of the random variable”, with the limitation that both of the terms mentioned always deal with real numbers.

For example, the results $x_1, x_2, \ldots, x_{10}$ of a series of 10 repeated measurements against a standard of length are the realization of a random variable $X$, which represents the population of all possible (infinitely many) measurement results against this standard.

In everyday speech, the measured values of a sample of real parts (set of measurement results) are also often called the sample.

*EXAMPLE 2.1:*

*The Figures 2.1 and 2.2 show a total population of 4941 balls each. The population according to Figure 2.1 consists of 4465 white balls and 476 black balls. We are going to consider the latter as a representative for non-conforming parts. The proportion $p'$ of the black balls is therefore $p' = \frac{476}{4941} \approx 0.096 = 9.6\%$.*

*In order to estimate the share of defects (which is the random variable of interest in this case), a sample of 490 balls is drawn. The area is illustrated through the area confined by the rectangle ($49 \cdot 10$ balls). The sample contains 48 black balls whose proportion in the sample is therefore $p = \frac{48}{490} \approx 0.098 = 9.8\%$.*

*In this case, the sample provides a relatively good estimation of the share of defects in the population.*



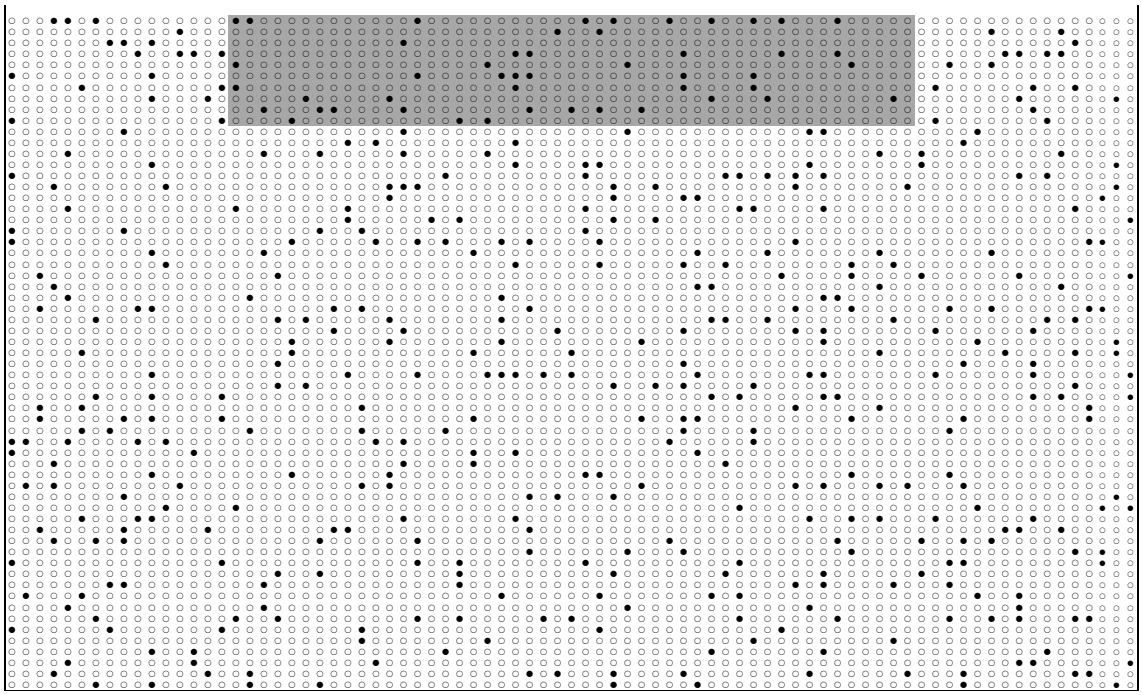


**Fig. 2.1:** Uniformly mixed population with approx. 10% share of non-conforming parts. The share of defects is estimated quite well using the sample (rectangle).

*The example in Figure 2.2 shows how such an estimation can lead to erroneous conclusions. Here, the population is not uniformly mixed. The share of defects decreases from bottom to top. This situation might occur, for example, if the share of defects on a production line constantly decreases within a limited period of time and the parts produced in the production series are placed into a container accordingly. In the present example, 499 balls from the population are black and 4442 are white.*

*There is therefore a share of defects $p' = \frac{499}{4941} \approx 0.101 = 10.1\%$ that is hardly distinguishable from Figure 2.1. However, it is obvious that the share of defects in sample $p = \frac{8}{490} \approx 0.0163 = 1.63\%$ leads to the wrong conclusion.*

*The difference in both sample results and the resulting estimation of proportions is given by the fact that the randomness principle was not adhered to (in both cases). Not every part of the population had the same chance to be drawn for the sample.*

*In reality, we live in a world where we do not know the share of errors among the total population and have to rely exclusively on the significance of the sample.*

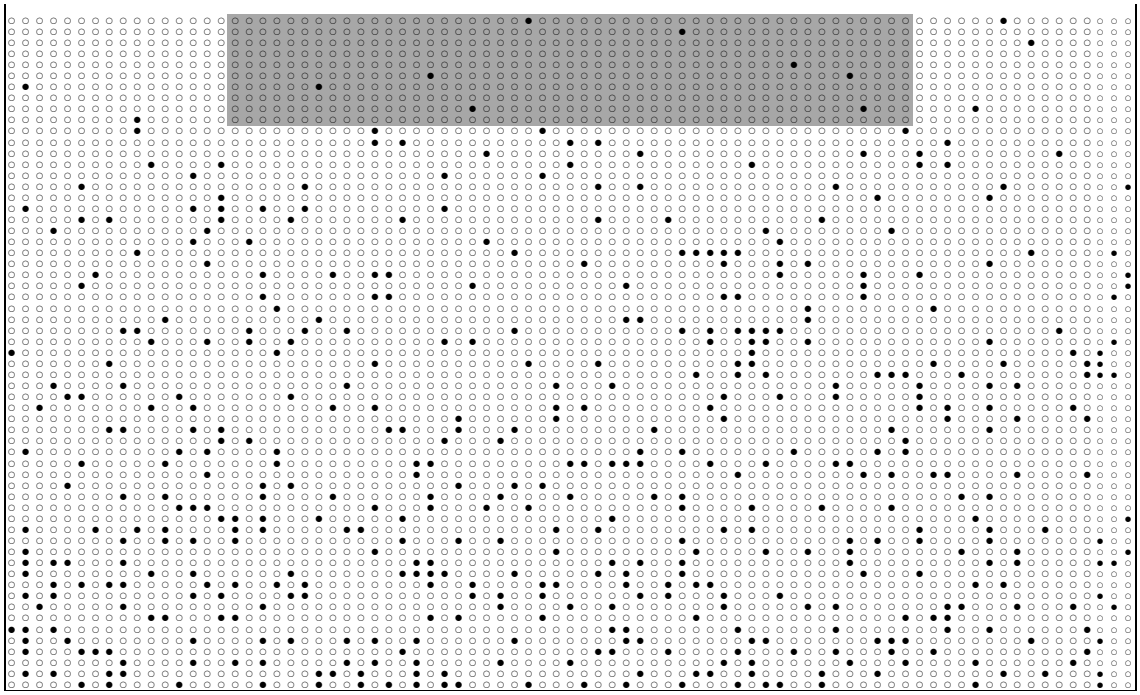


**Fig. 2.2:** Non-uniformly mixed population with approx. 10% share of non-conforming parts. The share of defects is estimated incorrectly using the sample (rectangle).

## 2.5 Probability

Mathematical probability is a number that is closely related to the results of a random experiment.

Its classic definition is derived from game theory, which is the origin for the theory of probability and statistics.

Flipping a coin is a random experiment that is commonly examined in statistical textbooks. It is generally accepted that the results from flipping a coin are impossible to predict and that, based on the (sufficient) symmetry of the coin, the results of “heads” and “tails” are just as probable.

According to the classic definition, the mathematical probability $P(A)$ of result A in a random experiment is given by

$$P(A) = \frac{g}{m}.$$

Where

$g$ is the number of (favorable) cases where $A$ occurs,

and

$m$ the number of all possible cases

in the experiment at hand. With regard to flipping a coin, this means:

the probability of the event "head" is $P(\text{"head"}) = \frac{1}{2} = 0.5 = 50\%$.

The number $g$ of cases where "tails" occurs (favorable result for the player who predicted "tails") is equal to 1, the number of all possible outcomes of the (one-time) coin roll is equal to 2 ("heads" and "tails"). The "symmetry of probability" obviously results from the symmetry of the coin. Both results are just as probable: $P(\text{"heads"}) = P(\text{"tails"}) = \frac{1}{2}$.

If one considers a random experiment with a finite number of possible results whose probability of occurrence cannot be directly derived from an examination of the symmetry, there is (at least theoretically) the possibility of repeating the experiment many times and determining the relative frequencies (see Chapter 5.3) of the occurrence of each result. One can then define the probability for a certain result as the limit, the event's relative frequency (a number between zero and one) is approaching for a large (towards infinity) number of repetitions of the random experiment.

For example, for repeated coin flips, one can determine that the relative frequencies (e.g. the number of "heads" divided by the total number of flips) for either event are approaching the value 0.5 (Figure 2.3) with increasing number of flips of the coin. Generally, this phenomenon is called the "law of large numbers".

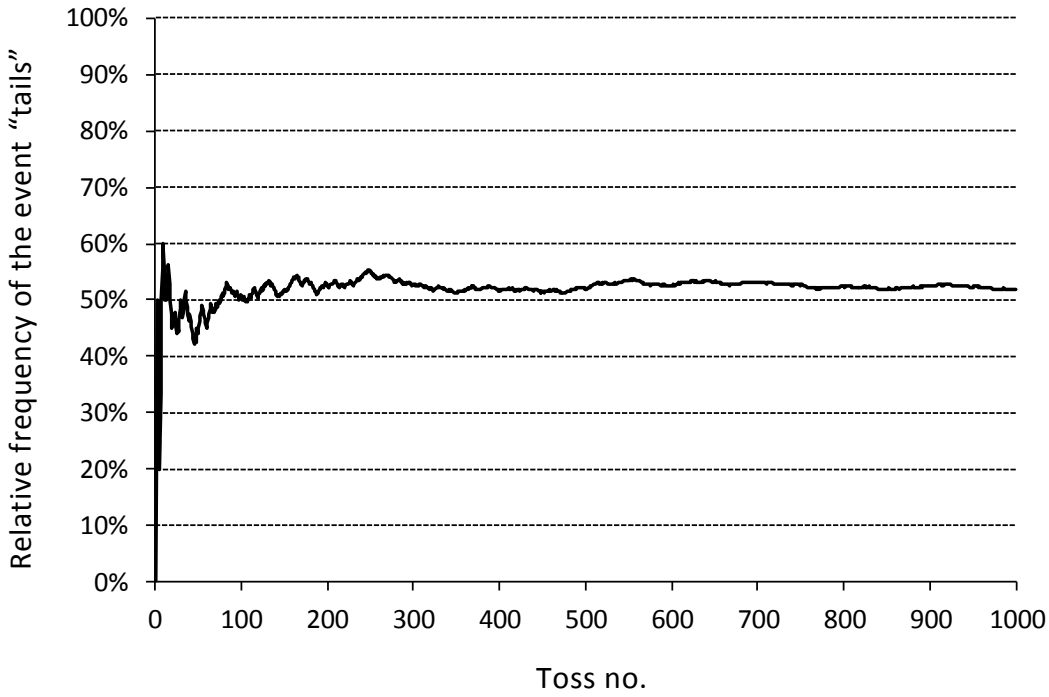



**Fig. 2.3:** Illustration of the law of large numbers. The relative frequency of the result "tail", when repeatedly flipping a coin, approaches the theoretic value of 0.5 after a sufficient number of rolls.

The following table shows how the relative frequencies in Figure 2.3 were calculated.

| Roll No. | Roll result | rel. frequency of the result “tail” | Roll No. | Roll result | rel. frequency of the result “tail” |
|---|---|---|---|---|---|
| 1 | H | 0/1 = 0.00 | 991 | T | 0.504 |
| 2 | T | 1/2 = 0.50 | 992 | T | 0.504 |
| 3 | H | 1/3 = 0.33 | 993 | T | 0.505 |
| 4 | H | 1/4 = 0.25 | 994 | T | 0.505 |
| 5 | H | 1/5 = 0.20 | 995 | T | 0.506 |
| 6 | T | 2/6 = 0.33 | 996 | H | 0.505 |
| 7 | H | 2/7 = 0.29 | 997 | H | 0.505 |
| 8 | T | 3/8 = 0.38 | 998 | H | 0.504 |
| 9 | T | 4/9 = 0.44 | 999 | T | 0.505 |
| 10 | H | 4/10 = 0.40 | 1000 | T | 0.505 |

**Table 2.1:** Calculation of the relative frequency for Figure 2.3

# 3 Statistical Measures

Essential properties of a data set are the “central tendency” of the individual values as well their “dispersion” on the number line from negative to positive infinity. This chapter explains measures suitable to describe these properties.

## 3.1 Median

Already during the collection of individual values, one must consider that their order contains essential information, namely the chronological sequence of their appearance (e.g. temperature range, trial order). These values are therefore recorded in the order of appearance. The resulting list is referred to as the master list.

If several parts are drawn subsequently from a serial production whose characteristics are to be measured and analyzed, it is advisable to number the parts according to the order in production.

This is especially important, if the measurement process is performed at a different location as the production site and there is a risk of losing the order. Of course, the values (measurement values) are recorded in correlation with the numbering of the parts.



*EXAMPLE 3.1:*

*The following nine measurements were collected: 5, 6, 6, 3, 5, 8, 6, 7, 4.*

One can assume that these are deviations from a specified desired value (mean of the tolerance range), e.g. 1/100 mm or mV.

Generally, one designates a value with the letter $x$ and the number of the values with $n$. A running index $i$ is added to the symbol $x$:

$$x_i;\ i = 1, 2, 3, \ldots, n,$$

meaning the values are designated with

$$x_1, x_2, x_3, \ldots, x_n.$$

If the values are arranged according to their size, beginning with the smallest value, the result is called an ordered list.

*According to Example 3.1: 3, 4, 5, 5, 6, 6, 6, 7, 8*

In mathematical terms:

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

The indices are placed in parentheses in order to differentiate them from the values of the master list.

In an ordered list, the first quantity corresponds to the minimum value and the last corresponds to the maximum value:

$$x_{(1)} = x_{min} \qquad x_{(n)} = x_{max}.$$

Plotting the values against the x-axis displays their frequency distribution (Figure 3.1).

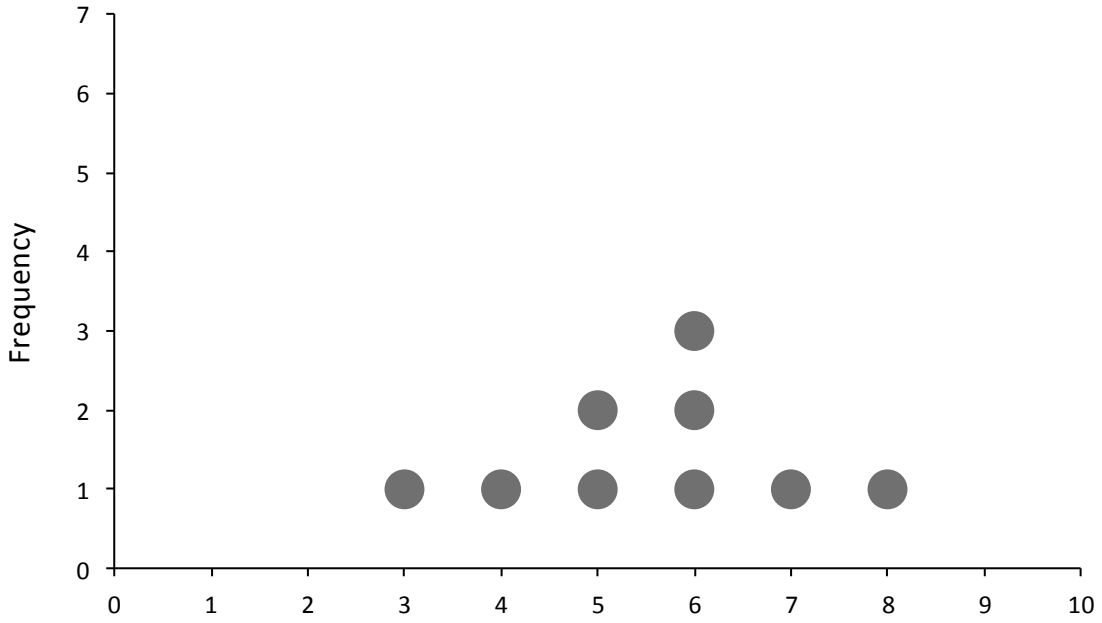



**Fig. 3.1:** Frequency diagram (dot diagram)

An easy to calculate location measure is the median, or central value. It divides the sample into two halves of equal number. The median, designated with $\tilde{x}$ (pronounce: "x-tilde"), is determined by counting the values in an ordered list:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} \qquad \text{if } n \text{ is odd,}$$

$$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \qquad \text{if } n \text{ is even.}$$

The median is therefore only included in the value set if there is an odd number of measured values; for an even number, the median equals the average of the sum of the two neighboring values $x_{\left(\frac{n}{2}\right)}$ and $x_{\left(\frac{n}{2}+1\right)}$.

*The ordered list given above:*

$x_{(1)} = 3,\ x_{(2)} = 4,\ x_{(3)} = 5,\ x_{(4)} = 5,\ x_{(5)} = 6,\ x_{(6)} = 6,\ x_{(7)} = 6,\ x_{(8)} = 7,\ x_{(9)} = 8$

*has a median of:* $\tilde{x} = x_{(5)} = 6$.

The key benefit of the median is its independence from the extreme values within a data set.

## 3.2 Arithmetic mean

The arithmetic mean is defined as the sum of all the individual values divided by the number of individual values:

$$\bar{x} = \frac{\text{The sum of all individual values}}{\text{The number of all individual values}} \qquad (\bar{x}\text{: pronounce: “x-bar”})$$

or, formulated mathematically:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad \text{(arithmetic mean).}$$

The sum of all individual values is represented in a simplified manner by a summation symbol (upper-case Greek letter sigma). This means that all $n$ values $x$, beginning with the first measured value $x_1$ (with $i = 1$) to the last measured value $x_n$ (with $i = n$), are summed up.

Written, it looks as follows:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ \ldots + x_n .$$

*For the nine measurement values ( $n = 9$ ) in Example 3.1 the sum is:*

$$\sum_{i=1}^{9} x_i = 5 + 6 + 6 + 3 + 5 + 8 + 6 + 7 + 4 = 50$$

*and the arithmetic mean:*

$$\bar{x} = \frac{50}{9} \approx 5.6 .$$

If $\bar{x}$ is not indexed, it always represents the arithmetic mean in this book.

As the following example shows, the mean is merely a starting point for the “central location of the values” on the number line. Without additional information, it can be almost useless.

*EXAMPLE 3.2:*

*The 20 students from a class have an average body height of 1.70 m.*

*Does this allow for a conclusion on the actual distribution of body size?*

*It is conceivable that the body sizes of 1.50 m, 1.60 m, 1.80 m, and 1.90 m were roughly equally represented, and the remaining students are 1.70 m tall.*

*However, it is also possible that half of the students are 1.60 m tall and the other half are 1.80 m tall.*

*It is also possible that 19 students are roughly 1.66 m tall and one is extremely tall at 2.46 m.*

This example shows that the explanatory power of the arithmetic mean is related to the associated distribution model (single peak, multiple peaks, symmetric, skewed distribution). It is especially clear that extreme values have a strong influence on the arithmetic mean.

## 3.3 Moving average

A moving average is created from a series of values by formally combining the values of this series in groups of $n$ values and calculating the mean for each of these $n$ values.

For each new value that is added to the series, one removes the first value from the last group of values so that a new group of values is created with the size $n$, from which the new moving average will be calculated, etc.

*Example for* $n = 5$ *:*

| | |
|---|---|
| ***3 7 4 9 1*** | $\bar{x}_1 = 4.8$ |
| *3* ***7 4 9 1 8*** | $\bar{x}_2 = 5.8$ |
| *3 7* ***4 9 1 8 5*** | $\bar{x}_3 = 5.4$ |
| *3 7 4* ***9 1 8 5 2*** | $\bar{x}_4 = 5.0$ |

Of course, the moving averages calculated in this manner are no longer independent from one another. This measure therefore only responds with a delay to sudden changes, which is most definitely intended.

For example, a long-term trend is more easily recognized in a depiction of the number of monthly car and station wagon registrations over time, if a moving average has been derived from the numbers of the previous 6 or 12 months (see Figure 3.2). Short-term deviations have almost no effect on the moving average.

Within the scope Statistical Process Control, it is possible to use a quality control chart with a moving average in order to control processes. However, in this case the delayed "response" of the moving average to sudden, undesired process conditions can be a disadvantage.

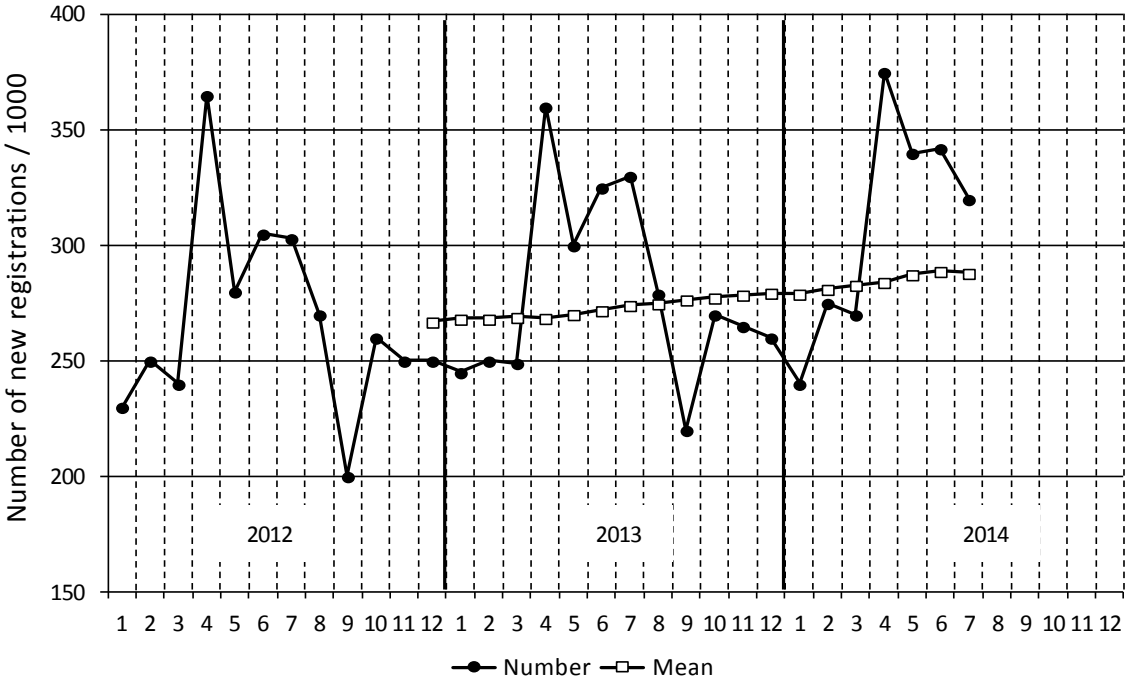



**Fig. 3.2:** Moving 12-month average

## 3.4 Geometric mean

The geometric mean corresponds to the $n$-th root of the product of all $n$ values in a series of numbers:

$$\bar{x}_g = \sqrt[n]{\text{Product of all n individual values}}$$

or in mathematical terms:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad \text{(geometric mean).}$$

The product of all the individual values is simply represented by the upper-case Greek letter pi. This means that all $n$ values $x$, beginning with the first measured value $x_1$ (with $i = 1$) to the last measured value $x_n$ (with $i = n$), are multiplied with each other:

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot \ldots \cdot x_n .$$

*The nine measured values from Example 3.1 will serve again as an example. Their product is:*

$$\prod_{i=1}^{9} x_i = 5 \cdot 6 \cdot 6 \cdot 3 \cdot 5 \cdot 8 \cdot 6 \cdot 7 \cdot 4 = 3628800 ,$$

*and the resulting geometric mean is:*

$$\bar{x}_g = \sqrt[9]{3628800} \approx 5.4 .$$

The geometric mean is used in conjunction with growth processes.

*EXAMPLE 3.3:*

*Let us assume that the population in a city has grown exponentially. We want to exclude sudden (discontinuous) changes from massive influxes or catastrophes.*

| *Time in years* | *Year* | *Population* |
|---|---|---|
| $t_1 = 0$ | *1990* | $N_1 = 100000$ |
| $t_2 = 10$ | *2000* | $N_2 = 141000$ |
| $t_3 = 20$ | *2010* | $N_3 = 200000$ |

*We assume that the population is known for years 1990 and 2010, and we want to determine the population in the year 2000 from this information.*

*The arithmetic mean would provide a population of 150,000 for the year 2000. However, with this type of calculation, one would not be considering the exponential growth and would falsely calculate the value that results from linear growth.*

*The geometric mean provides the correct estimation in this example:*

$$\bar{x}_g = \sqrt[2]{100000 \cdot 200000} \approx 141000 .$$

*The reason for this relationship becomes clear when one considers what exponential growth is. The population grows according to the function*

$N = N_0 \cdot e^{a \cdot t}$

*in time $t$. Using this law of growth and the information from the years 1990 and 2010, the growth parameter $a$ can be calculated:*

$$a = \frac{1}{t} \cdot ln\left(\frac{N_3}{N_1}\right) = \frac{1}{20} \cdot ln\left(\frac{200000}{100000}\right) = 0.03466\,.$$

*The population in the year 2000 can then be determined by substituting t with the mean time* $t_2 = \frac{t_1 + t_3}{2}$*:*

$$N_2 = N_1 \cdot e^{a \cdot \frac{t_1 + t_3}{2}} = 100000 \cdot e^{0.03466 \cdot 10} \approx 141000\,.$$

*This corresponds to the result obtained using the geometric mean.*



## 3.5 Harmonic mean

Provided that the measurement values $x_i$ are ratios (or reciprocals), calculating the arithmetic mean will lead to an incorrect result.

*EXAMPLE 3.4:*

*A motorist travels 200 km on the highway. He covers the first half of the trip $s_1 = 100\,km$ at $v_1 = 80\,km/h$, the second half $s_2 = 100\,km$ at $v_2 = 160\,km/h$. What is the average speed?*

*The obvious answer* $\bar{v} = \frac{80 + 160}{2} km/h = 120\,km/h$ *is wrong!*

*The correct result is obtained by dividing the complete distance by the time required:* $\bar{v} = \frac{s_1 + s_2}{t_1 + t_2} = \frac{s_1 + s_2}{\frac{s_1}{v_1} + \frac{s_2}{v_2}}$.

*Since both segments have the same length ( $s_1 = s_2$ ), this gives:*

$$\bar{v} = \frac{2}{\frac{1}{80\,km/h} + \frac{1}{160\,km/h}} \approx 107\,km/h$$

In general, the value to be considered

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

is called the harmonic mean value (harmonic mean).

## 3.6 Standard deviation

The data sets pictured in Figure 3.3 consist of 7 measurement values each and are all characterized by the same arithmetic mean $\bar{x} = 5$.

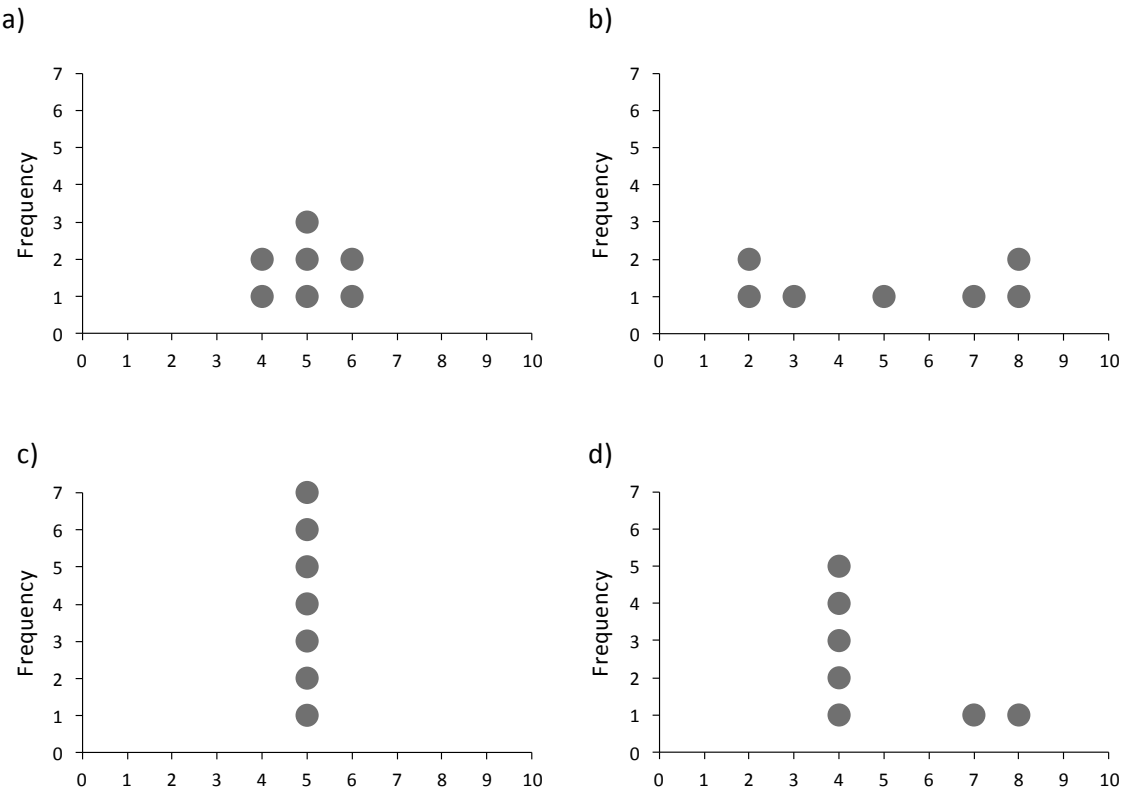



**Fig. 3.3:** Dot diagram for data sets with the same mean ( $\bar{x} = 5$ )

Although the arithmetic mean is the same in all cases, the individual values are obviously dispersed differently around the mean. This means that a more or less large deviation of the individual values around the mean. In Figure 3.3c, the deviation is smallest, in 3.3b it is largest.

Therefore it appears to be useful to calculate an average deviation from the arithmetic mean by dividing the sum of the individual deviations $\sum_{i=1}^{n} (x_i - \bar{x})$ by the number of the individual values $n$: Mean deviation = $\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})$.

However, the problem occurs that the sum of all individual deviations becomes zero:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x})$$

$$= (x_1 + x_2 + x_3 + \ldots + x_n) - n \cdot \bar{x}$$

$$= \left( \sum_{i=1}^{n} x_i \right) - n \cdot \bar{x} .$$

Because of the relationship $\sum_{i=1}^{n} x_i = n \cdot \bar{x}$ (definition of the mean), it follows that

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 .$$

Obviously, the sum of the deviations of the individual values from the mean is not a useful measure of dispersion.

The alternative is to add up the absolute values of the individual deviations from the mean and then to divide the sum by the sample size. The measure of dispersion defined in this way is called the mean linear deviation:

$$D = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| x_i - \bar{x} \right| .$$

However, this measure is not commonly used.

On the contrary, a very important and often used measure of dispersion can be obtained by summing up the squares of the individual deviations instead of the absolute values, and using the result as the measure:

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

On the one hand, by squaring, the individual summands of the total deviation become positive and on the other hand, the individual values that lie farther away from the mean have a greater influence. A suitable measure of dispersion is finally obtained by dividing the sum of the squared deviations by the sample size minus one.

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

This measure $s^2$ of deviation is called “variance”.

It should be noted that the division is not performed with the number of individual values $n$, but rather with that very number minus one $n-1$. The reason for this is that the sample variance defined in this way is a “good” (in mathematical terms: unbiased) estimation of the unknown variance of the examined population.

The value that is derived by taking the square root of the variance $s^2$ is called (empirical) standard deviation $s$:

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

Because it is calculated using a sum of squared expressions, it is always a positive number (greater than or equal to zero).

With regard to repeated measurements in order to estimate the measurement uncertainty, measurements are occasionally indicated in this way $(23 \pm 0.2)$ mm.

The number 23 corresponds to the mean $\bar{x}$ calculated from the individual measurement values, and the number 0.2 corresponds, for example, to three times the standard deviations of these individual values.

So the statement $\bar{x} \pm 3 \cdot s$ includes besides the measured mean also information about the "range" of the measured values.

Since each measured value (e.g. 23 mm) consists of a value (the number 23) and a unit (mm), it is obvious that the variance is not suitable for indicating the measurement uncertainty (a statement in the form of 23 mm $\pm$ 0.04 mm$^2$ wouldn't make sense).

*The manual calculation of the standard deviation s will be explained using the values from Example 3.1. The following table will suit this purpose:*

| *Running index* $i$ | *Individual values* $x_i$ | *Deviation* $x_i - \bar{x}$ | *Squared deviation* $(x_i - \bar{x})^2$ |
|---|---|---|---|
| *1* | *5* | *-0.55* | *0.303* |
| *2* | *6* | *0.45* | *0.203* |
| *3* | *6* | *0.45* | *0.203* |
| *4* | *3* | *-2.55* | *6.503* |
| *5* | *5* | *-0.55* | *0.303* |
| *6* | *8* | *2.45* | *6.003* |
| *7* | *6* | *0.45* | *0.203* |
| *8* | *7* | *1.45* | *2.103* |
| *9* | *4* | *-1.55* | *2.403* |
| *Sum* | *50* | | *18.227* |

**Table 3.1:** Evaluation of the values from Example 3.1

$$\bar{x} = \frac{50}{9} = 5.\overline{5} \qquad s = \sqrt{\frac{1}{9-1} \cdot 18.227} = 1.51$$

## 3.7 Variation coefficient

The variation coefficient v represents a fairly important measure for the evaluation of populations. The variation coefficient relates the extent of the dispersion of the individual values to the value of the arithmetic mean:

$$v = \frac{s}{\bar{x}} \cdot 100\,\%\,.$$

The application of this measure is always beneficial when comparing two data sets, which share a similar distribution type, but whose means are significantly apart.

## 3.8 Range

Another useful measure of dispersion is the range $R$. The range is the difference between the last and the first value of an ordered series of values:

$$R = x_{(n)} - x_{(1)}$$

or, with regard to any unstructured value set, the difference between the biggest and the smallest values:

$$R = x_{max} - x_{min}.$$

The range is always a positive number (greater than or equal to zero).

*EXAMPLE 3.5:*

*The result for the value set (2, 3, 7, 5, 3, 2, -2, 0, 4, 3) is:*

$x_{max} = 7 \quad x_{min} = -2 \quad R = 7 - (-2) = 9.$



## 3.9 Range method to determine the standard deviation

The range method is a simplified calculation to quickly determine a standard deviation $s_R$. This measure of dispersion $s_R$ is a good approximation for $s$ and is sufficiently accurate for many practical applications. The requirements for this simple procedure are: the data set must be derived from a normal distribution and especially not contain any outliers.

The values from the measurement series $m$ are divided into groups (samples) with $n$ individual values each. The data set consists, therefore, of a total of $m \cdot n$ individual values. In general, this procedure is applied when the measurement values occur in groups anyway, e.g. with the median r chart within an SPC in the form of 5-piece samples.

The respective group mean $\bar{x}_j$ is then

$$\bar{x}_j = \frac{1}{n} \cdot \sum_{i=1}^{n} x_{i,j}$$

With $i$: running index within a group,

$j$: running index for the groups ( $j = 1, 2, \ldots, m$).

The range of each group is $R_j = R_{j,max} - R_{j,min}$.

The mean $\bar{R}$ of the range of all groups is

$$\bar{R} = \frac{1}{m} \cdot \sum_{j=1}^{m} R_j .$$

From $\bar{R}$ and the use of a tabulated auxiliary quantity $d_2^*$ finally the standard deviation $s_R$ can be calculated:

$$s_R = \frac{\bar{R}}{d_2^*}.$$

$d_2^*$ is dependent on the number $n$ of the individual values per group as well as the number m of the groups (see Table 3.2).

| | | Sample size n | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Number m of groups | 1 | 1.414 | 1.912 | 2.239 | 2.481 | 2.673 | 2.830 | 2.963 | 3.078 | 3.179 |
| | 2 | 1.279 | 1.805 | 2.151 | 2.405 | 2.604 | 2.768 | 2.906 | 3.024 | 3.129 |
| | 3 | 1.231 | 1.769 | 2.120 | 2.379 | 2.581 | 2.747 | 2.886 | 3.006 | 3.112 |
| | 4 | 1.206 | 1.750 | 2.105 | 2.366 | 2.570 | 2.736 | 2.877 | 2.997 | 3.103 |
| | 5 | 1.191 | 1.739 | 2.096 | 2.358 | 2.563 | 2.730 | 2.871 | 2.992 | 3.098 |
| | 6 | 1.181 | 1.731 | 2.090 | 2.353 | 2.558 | 2.726 | 2.867 | 2.988 | 3.095 |
| | 7 | 1.173 | 1.726 | 2.085 | 2.349 | 2.555 | 2.723 | 2.864 | 2.986 | 3.092 |
| | 8 | 1.168 | 1.721 | 2.082 | 2.346 | 2.552 | 2.720 | 2.862 | 2.984 | 3.090 |
| | 9 | 1.164 | 1.718 | 2.080 | 2.344 | 2.550 | 2.719 | 2.860 | 2.982 | 3.089 |
| | 10 | 1.160 | 1.716 | 2.077 | 2.342 | 2.549 | 2.717 | 2.859 | 2.981 | 3.088 |
| | … | … | … | … | … | … | … | … | … | … |
| $d_2$ | | 1.128 | 1.693 | 2.059 | 2.326 | 2.534 | 2.704 | 2.847 | 2.970 | 3.078 |

**Table 3.2:** Values of $d_2^*$ depending on n and m. $d_2$ is the limit of $d_2^*$ for infinitely many groups (i.e. $m \to \infty$).

*EXAMPLE 3.6:*

| | *Group No.* | | | | | |
|---|---|---|---|---|---|---|
| *i* | *1* | *2* | *3* | *4* | *5* | *6* |
| *1* | *70* | *71* | *68* | *72* | *72* | *72* |
| *2* | *68* | *67* | *72* | *76* | *66* | *69* |
| *3* | *69* | *66* | *69* | *67* | *63* | *63* |
| *4* | *69* | *64* | *67* | *68* | *73* | *68* |
| *5* | *75* | *72* | *69* | *69* | *72* | *68* |
| $\tilde{x}_j$ | $\tilde{x}_1 = 69$ | $\tilde{x}_2 = 67$ | $\tilde{x}_3 = 69$ | $\tilde{x}_4 = 69$ | $\tilde{x}_5 = 72$ | $\tilde{x}_6 = 68$ |
| $R_j$ | $R_1 = 7$ | $R_2 = 8$ | $R_3 = 5$ | $R_4 = 9$ | $R_5 = 10$ | $R_6 = 9$ |

$$\overline{R} = \frac{1}{6} \cdot \sum_{j=1}^{6} R_j = \frac{7+8+5+9+10+9}{6} = 8$$

$$s_R = \frac{\overline{R}}{d_2^*} = \frac{8}{2.353} \approx 3.4$$

# 4 Statistical calculations in EXCEL

Measurements are often recorded in EXCEL tables. This often leads to the need to calculate simple statistical measures directly in these tables.

One can calculate a mean by, for example, adding up all of the individual values in column A using the “=SUM(A1:An)” function and then dividing the result by the number n. This would correspond to the process explained in 3.2.

To calculate the variance, one can tediously copy the steps described in Chapter 3.6, and subtract the calculated mean from every value in column A, enter the results into the respective lines in column B, then multiply every field in column B with itself, enter the results into the respective lines in column C, and then finally sum up column C and divide the result by $n-1$. This way statistical measures are determined using basic arithmetic methods like addition, subtraction, multiplication and division. To calculate the standard deviation, one also has to take the square root.

These “programming exercises” can be easily understood and handled using simple mathematical EXCEL functions. However, to quickly calculate statistical measures, the practiced Excel user will use the available statistical functions.

For example, the mean of the values in lines 1 to 10 in column A can be calculated by entering "=AVG(A1:A10)" into cell A11. Similarly, the standard deviation can be calculated by entering "=STDEV(A1:10)" into cell A12. Alternatively, one can use the corresponding functions on the graphical interface.

Note the following feature, with regard to the standard deviation. The EXCEL function "STDEV” assumes that the values entered originate from sample results, from which the standard deviation $\sigma$ of a larger population will be calculated. The calculation of s (as the estimator for $\sigma$) is realized by the following formula:

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

In this case, the sum of the squared deviations is divided by the sample size minus one.

The EXCEL function “STDEV” assumes that the provided numbers already correspond to the population (e.g. body sizes of students in a class), and the sum of the squared deviations is divided by the population size (number of entered values). So the calculation uses the formula:

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

When dealing with larger data sets, starting at $n = 50$, the difference between the factors $\frac{1}{n-1}$ and $\frac{1}{n}$ becomes meaningless (the relative “error” in the case of $n = 50$ is approximately 2%).

## Measures should be used with caution!

The following example is going to illustrate that using a statistical measure alone is not enough to provide a clear conclusion regarding the fraction nonconforming with respect to a tolerance limit (upper limit - UL).

*EXAMPLE 4.1:*

| *Measurement series 1, measured values $x_i$* | | | | |
|---|---|---|---|---|
| *22* | *25* | *25* | *26* | *26* |
| *26* | *22* | *22* | *23* | *25* |
| *24* | *23* | *24* | *23* | *23* |
| *23* | *24* | *22* | *26* | *25* |
| *26* | *25* | *24* | *22* | *24* |

*Mean: $\bar{x} = 24.0$ Standard deviation: $s_x = 1.4434$*

| *Measurement series 2, measured values $y_i$* | | | | |
|---|---|---|---|---|
| *24* | *23* | *22* | *25* | *23* |
| *23* | *25* | *21* | *24* | *24* |
| *24* | *23* | *22* | *26* | *24* |
| *23* | *26* | *24* | *24* | *25* |
| *24* | *25* | *28* | *24* | *24* |

*Mean: $\bar{y} = 24.0$ Standard deviation: $s_y = 1.4434$*

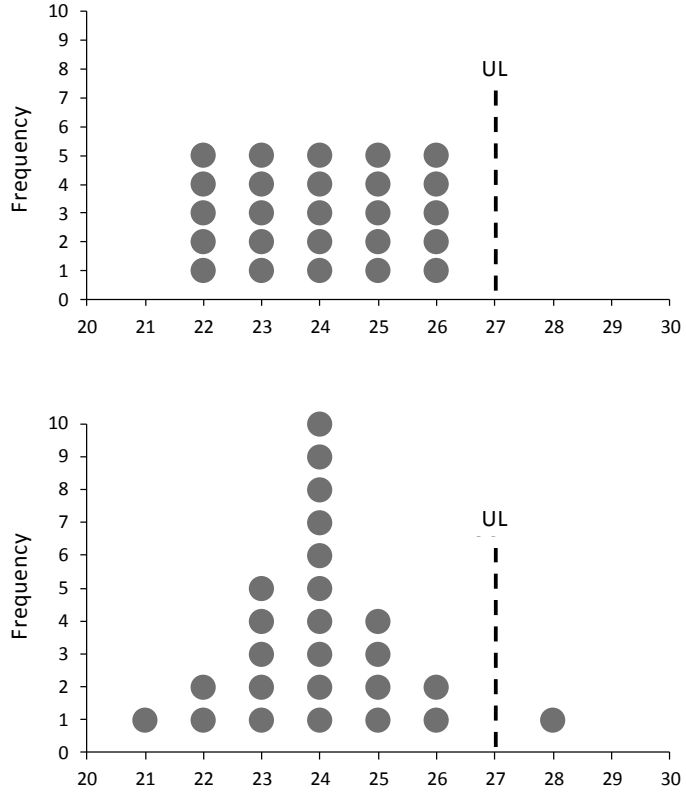



**Fig. 4.1:** Dot diagram for measurement series 1 (above) and measurement series 2 (below)

Although the means and standard deviations of both measurement series are each the same, in the first case, there is not a single value above the upper limit UL ("upper tolerance limit"), while in the second case, one measurement UL exceeds this.

This example shows that it is unavoidable, when evaluating series of measurements, to embrace a comprehensive, integrated approach and not to draw conclusions based on a few pieces of information. Obviously, the "distribution" of the measurements in the previous example should not be disregarded (see Chapter 6).

*NOTE:*

*It is often assumed that the terms "mean" and "standard deviation" always refer to the normal distribution (see Chapter 6). This assumption is not correct.*

# 5 Graphical display of data

A graphical display of data allows the viewer to quickly ascertain the essential properties of a data set. It assists and facilitates when examining series of measurements.

For instance, by using an original chart it is extremely easy to recognize special characteristics like the starting point, end point, trends, periodicity, aggregations of points or individual points that lie well outside the majority of the rest of the points, so-called outliers, etc.

Moreover, the majority of graphical displays serves the viewer to determine a data set's statistical properties and measures without the use of a computer or calculator.



## 5.1 Original value chart

Using an original value chart, one lists measurement values in series based on their occurrence, the abscissa (x-coordinate) usually corresponds to time. When examining processes, data is normally recorded in increments of minutes, hours, shifts or days. The time, therefore, can be indicated in these cases by entering the date and time of day. When examining parts that were sampled from a production process, there may be a great difference between the respective points of time of the sampling and the measurement of the value of the part's characteristic. This may play a role, for instance, with products that are subject to alterations over time (e.g. plastic parts, adhesive bonds).

On the other hand, when conducting experimental studies it may be of interest to define the beginning as point zero on the time scale (e.g. the transient response of a control device). The following example shows the temperature profile in a drying chamber with a simple two-step control.

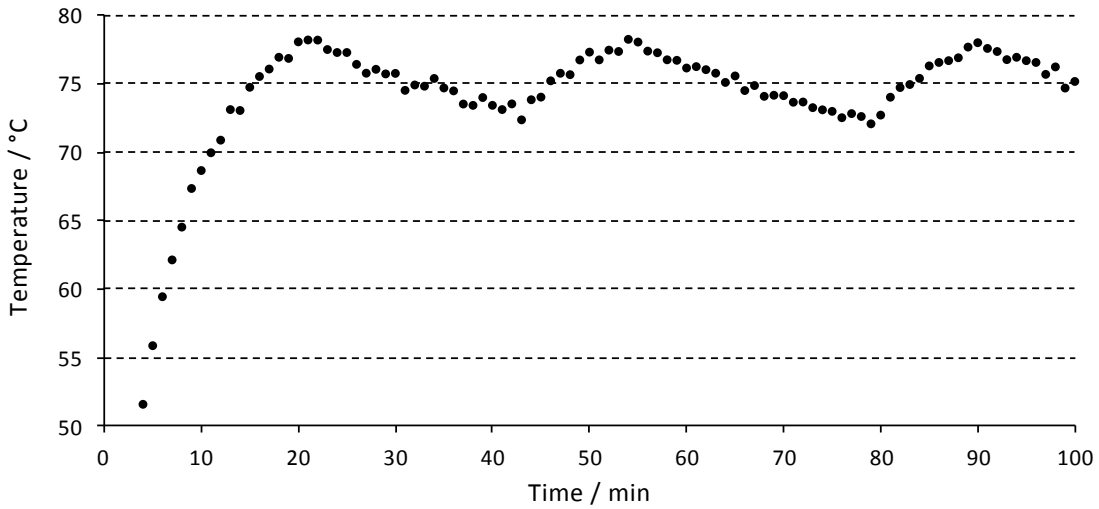



**Fig. 5.1:** Example of an original value chart

## 5.2 Tally chart, dot diagram

Examples of dot diagrams are found in the previous Chapters 3.1, 3.6 and 4. These dot diagrams as well as frequency charts or histograms are going to be used to explain grouping in following sections.

As illustrated in the following example, a dot diagram originates from a tally chart when each group of five (obviously other groupings are also possible) is represented by a dot. Compare Figure 5.2 to Figure 4.1. (below). In both cases, a so-called natural grouping arises (see Chapter 5.3) since the values only occur as integers. The height of a "pillar" is a measure for the absolute frequency of the associated value.

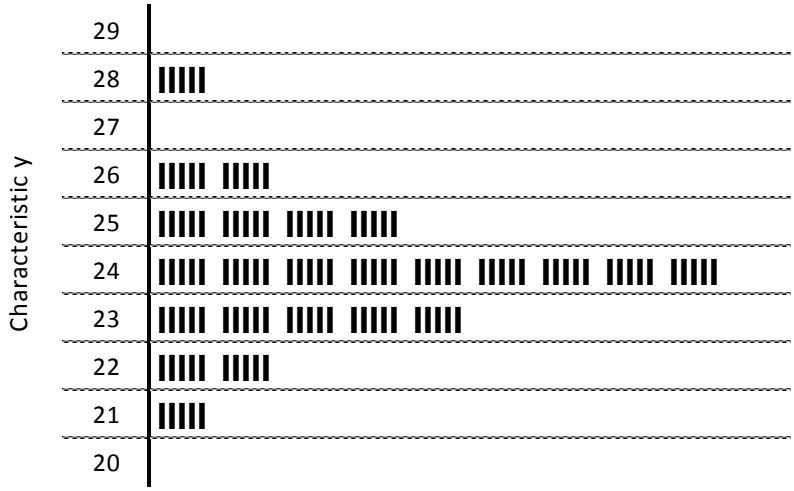



**Fig. 5.2:** Example of a tally chart

## 5.3 Grouping, histogram

If the number of measurement values in a sample is greater than about 25, it is expedient to group these values.

The grouping procedure will be explained in an example with the following list of measurement values. The measurement values, given in millimeters (mm) are hypothetical. However, they could very well originate from a production process like cutting bar stock.

*EXAMPLE 5.1:*
*The following master list contains 50 values:*

| | | | | |
|---|---|---|---|---|
| *8.0* | *7.0* | *7.4* | *8.0* | *7.0* |
| *7.4* | *7.8* | *7.5* | *7.7* | *6.9* |
| *6.5* | *7.5* | *7.6* | *7.3* | *8.0* |
| *7.0* | *7.5* | *7.1* | *7.4* | *8.6* |
| *6.0* | *8.0* | *7.0* | *8.0* | *6.9* |
| *7.5* | *8.4* | *6.8* | *8.3* | *8.0* |
| *8.3* | *7.3* | *7.0* | *7.5* | *7.9* |
| *8.0* | *7.5* | *7.0* | *6.5* | *7.8* |
| *5.8* | *7.8* | *6.3* | *7.5* | *7.9* |
| *9.0* | *8.0* | *7.1* | *7.0* | *7.4* |

*If one enters the measurement values into a frequency diagram with a grouping of 7 classes in mind, the following illustration results:*

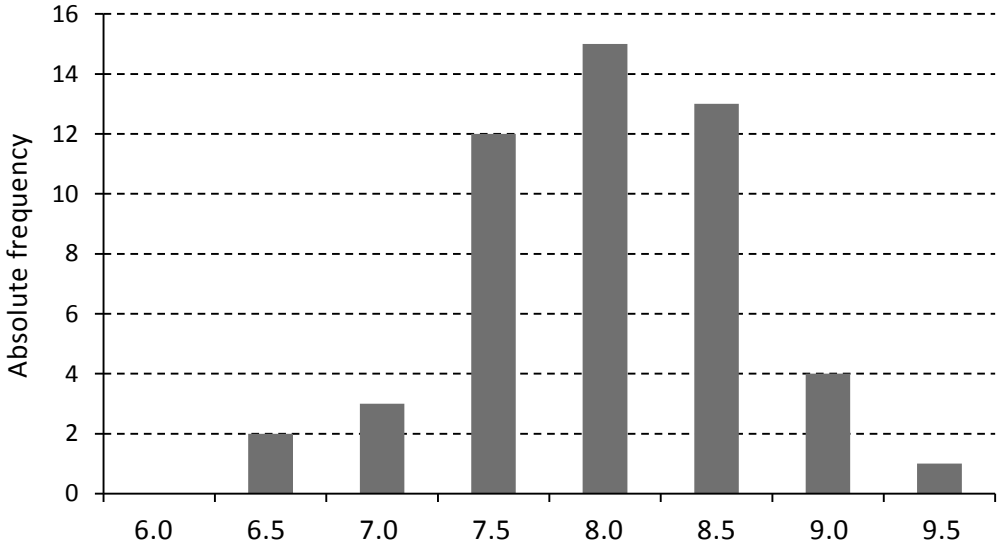



**Fig. 5.3:** Frequency diagram for Example 5.1; $k = 7$, $w = 0{,}5$

*A grouping with 22 classes, on the other hand, gives the following frequency diagram:*

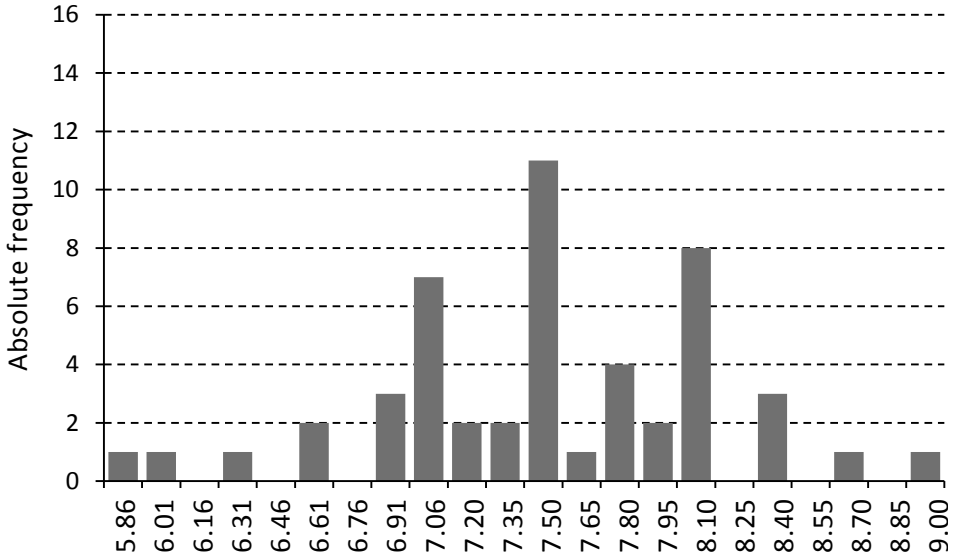



**Fig. 5.4:** Frequency diagram for Example 5.1; $k = 22$, $w = 0{,}15$

Obviously, the number of classes and their width has a strong influence on the "appearance" of the frequency diagram.

The following formula (rule of thumb) provides a clue for the suitable number $k$ of classes:

$$k \approx \sqrt{n} \quad \text{for } 25 \le n \le 100.$$

The limitation for this sample size indicates that this formula only applies for measurement series of up to 100 values. For less than 25 values, it is usually no longer sensible to create a frequency diagram.

On the other hand, if more than 100 measurement values exist, it is generally recommended to choose the number of classes using the formula:

$$k \approx 5 \cdot \log(n) \quad \text{for } n > 100.$$

The following number of classes result with these rules in mind:

| Number of measurement values n | Number of classes k |
|---|---|
| from 25 | 5 |
| 30 | 6 |
| 45 | 7 |
| 60 | 8 |
| 75 | 9 |
| 100 | 10 |
| 200 | 12 |

**Table 5.1:** Number of classes k depends on the number of measurement values n



As one can see from the frequency diagram, each class contains a certain range of values. The limits of each range (interval) are called the lower and upper class limits. The length of such an interval is called the class width $w$.

One possible way to determine this class width is to use the formula $w = \frac{x_{max} - x_{min}}{k - 1}$ for the range. This, however, usually yields class limits with several decimal places and may result in empty classes.

*The frequency diagram from Figure 5.3 was created using the $n = 50$ values from Example 5.1. The following values result when using the above rules for choosing the quantities $k$ and $w$ :*

$$k \approx \sqrt{50} \approx 7 \quad \text{and} \quad w = \frac{x_{max} - x_{min}}{k - 1} = \frac{9.0 - 5.8}{7 - 1} \approx 0.5.$$

In Figures 5.3 and 5.4, the class limits were determined by choosing the second decimal point so that each value can be clearly assigned to a class. Another option to clearly assign classes is to include the right class limit with the interval, or to evenly divide the values that are identical to the class limit into the neighboring classes, respectively.

*NOTE:*

*There are conceivable situations where it would be advantageous to select different class widths. For example, the above rule usually fails to determine the class width when the data set contains an outlier.*

*Strictly following this rule could lead to a situation where only the outer classes are occupied (one of the classes containing only the outliers) and all the others are empty. This can be avoided by, for example, disregarding individual extreme values when creating the classes, and then assigning these to the corresponding outer classes (first or last class), after determining classes that would be sensible for the situation. This means that the first class on the left and the last class on the right are open ended with no lower or upper class limits, respectively.*

*One cannot avoid the fact that statistical programs that create groupings according to a few, simple rules (when, for example, creating histograms) might deliver unusable illustrations de-*

*pending on how "exotic" the data set is, due to reasons mentioned. Which is why they normally provide the users with the option of correcting the grouping at their own discretion.*

Before we begin calculating the mean $\bar{x}$ and the standard deviation $s$ from a grouping, certain important terms need to be explained.

Class limit:
Each class from a given grouping is limited by its lower class limit $x'_{j-1}$ and its upper class limit $x'_j$.

Class midpoint $x_j$:

The class midpoint corresponds to the arithmetic mean of the lower and upper class limits:

$$x_j = \frac{x'_{j-1} + x'_j}{2}.$$

Class width $w_j$:

The class width corresponds to the distance between the lower and upper class limits:

$$w_j = x'_j - x'_{j-1}.$$

Generally, all classes have the same class width, i.e. $w_j = w$ for all classes.

Absolute frequency $n_j$:

The number of values allotted to the ($j$-th) class (one can also speak of "absolute class frequency").

Relative frequency $h_j$:

Absolute frequency divided by the total number $n$ of the values from the data set:

$$h_j = \frac{n_j}{n} \quad \text{with } n = \sum_{j=1}^{k} n_j = n_1 + n_2 + n_3 + \ldots + n_k.$$

Absolute cumulative frequency $G_j$:

Sum of the absolute frequencies $n_j$ from the first until the $j$-th class (inclusive).

$$G_j = \sum_{i=1}^{j} n_i = n_1 + n_2 + n_3 + \ldots + n_j$$

Relative cumulative frequency $H_j$:

Relative share of all values below the upper class limit of the $j$-th class:

$$H_j = \sum_{i=1}^{j} h_i = h_1 + h_2 + h_3 + \ldots + h_j$$

or, which is easier when calculating by hand: $H_j = \frac{G_j}{n}$.

A frequency diagram shows the distribution of the measured values, therefore the relationship between a variable $x$ and the frequency of its occurrence. If one plots the absolute frequencies on the x-axis, one obtains a frequency diagram (see Figures 5.3 and 5.4).

On the other hand, if one enters the relative frequencies, one obtains a so-called histogram (also called a bar graph).

*The following chart is created from the values in Example 5.1:*

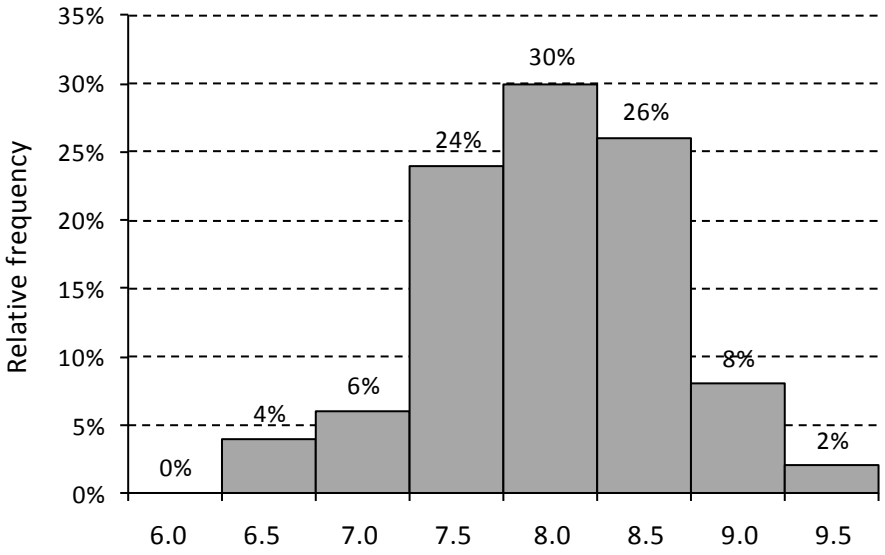



**Fig. 5.5:** Histogram for the values from Example 5.1

Here, rectangles are drawn above the classes of the characteristics, whose heights correspond to the frequencies $h_j$ (with constant class width). The following table includes almost all of the essential statistical measures:

| *Class No.* | *Lower class limit* $x'_{j-1}$ | *Upper class limit* $x'_j$ | *Absolute frequency* $n_j$ | *Relative frequency* $h_j$ | *Relative cumulative frequency* $H_j$ | *Auxiliary value (see text)* $n_j \cdot x_j$ | *Auxiliary value (see text)* $n_j \cdot x_j^2$ |
|---|---|---|---|---|---|---|---|
| *1* | *5.75* | *6.25* | *2* | *4%* | *4%* | *12.0* | *72.0* |
| *2* | *6.25* | *6.75* | *3* | *6%* | *10%* | *19.5* | *126.8* |
| *3* | *6.75* | *7.25* | *12* | *24%* | *34%* | *84.0* | *588.0* |
| *4* | *7.25* | *7.75* | *15* | *30%* | *64%* | *112.5* | *843.8* |
| *5* | *7.75* | *8.25* | *13* | *26%* | *90%* | *104.0* | *832.0* |
| *6* | *8.25* | *8.75* | *4* | *8%* | *98%* | *34.0* | *289.0* |
| *7* | *8.75* | *9.25* | *1* | *2%* | *100%* | *9.0* | *81.0* |
| $\Sigma$ | | | $n = 50$ | *100%* | | *375.0* | *2832.6* |

**Table 5.2**

## 5.4 Cumulative curve

If one plots the relative cumulative frequencies above the upper class limits, one obtains a s-shaped curve, the so-called cumulative curve.

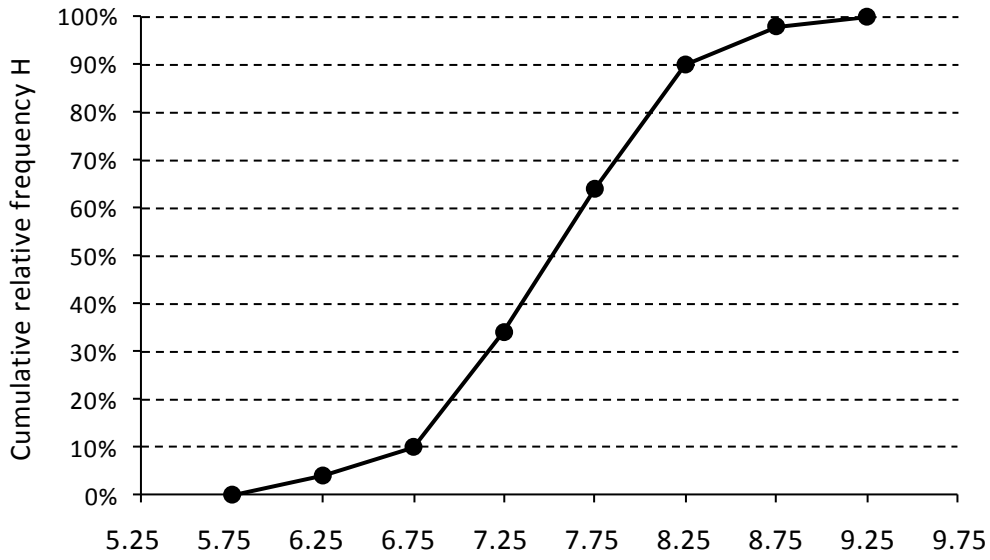



**Fig. 5.6:** Cumulative curve for the values from Example 5.1



The benefit of a cumulative curve versus a frequency diagram is easy to see. One can read, without much trouble, the percentage of the measured values that lie below a certain value on the x-axis (e.g. when estimating a rejection rate). In the example shown, 90% of the data lies below the value 8.25, or 10% of the data lies above this value. If the original values of the data set are not known, the following formulas can be of use to calculate the mean and the standard deviation using the information included in the histogram (remember: here $x_j$ designate the class midpoints) .

Mean: $\bar{x} = \frac{1}{n} \cdot \sum_{j=1}^{k} (n_j \cdot x_j) = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + \ldots + n_k \cdot x_k}{n}$

Variance: $s^2 = \frac{1}{n-1} \cdot \sum_{j=1}^{k} n_j \cdot (x_j - \bar{x})^2 = \frac{1}{n-1} \cdot \left( \sum_{j=1}^{k} (n_j \cdot x_j^2) - n \cdot \bar{x}^2 \right)$

Standard deviation: $s = \sqrt{s^2}$

The absolute frequencies $n_j$ can be calculated using the relative frequencies $h_j$ : $n_j = n \cdot h_j$.

*In the above example (see Table 5.2), one finds:*

$\bar{x} = \frac{375.0}{50} = 7.5$ *and* $s^2 = \frac{1}{50-1} \cdot (2832.6 - 50 \cdot 7.5^2) = 0.41$ *and finally* $s = 0.64$.

*When calculating these quantities using the original values, one obtains:* $\bar{x} = 7.454$ *and* $s = 0.6399$.

# 6 Statistical distributions

## 6.1 Gaussian normal distribution

If one continuously increases (under constant measurement conditions) the size $n$ of a series of measurements (i.e. the number of the measurement values becomes theoretically infinitely large), and one simultaneously reduces the class width (to zero), the cumulative curve (see 5.4) approaches a limit curve that corresponds to the distribution (distribution function) of the (infinite) population. Similarly, the outline of the step function representing the upper edge of the histogram approaches a limit curve that is the graphical representation of a probability density function (see Figure 6.1, $n \to \infty$, class width $w \to 0$).



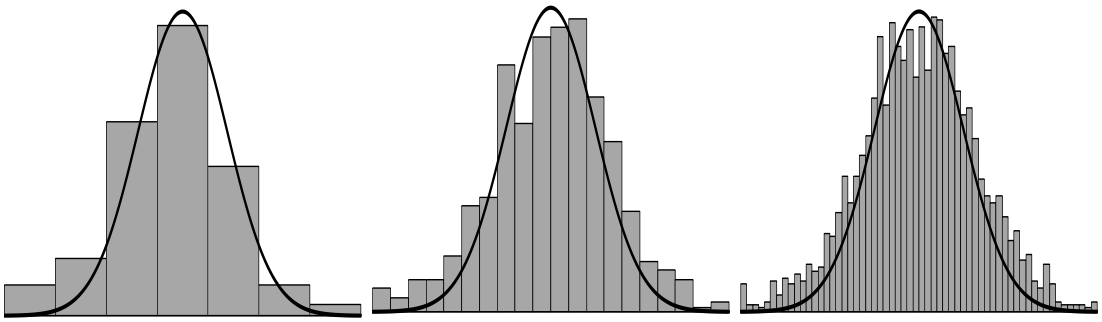


**Fig. 6.1:** Illustration of the transition from histogram to density function using the example of a normal distribution

Each point on the x-axis corresponds to a number with theoretically infinitely many decimal points, e.g. $x = 73.2645271\mathbf{8}\ldots$. The probability that a characteristic (with known distribution) assumes this value exactly is zero. On the other hand, the probability that a value lies within the range (interval) between 73.2 and 73.4 is a finite number greater than zero. One obtains such a probability by multiplying a value on the probability density function with the interval width. The probability density function is the generalization of the relative frequency when the class width shrinks to zero, so to speak.

The term “density” touches upon an analogy between the calculus of probabilities and the mechanics of rigid bodies (see e.g. [3]).

The area confined by the probability density function and a definite interval on the x-axis corresponds to the probability with which values of the population fall into this very interval. This area is therefore a graphical analog to the probability. The total area limited by a chosen probability density function and the characteristic axis (between negative and positive infinity) always corresponds to the value 1 (=100%).

The past has shown that experimental studies and statistical observations often find distributions of characteristics that result in histograms with a similar appearance. The mathematician C. F. Gauss examined this phenomenon using land survey data. This type of distribution is called the “normal distribution” and often serves as a distribution model for technical and statistical phenomena. Due to their characteristic shape, illustrations of this distribution’s density function are also called the “Gaussian bell curve”.

### 6.1.1 Properties and measures of the normal distribution

The bell curve is the graphical representation of the density function of the normal distribution, which is described by the mathematical relationship

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

The function (and the curve) is clearly defined by the parameters $\mu$ and $\sigma$. Thereby, $\mu$ represents the mean of the distribution and $\sigma$ its standard deviation.

Looking at the functional equation or its graphical representation, several special characteristics can be found:

- The curve is symmetrical about the mean $\mu$.
- The curve has an inflection point at the points $\mu - \sigma$ and $\mu + \sigma$ respectively. This means that, e.g. at $\mu - \sigma$ the curve changes its orientation from convex (away from the $x$-axis) to concave (toward the $x$-axis).
- The curve runs from $x = -\infty$ to $x = +\infty$. However, that is only interesting from a theoretical point of view. Practically speaking, the curve is only meaningful at a distance of three to four standard deviations to the left and right of the mean $\mu$ in Figure 6.2. There, the curve already approaches the $x$-axis.

As already explained, the area under the Gaussian curve corresponds to an infinitely large number of measurement values from a normally distributed population. If this area is set to 1 (corresponds to 100%), a proportion that lies between two points can be determined (in %).

If on the x-axis around the mean $\mu$ equal distances are marked in multiples of the standard deviation, the proportion of the distribution can be given depending on $\sigma$. In Figure 6.2, these proportions are marked in gray for the areas $\mu \pm 1 \cdot \sigma$, $\mu \pm 2 \cdot \sigma$ and $\mu \pm 3 \cdot \sigma$.

Accordingly, the following is obtained for the interval

$\mu \pm 1 \cdot \sigma$ a percentage of 68.3%,

$\mu \pm 2 \cdot \sigma$ a percentage of 95.4%,

$\mu \pm 3 \cdot \sigma$ a percentage of 99.7%.

One can see that outside $\mu \pm 3 \cdot \sigma$ there is only an infinitesimally small percentage of the distribution, namely just 0.3% (= 100% - 99.7%) (see Figure 6.2).

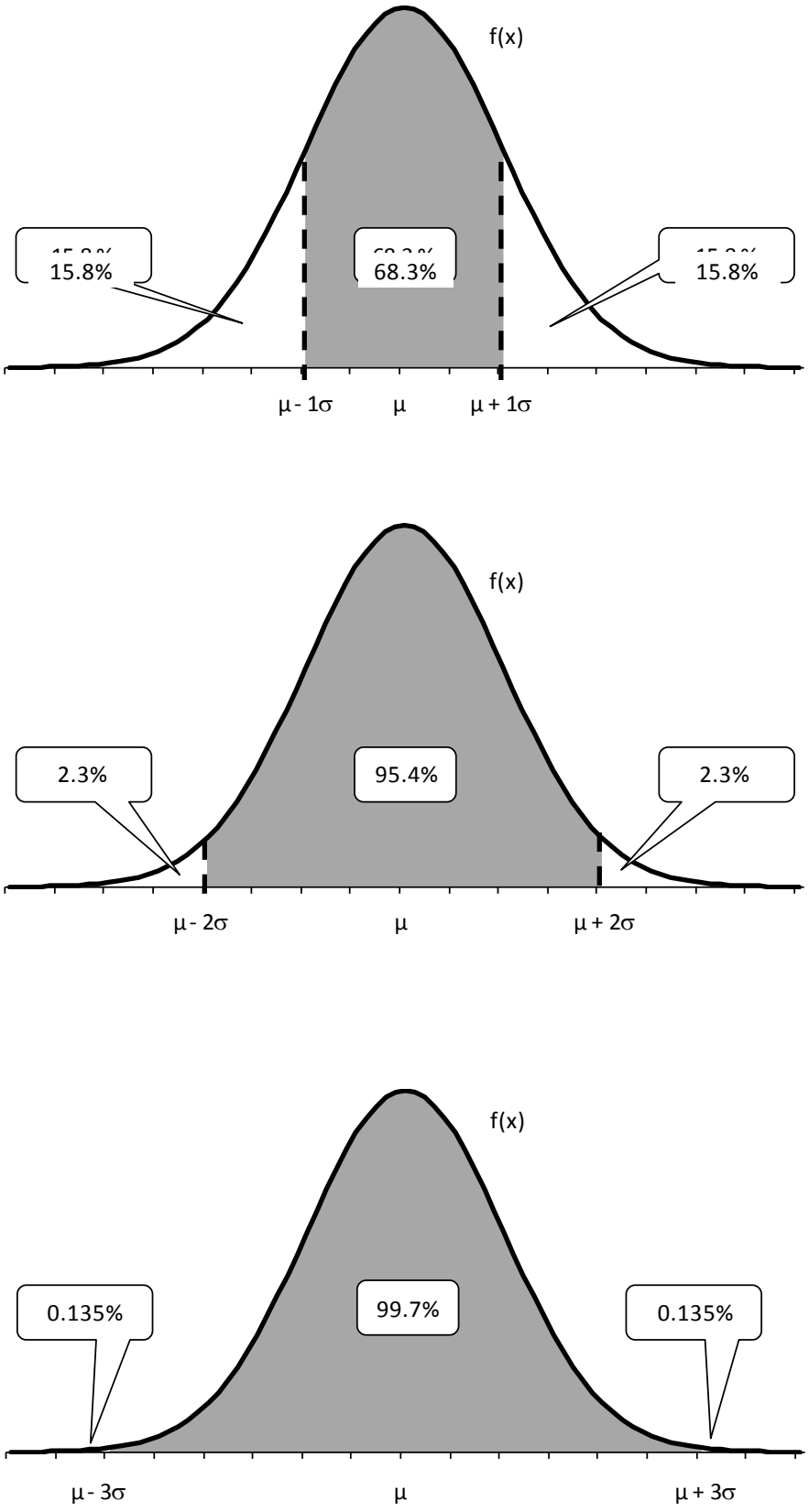



**Fig. 6.2:** Surface area under the bell curve

### 6.1.2 Distribution function

Chapter 5.4 describes how to determine the cumulative curve of a frequency distribution. The relative cumulative frequencies are determined by adding the individual relative class frequencies and then they are plotted over the upper class limits. The corresponding points are connected piecewise using lines. All of the values (= 100% of the distribution) are then accounted for at the last upper class limit.

The cumulative curve of the Gaussian distribution is also determined this way, in principle. The frequency curve created using the characteristic value $x$, which corresponds to a certain section of the area below the Gaussian curve, must now be calculated with a special mathematical procedure: integration.

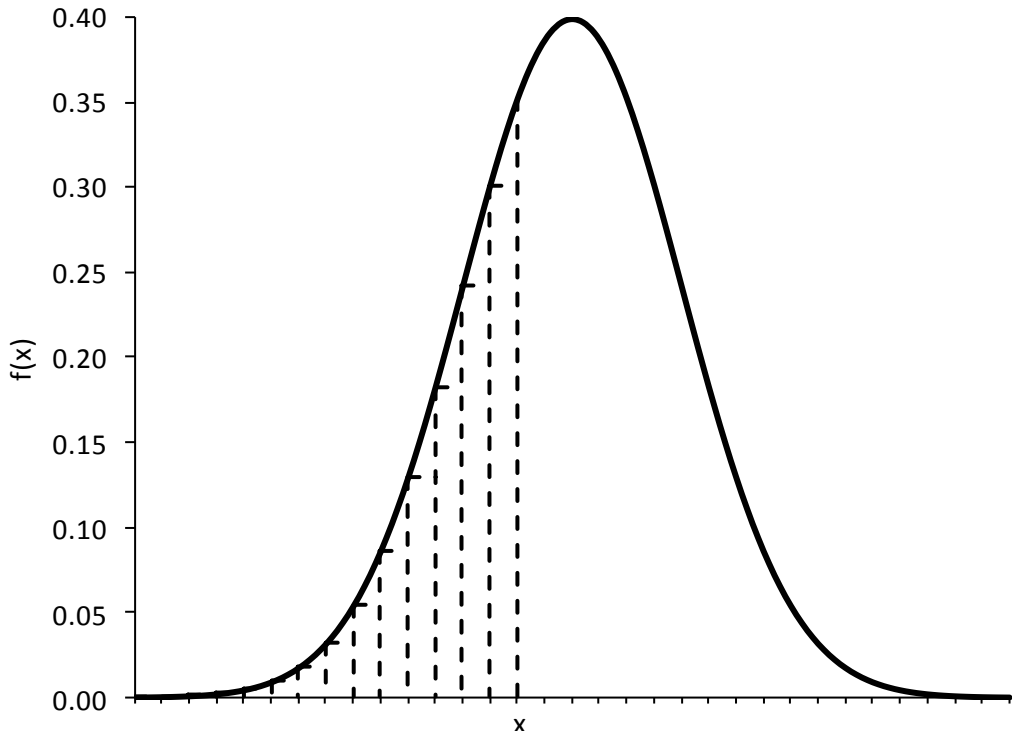



**Fig. 6.3:** Illustration of integration

The function that describes the cumulative curve of a probability distribution is called the distribution function $F(x)$. It provides for every $x$ the probability that a randomly measured value is less than or equal to $x$.

Mathematically formulated, the cumulative probability up to the point $x$ is given by the distribution function:

$$F(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2} \cdot \left(\frac{v-\mu}{\sigma}\right)^2} dv .$$

$F(x)$ corresponds to the area under the Gaussian bell curve up to the value $x$.

Figure 6.3 illustrates the meaning of integration. The area under the curve up to the point $x$ is calculated approximately by defining and summing up the areas of narrow rectangles (width $\Delta x$). The result becomes more accurate the narrower the rectangles are (passage to the limit $\Delta x \rightarrow 0$).

The distribution function is given as a curve in Figure 6.4 (below).

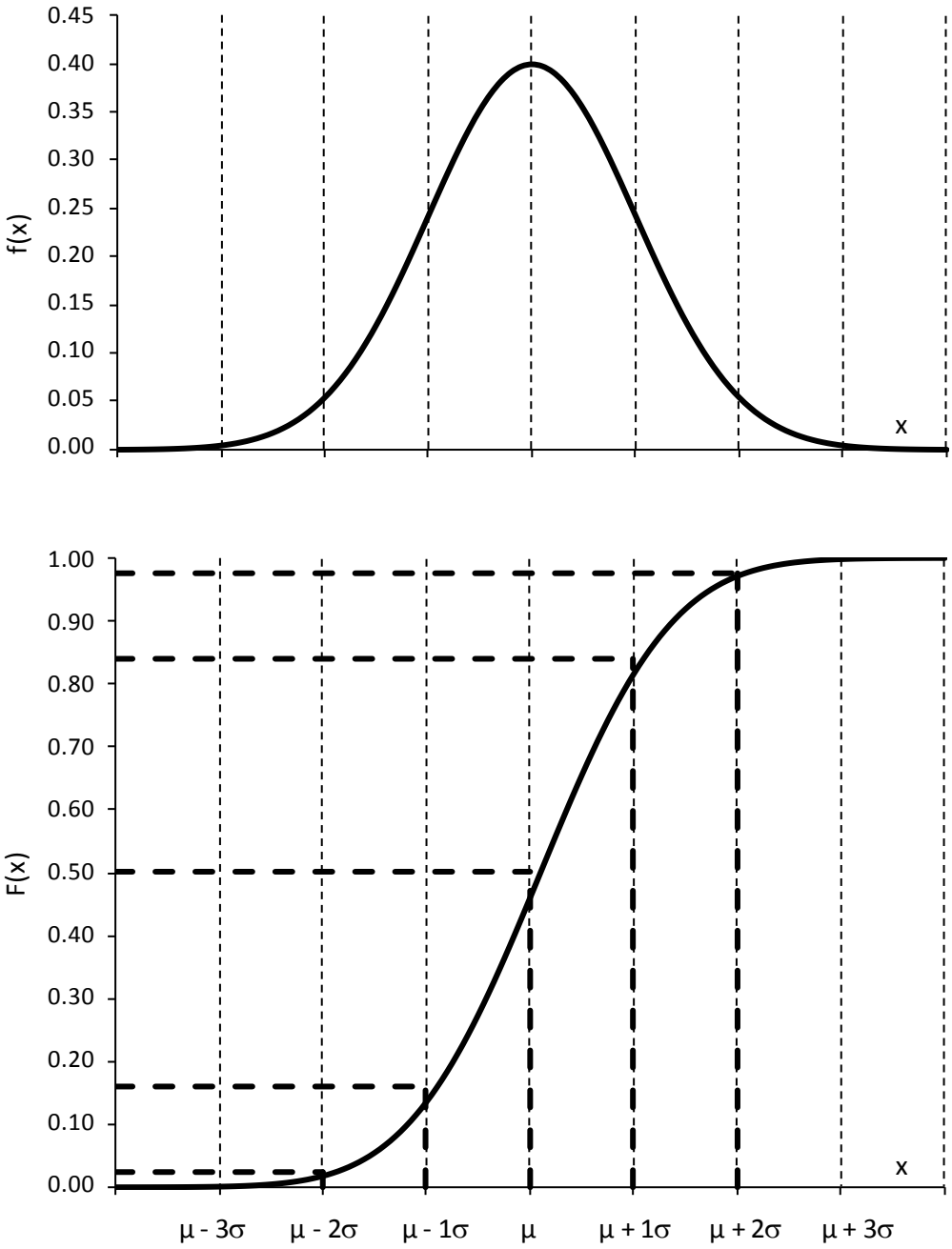



**Fig. 6.4:** Comparison between the probability density function (above) and the distribution function (below) of the normal distribution

From this illustration, it follows that the cumulative frequency is 50% at the mean $\mu$. The 0% line as well as the 100% line are only touched by the curve in infinite, theoretically speaking. At $\mu - 3 \cdot \sigma$ or $\mu + 3 \cdot \sigma$, however, the corresponding lines have already almost been reached. At $\mu - 3 \cdot \sigma$ the cumulative probability is 0.135%, at $\mu + 3 \cdot \sigma$ it is 99.865%.

One can easily see that the percentage of the distribution is approximately 99.73% between $\mu - 3 \cdot \sigma$ and $\mu + 3 \cdot \sigma$, namely 99.863% - 0.135%.

### 6.1.3 Standard normal distribution

The Gauss curves are only really practical when they have been standardized. That is understandable if one considers that for every chosen normal distribution, the associated Gauss curve can be drawn.

Standardization turns all Gaussian curves into a standard curve with mean $\mu = 0$ and standard deviation $\sigma = 1$. This is achieved through the following transformation:

$$u = \frac{x - \mu}{\sigma}.$$

By subtracting $x - \mu$ the mean is shifted toward the zero point. Dividing by the standard deviation effectively compresses (or stretches) the x-axis so that the standard deviation becomes 1. Figure 6.6 illustrates this transformation in scale.



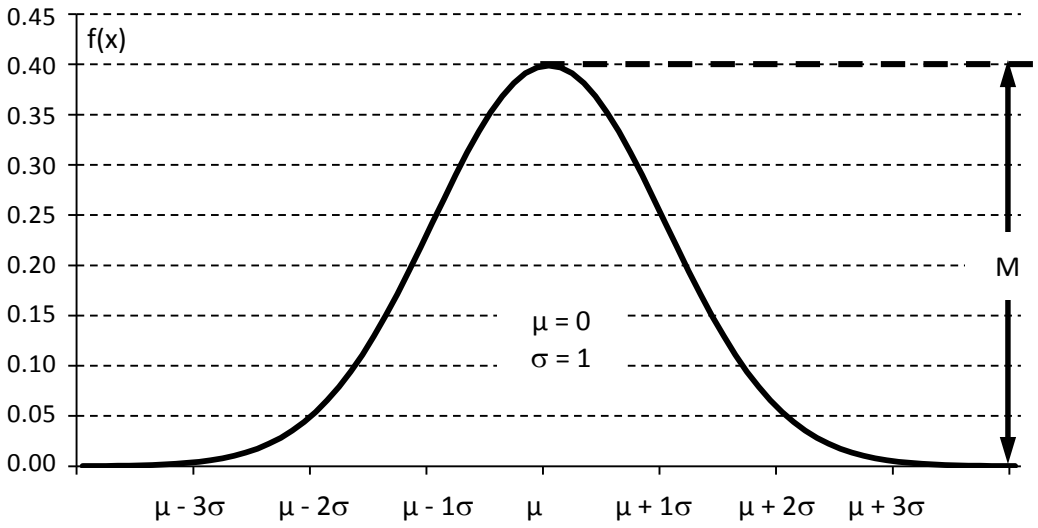



**Fig. 6.5:** Bell curve for standard normal distribution

When drawing the bell curve, one can use the following approximate values for a given choice of the maximum M in mm (see Figure 6.5):

| Abscissa | $\mu \pm 0.5 \cdot \sigma$ | $\mu \pm 1 \cdot \sigma$ | $\mu \pm 1.5 \cdot \sigma$ | $\mu \pm 2 \cdot \sigma$ | $\mu \pm 3 \cdot \sigma$ |
|---|---|---|---|---|---|
| Ordinate | $\frac{7}{8} \cdot M$ $\approx 0.88 \cdot M$ | $\frac{5}{8} \cdot M$ $\approx 0.63 \cdot M$ | $\frac{2.5}{8} \cdot M$ $\approx 0.33 \cdot M$ | $\frac{1}{8} \cdot M$ $\approx 0.13 \cdot M$ | $\frac{0.1}{8} \cdot M$ $\approx 0.01 \cdot M$ |

**Table 6.1:** Approximate values to graphically display the Gaussian curve

The advantage of standardization is that, for Gauss distributions with any $\mu$ and $\sigma$, the probability density and therefore also the cumulative function (cumulative frequency) only depend on the values of the variable $u$.

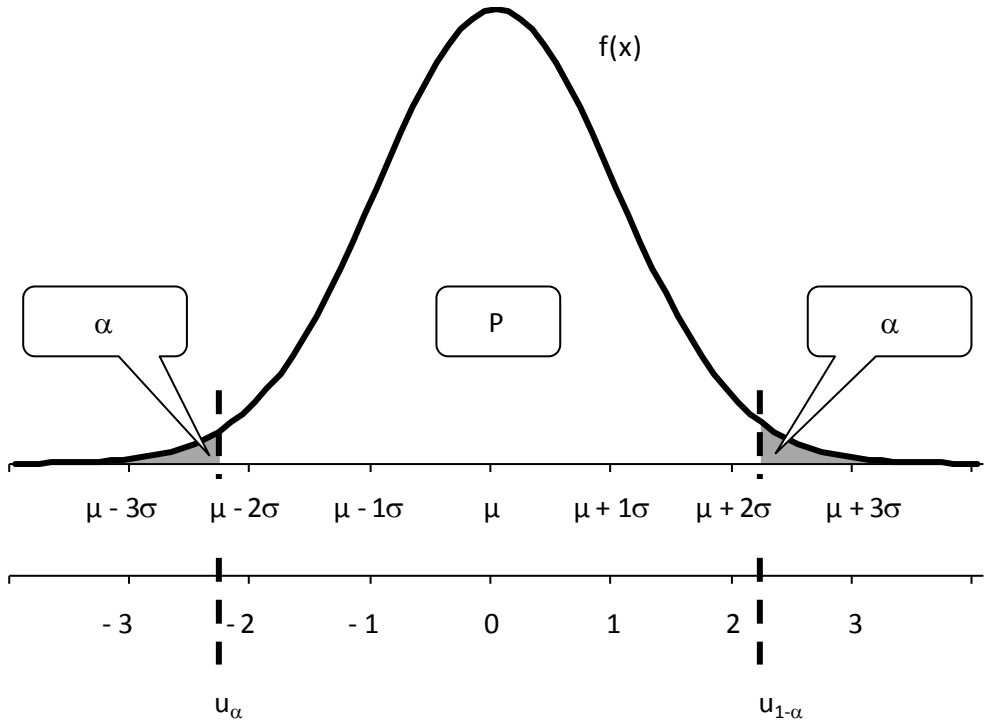



**Fig. 6.6:** Illustration of fractions non-conforming

Two new terms appear with regard to this representation:

- The quantity $P$, which gives the probability that a randomly measured value lies between $x_\alpha$ and $x_{1-\alpha}$, and
- the non-conforming fraction $\alpha$. This quantity corresponds to the probability of which the measured value is smaller than $x_\alpha$. Due to the symmetry, $\alpha$ is also equal to the probability that the measured value is greater than $x_{1-\alpha}$.

The area under the Gauss curve corresponds to $P + 2 \cdot \alpha = 1 = 100\%$.

For the interval $\mu \pm 3 \cdot \sigma$, the probability is $P = 99.73\%$ and therefore the one-sided non-conforming fraction is $\alpha = 0.135\%$ (because $2 \cdot \alpha = 0.27\%$).

So-called limits of variation located on the x-axis are assigned to the non-conforming fraction $\alpha$; the lower limit is designated with $x_\alpha$ and the upper limit with $x_{1-\alpha}$.

By rearranging the standardizing equation

$$u = \frac{x - \mu}{\sigma}$$

these limits of variation can be calculated easily. It follows: $x_{1-\alpha} = \mu + u_{1-\alpha} \cdot \sigma$.

Since the normal distribution is symmetrical, it also follows:

$x_\alpha = \mu + u_\alpha \cdot \sigma$ with $u_\alpha = -u_{1-\alpha}$ $\Rightarrow$ $x_\alpha = \mu - u_{1-\alpha} \cdot \sigma$.

The practical use of standardization can now be understood: For every Gaussian distribution with any given $\mu$ and $\sigma$, the proportion P of the distribution lies between the limits of variation $x_\alpha$ and $x_{1-\alpha}$.P

This general relationship between u and P is visible in the following table:

| u | 1.0 | 1.28 | 1.64 | 1.96 | 2.0 | 2.33 | 2.58 | 3.0 | 3.2 |
|---|---|---|---|---|---|---|---|---|---|
| P(±u) % | 68.3 | 80.0 | 90.0 | 95.0 | 95.4 | 98.0 | 99.0 | 99.7 | 99.9 |
| $\alpha$ % | 15.9 | 10.0 | 5.0 | 2.5 | 2.28 | 1.0 | 0.5 | 0.135 | 0.05 |

**Table 6.2:** Relationship between u and P

Chapter 10 includes another detailed table for the standard normal distribution. However, somewhat different designations are used there. D(u) corresponds to the quantity designated here with P and $\Phi(-u)$ is identical to $\alpha$.



## 6.2 The normal probability plot

The normal probability plot is a guide to visually determine statistical characteristics of a data set. It also offers the possibility to verify whether or not the values of a data set originate from a normal distribution.

The division of the y-axis of the normal probability plot is chosen in such a way that the data sets origination from a normal distribution provide a sequence of points that lie on a straight line. In visual terms, the cumulative curve from Figure 6.4 is “straightened” by the corresponding distortion of the y-axis. This is achieved by equally spacing the cumulative frequencies belonging to the integral multiples of the (dimensionless) quantities u on the y-axis. The abscissa axis, that is the x-axis (represented here for u, see Chapter 6.1.3), has a linear division.

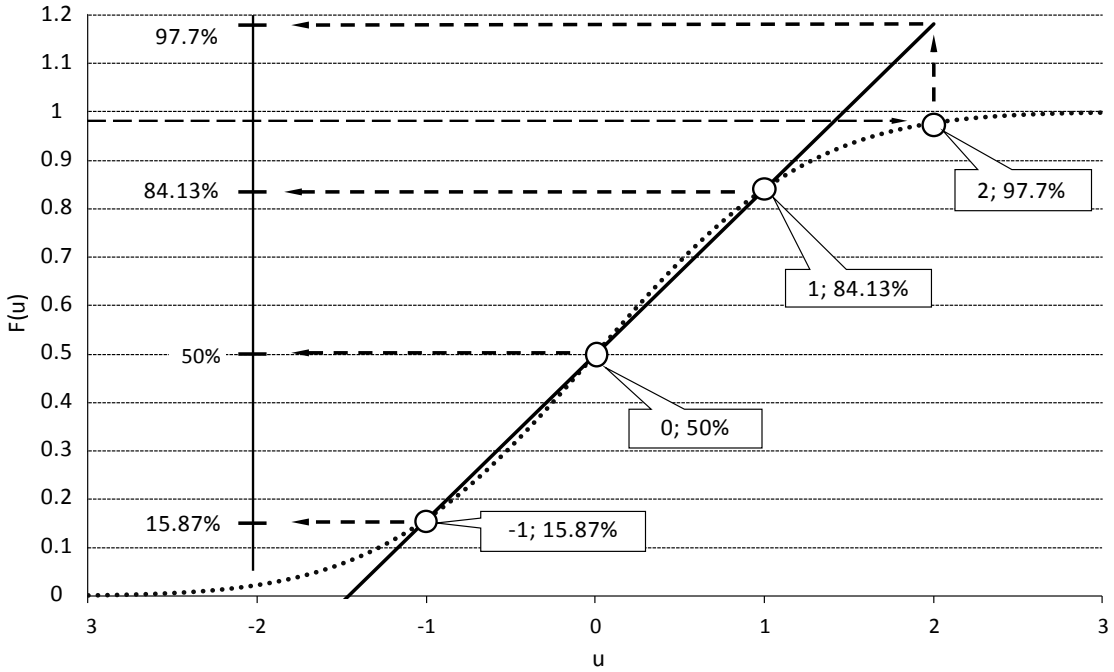



**Fig. 6.7:** Formation of the normal probability plot

Since the values for the relative cumulative frequency determined using a sample normally only approximate the theoretical normal distribution, the corresponding points in the probability plot are approximated using a straight line.

The use and practical benefit of the probability plot are explained in the following chapter.

**Frequency diagram with probability plot**

Using a normal probability plot, one can verify whether or not it is justifiable to adopt this distribution model for the data set in consideration. Moreover, the statistical measures of the data set can be identified/read easily.

| The relative cumulative frequency | equates to |
|---|---|
| 50% | the mean $\bar{x}$ |
| 99.865% | the value $\bar{x} + 3 \cdot s$ |
| 0.135% | the value $\bar{x} - 3 \cdot s$ |

These relationships form the basis for using the probability plot.

For series of measurements that consist of more than 25 measurements and are to be analyzed using grouping, one plots the values of the relative cumulative frequency on the upper class limit in the probability plot.

For smaller series of measurements, one first creates an ordered list (arranging the measurements according to size) and then assigns cumulative frequencies $H_i(n)$ to these values that can be calculated using the approximation formula $H_i(n) = \frac{i - 0.3}{n + 0.4}$ $(i = 1, 2, \ldots, n)$.

$x_{(1)}$, $x_{(2)}$, ... , $x_{(n)}$

$H_1(n)$, $H_2(n)$, ... , $H_n(n)$.

Finally, the points ( $x_{(i)}$, $H_i(n)$ ) are plotted on the probability plot.

A best-fit line is drawn through the series of points in order to minimize the distances of the points from the line and to place approximately as many points above and below the line. The better this line approximates this series of points, the more suitable the model of normal distribution is to describe the data set being evaluated.

*Example 6.1 provides more explanation about how to graphically determine statistical measures.*

*Graphical determination of the mean*

*One finds the mean $\bar{x}$ by finding the point where the horizontal line at 50% cumulative frequency intersects with the best-fit line and reading the associated value on the x-axis.*

*One finds $\bar{x} = 346$.*

*Graphical determination of the standard deviation*

*The standard deviation $s$ can be determined, by reading the $x$-values belonging to the cumulative frequencies 99.865% (corresponds to $\bar{x}+3\cdot s$) and 0.135% (corresponds to $\bar{x}-3\cdot s$), in the example at hand, these values are $x = 366$ and $x = 324$, and dividing their difference by 6.*

*One therefore finds: $s = \frac{366 - 324}{6} = \frac{42}{6} = 7$.*

*Graphical determination of the fraction non-conforming*

*In the present case, the upper limit is $UL = 360$ ("upper tolerance limit"). The theoretical proportion of the population that exceeds $UL$ is found by drawing a perpendicular line at this point and finding its intersection with the best-fit line. The cumulative frequency associated with this intersection corresponds to the proportion of the population that lies below the $UL$; here about 97.5%.*

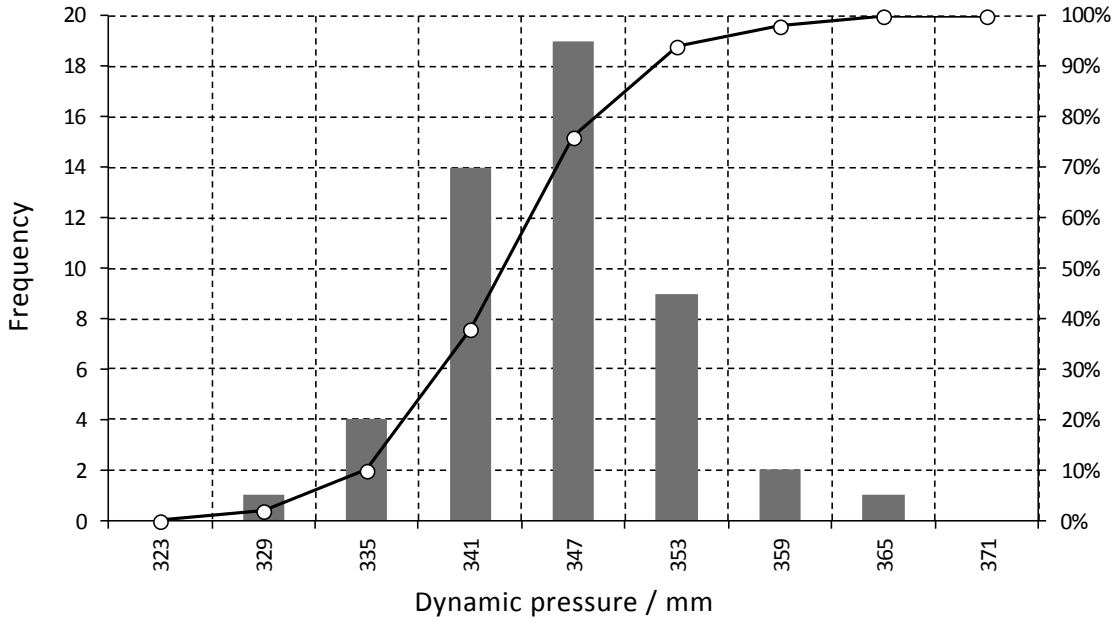
*The desired fraction non-conforming (right-hand side) equals the difference between that and 100%, therefore 2.5%.*

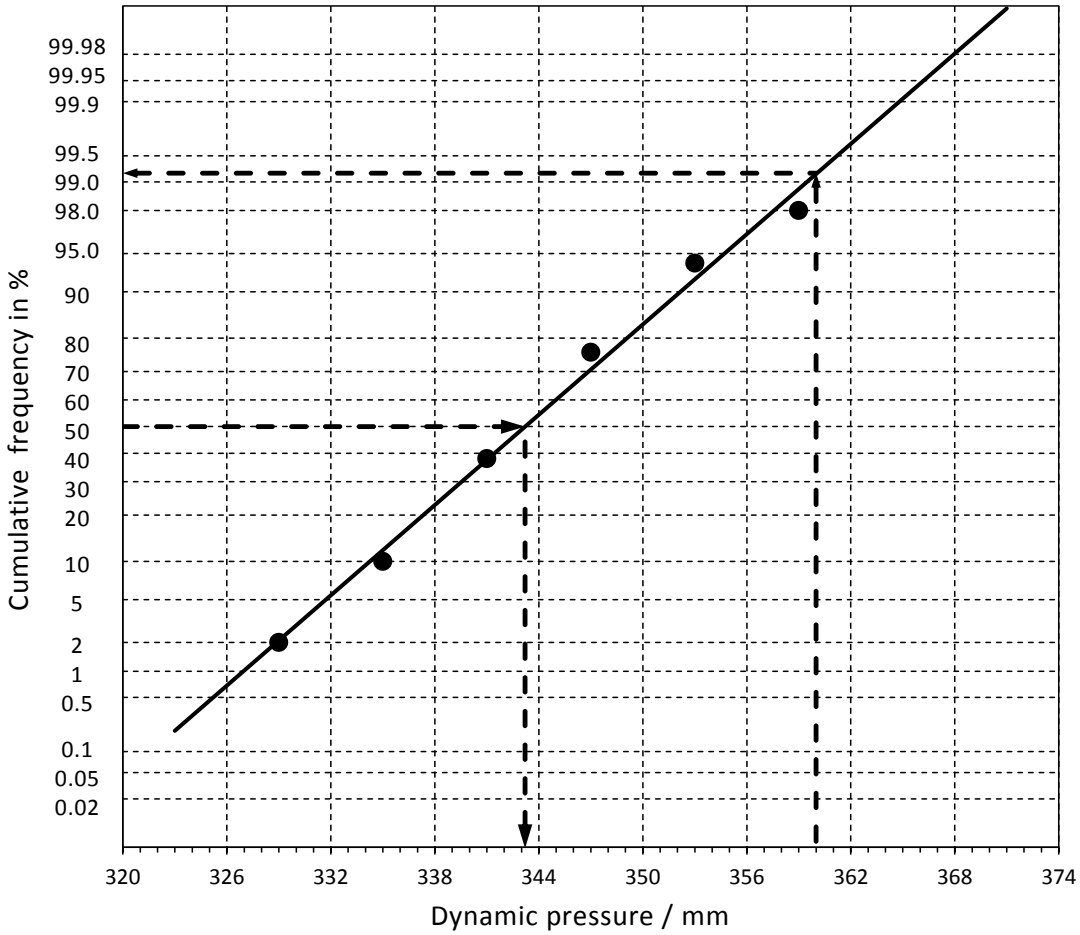
*The proportion that lies below a certain lower limit $LL$ ("lower tolerance level") is given in a similar manner. One therefore draws a perpendicular line at $LL$ and reads the associated cumulative frequency at the intersection with the best-fit line. This values corresponds directly to the desired fraction non-conforming (left-hand side).*

*The so-called two-sided fraction non-conforming is the sum of the lower (left-hand side) and upper (right-hand side) fractions non-conforming. It gives the proportion of units from the population whose values lie outside the tolerance range.*

*Example 6.1: Dynamic pressure values of valve plugs*

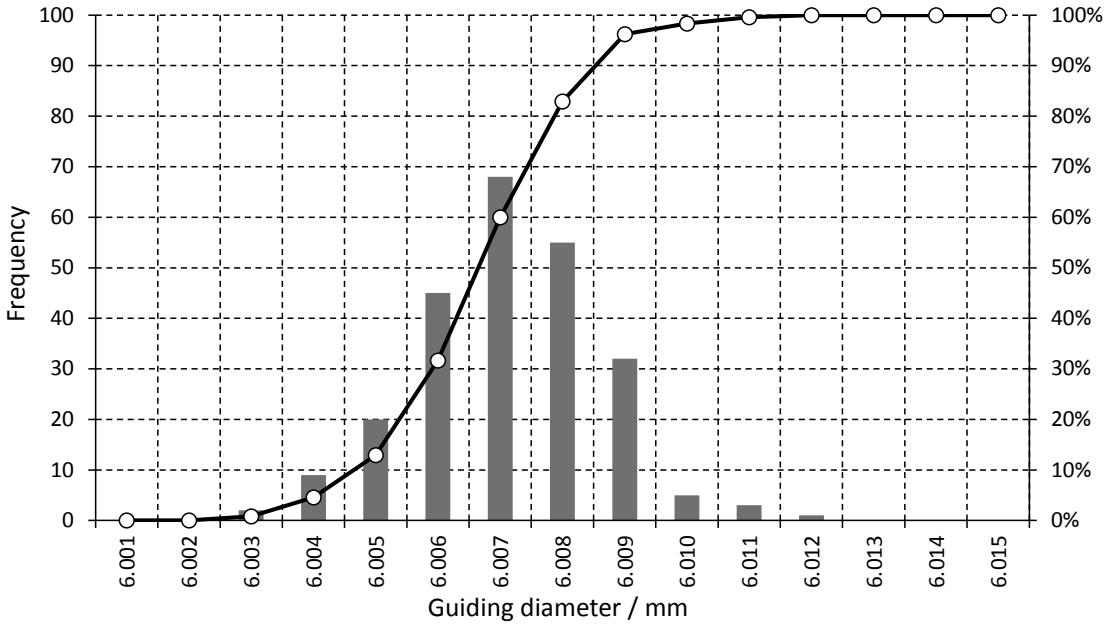
*The frequency distribution of a sample of 50 measured values is displayed here.*

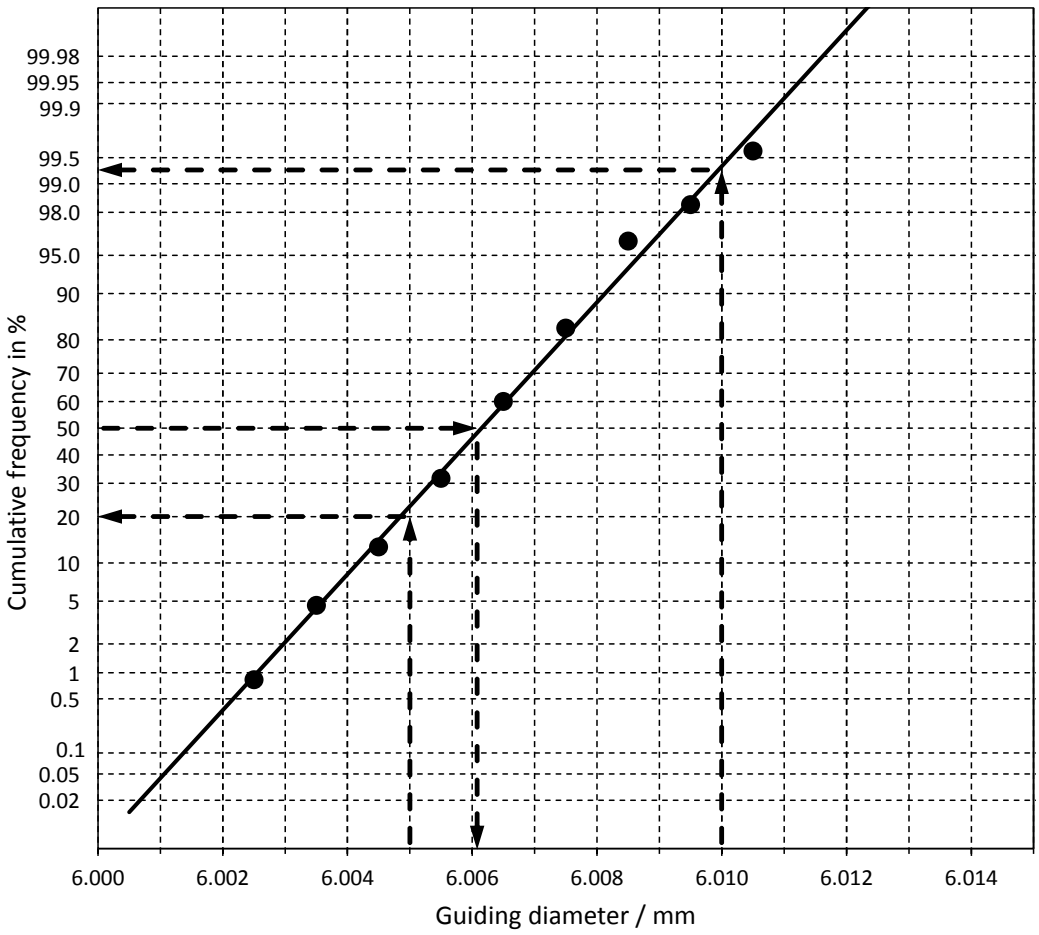


*Example 6.2: Guide diameter of injection needles*

*This "sample" includes $n = 240$ measured values.*



*EXAMPLE 6.3: Dynamic pressure values for valve plugs*

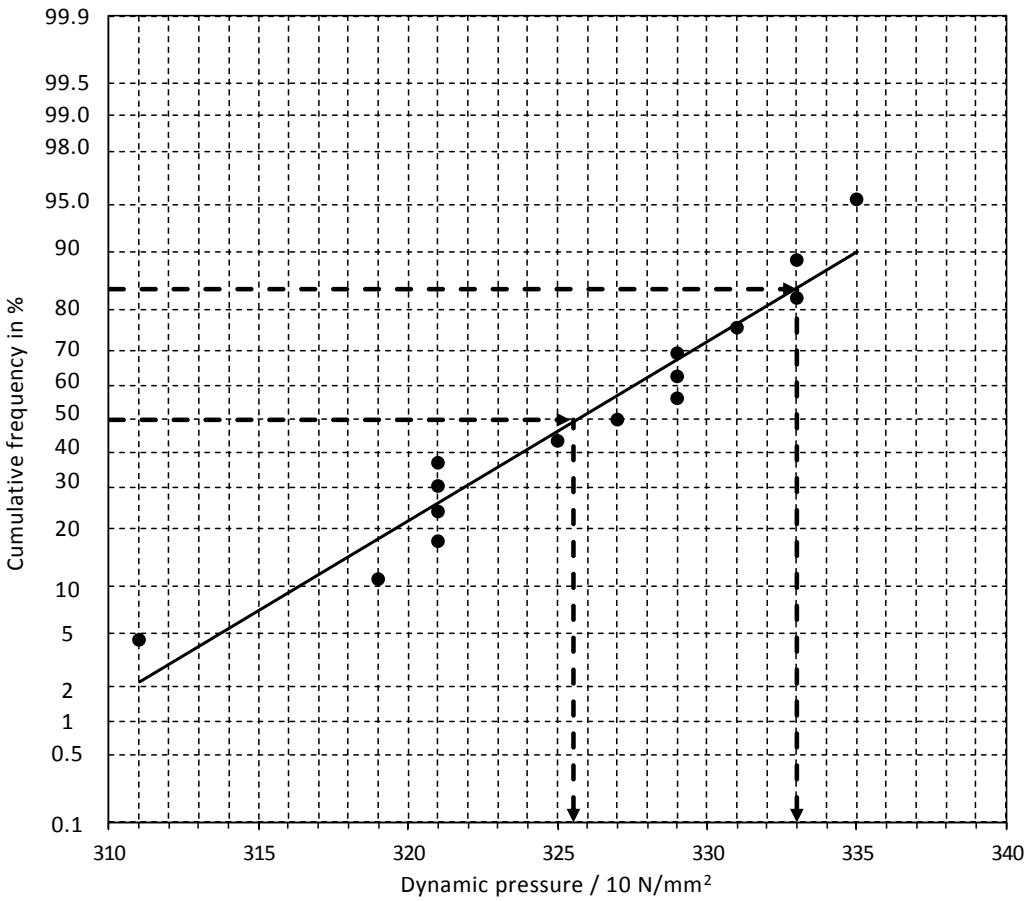
*This example shows how to evaluate a "small series of measurements" ( $n = 15$ measurements) in a probability plot. The following table contains the measurements $x_i$ (10 N/mm²) and the associated relative cumulative frequencies $H_i(n)$ to record the points ( $x_{(i)}$, $H_i(n)$ ) into the probability plot (in ascending order).*

*The latter were calculated with the approximation formula* $H_i(n) = \dfrac{i - 0.3}{n + 0.4}$ *for* $i = 1, 2, \ldots, 15$.

*Example:* $H_2(15) = \dfrac{2 - 0.3}{15 + 0.4} \approx 0.11 = 11\,\%$.

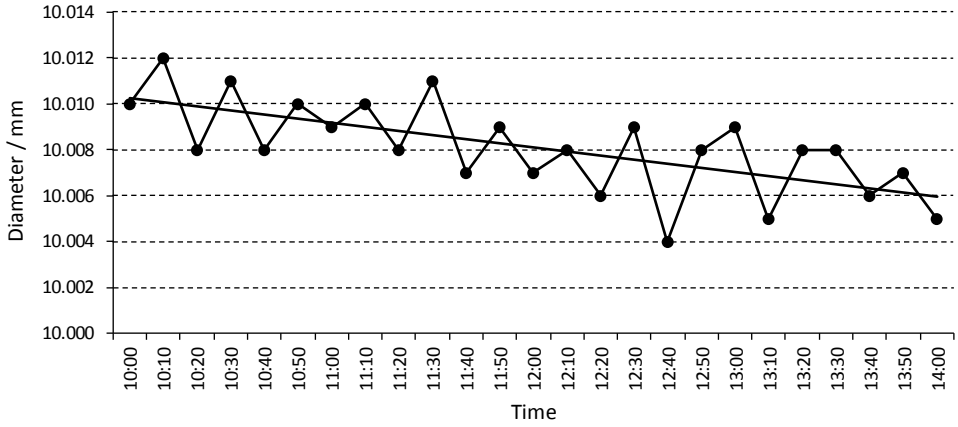
| No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x_i$ | 311 | 319 | 321 | 321 | 321 | 321 | 325 | 327 |
| $H_i(n)$ | 4.5 | 11.0 | 17.5 | 24.0 | 30.5 | 37.0 | 43.5 | 50.0 |
| No. | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| $x_i$ | 329 | 329 | 329 | 331 | 333 | 333 | 335 | |
| $H_i(n)$ | 56.5 | 63.0 | 69.5 | 76.0 | 82.5 | 89.0 | 95.5 | |

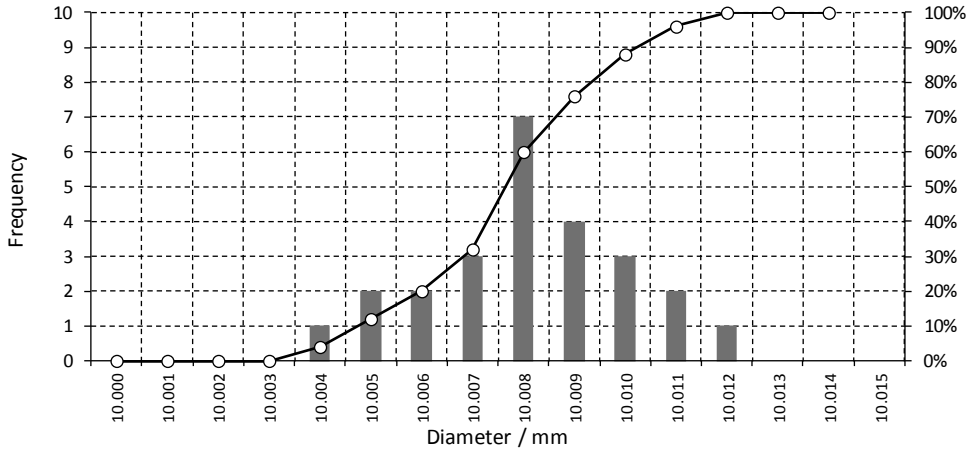
*The scale on the x-axis must be chosen carefully. On one hand, one should be able to draw the best-fit line completely so that it is possible to read the points of intersections on the upper and lower edge of the plot. On the other hand, the line should not run too steeply, in the interest of readability.*

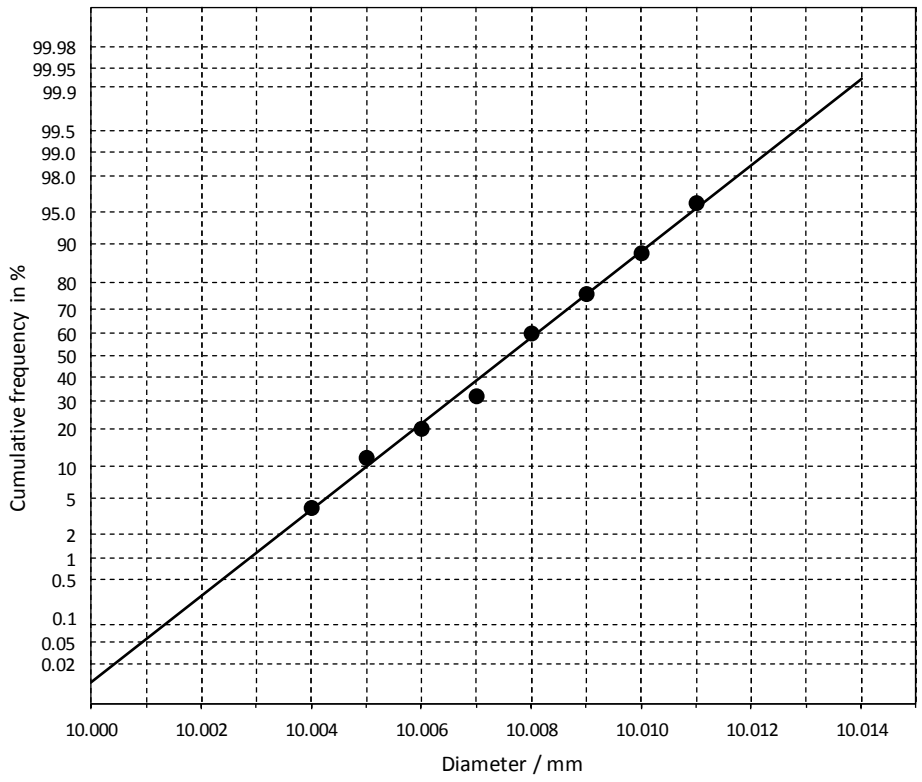


*EXAMPLE 6.4: Analysis of the process "bearing bush grinding"*

*In this case, the measurements were entered similar to an original value chart in chronological order. The frequency diagram is created in a simple manner from this representation.*



## 6.3 Lognormal distribution

If a characteristic cannot go below or above of a certain limit due to physical reasons, the result is commonly a so-called skewed (asymmetrical) distribution of the values. This is the case, for example, with characteristics like deflection and concentricity where the lower limit is “zero”, or the Rockwell hardness test for steel, where the value won't be less than a defined minimum hardness. Evaluating the frequency diagram of such a skewed distribution with the normal probability plot produces a curved line (Figure 6.8).

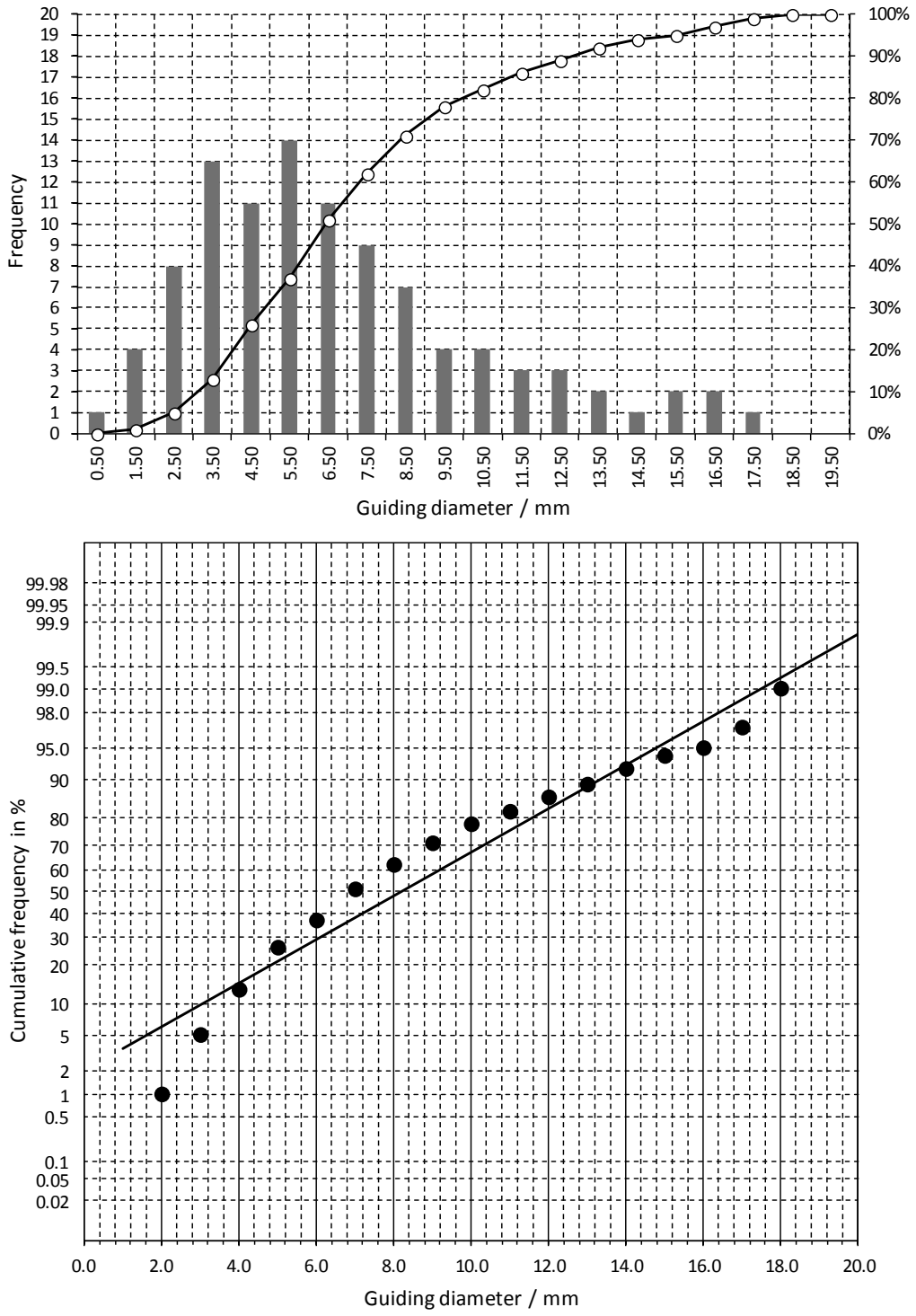



**Fig. 6.8:** Evaluating the skewed distribution using a normal probability plot

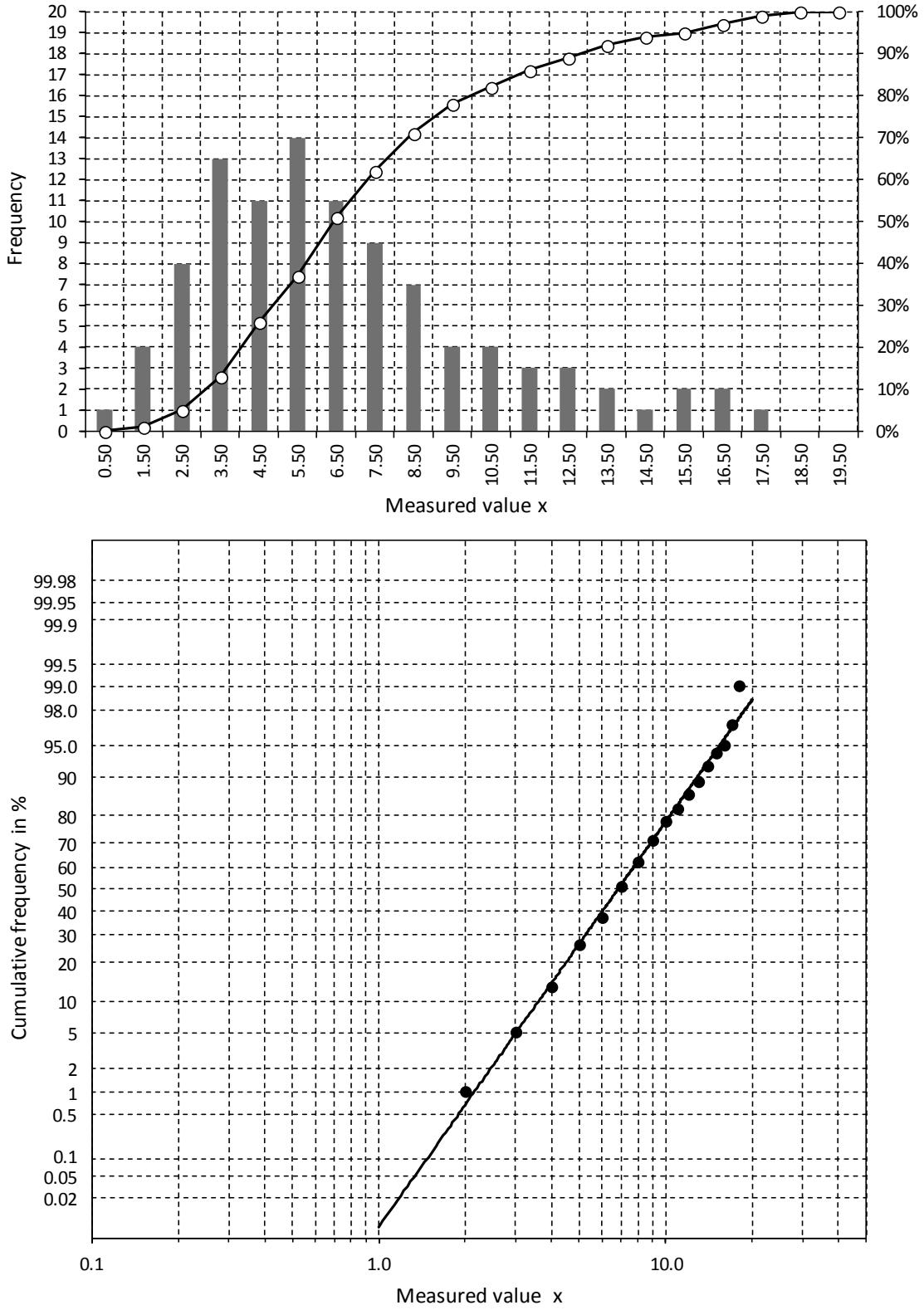



**Fig. 6.9:** The same data set using a lognormal probability plot

However, if the same data set is presented in the lognormal probability plot, the result is approximately a straight line of points (see Figure 6.9).

### 6.3.1 Lognormal probability plot

The ordinate scale of the lognormal probability plot is identical to the scale of a normal probability plot. The plots only differ in the scale division of the abscissa ( $x$ -axis). In case of the lognormal probability plot, this scaling is logarithmic. Because of the relationships given in the following table (the first two columns contain a different notation for the same number $x$ ), the intervals between the values corresponding to 10, 100, 1000, for example, are identical on an axis with a logarithmic scale.

| x | x | log(x) |
|---|---|---|
| $\frac{1}{100} = 0.01$ | $10^{-2}$ | -2 |
| $\frac{1}{10} = 0.1$ | $10^{-1}$ | -1 |
| 1 | $10^{0}$ | 0 |
| 10 | $10^{1}$ | 1 |
| 100 | $10^{2}$ | 2 |
| 1000 | $10^{3}$ | 3 |

This provides a simple way to display data sets that span over two or more orders of magnitude (see Figure 6.9). Note that a logarithmic scale cannot contain zero, since the corresponding value for zero would be "negative infinity".

The points are plotted like in a "normal" probability plot (see Chapter 6.2).

The measures for the lognormal distribution are the geometric mean $\bar{x}_g$ and the geometric standard deviation $\varepsilon$ ("epsilon"). $\bar{x}_g$ is the median, meaning that 50% of the individual data is lower and the other 50% is larger than this number. It is easy to determine graphically by looking for the intersection of the drawn best-fit line with the horizontal line at 50% and determining the corresponding $x$ -value.

The mean $\bar{x}$ is not identical to the most common value (mode) due to the asymmetry of the distribution, it lies between the mode and the median $\bar{x}_g$.

When determining the geometric standard deviation $\varepsilon$, one first determines the intersection of the best-fit line with the horizontal line on the probability plot at $\bar{x}_g \cdot \varepsilon$ (corresponds to a cumulative frequency of 84.13%), then reads the associated value on the $x$ -axis and finally divides this number by $\bar{x}_g$.

Similar to a normal distribution, one can determine areas using $\bar{x}_g$ and $\varepsilon$, which are limited by certain intervals of the curve of the density function and the $x$ -axis.

68.3% of all values of a lognormal distribution lie between $\frac{\bar{x}_g}{\varepsilon}$ and $\bar{x}_g \cdot \varepsilon$.

95.4% of all values of a lognormal distribution lie between $\frac{\bar{x}_g}{\varepsilon^2}$ and $\bar{x}_g \cdot \varepsilon^2$.

99.7% of all values of a lognormal distribution lie between $\frac{\bar{x}_g}{\varepsilon^3}$ and $\bar{x}_g \cdot \varepsilon^3$.

Generally, only the upper limit (UL) $UL$ is of interest with regard to zero-limited characteristics. In this case, the associated non-conforming fraction can be determined using the logarithmic probability plot. To do so, one draws a perpendicular line at the point $UL$ (on the $x$-axis) and determines from the intersection with the best-fit line the corresponding cumulative frequency $H(UL)$ in percent. The value $100 - H(UL)$ corresponds to the theoretical proportion of the distribution (in percent) that exceeds the limit.

Figure 6.10 shows an example of such an analysis. Note that the frequency diagram (above) and the probability plot (below) have different x-axis scales (linear and logarithmic division), and therefore the points in the probability plot won't lie perpendicular under the positions of the upper class limit of the frequency diagram.

Following the horizontal line at 50% cumulative frequency towards the right up to the intersection with best-fit line and then down perpendicularly, one finds the geometric mean $\bar{x}_g = 0.31$. Correspondingly, one finds the value $\bar{x}_g \cdot \varepsilon = 0.51$ at the intersection of the 84.14%-line with the best-fit line. Dividing both of these numbers results in the geometric standard deviation $\varepsilon = \frac{0.51}{0.31} = 1.65$. Following the perpendicular line at the upper limit $OGW = 1{,}0$, one finds the intersection with the best-fit line at 99 % cumulative frequency. The fraction non-conforming with regard to this limit is therefore 1%.

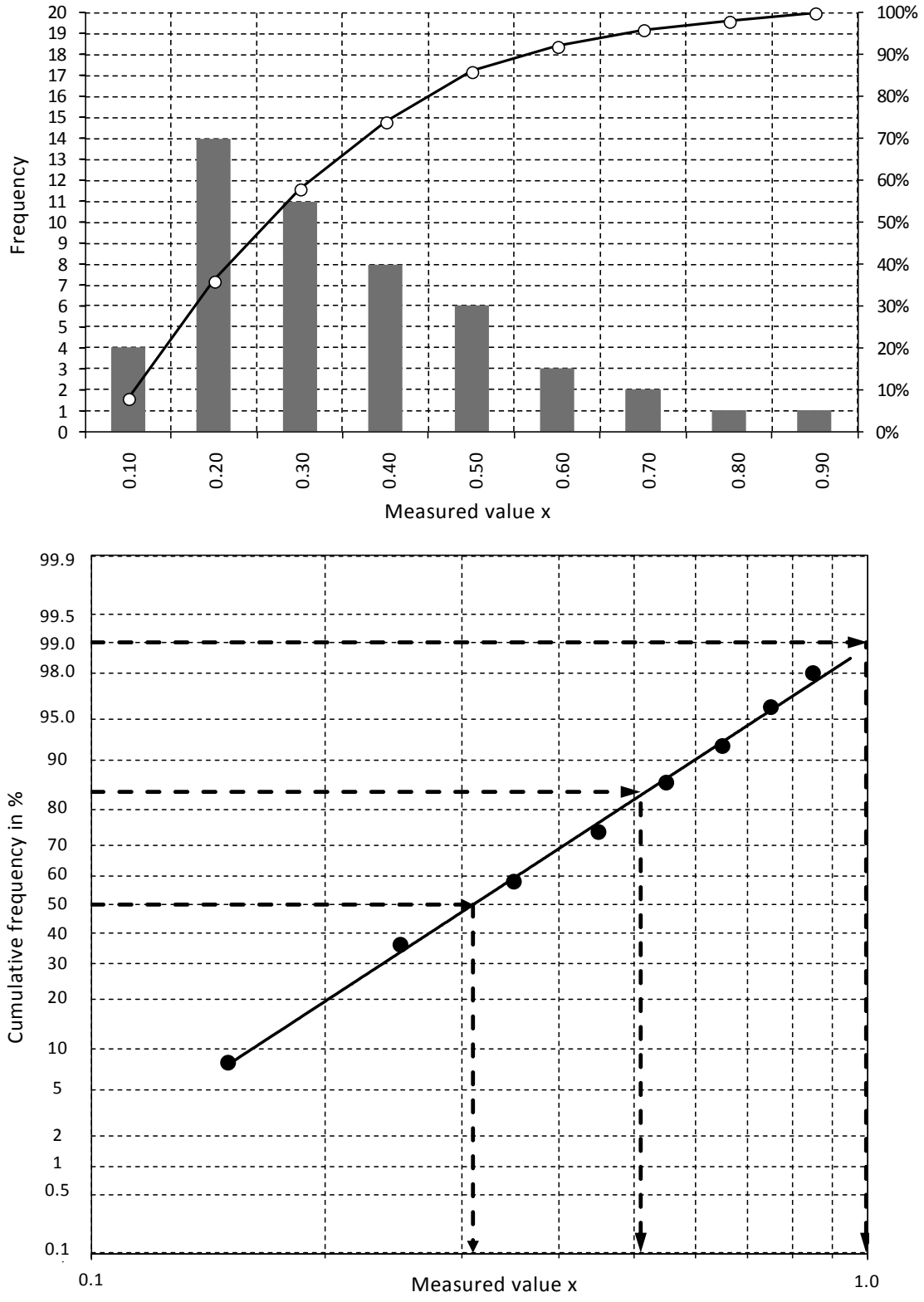



**Fig. 6.10:** Graphical analysis in a logarithmic probability plot

### 6.3.2 Relationship between normal distribution and lognormal distribution

By taking the logarithm, a lognormal distributed characteristic $x$ is transformed into a normal distributed characteristic $z$. Through the use of logarithms, the upper portion of the lognormal distribution is strongly compressed and the area between the geometric mean $\bar{x}_g$ and the zero point is strongly stretched. This corresponds to a "mirroring" of the curve of the function $z = \ln(x)$. Image 6.11 shows that $\bar{x}_g$ is mapped to $\bar{z}$ and the point $\bar{x}_g \cdot \varepsilon$ corresponds to the value $\bar{z} + s_z$. It is:

$$\bar{z} = \ln(\bar{x}_g) \quad \Leftrightarrow \quad \bar{x}_g = e^{\bar{z}} \quad \text{and} \quad s_z = \ln(\varepsilon) \quad \Leftrightarrow \quad \varepsilon = e^{s_z}.$$

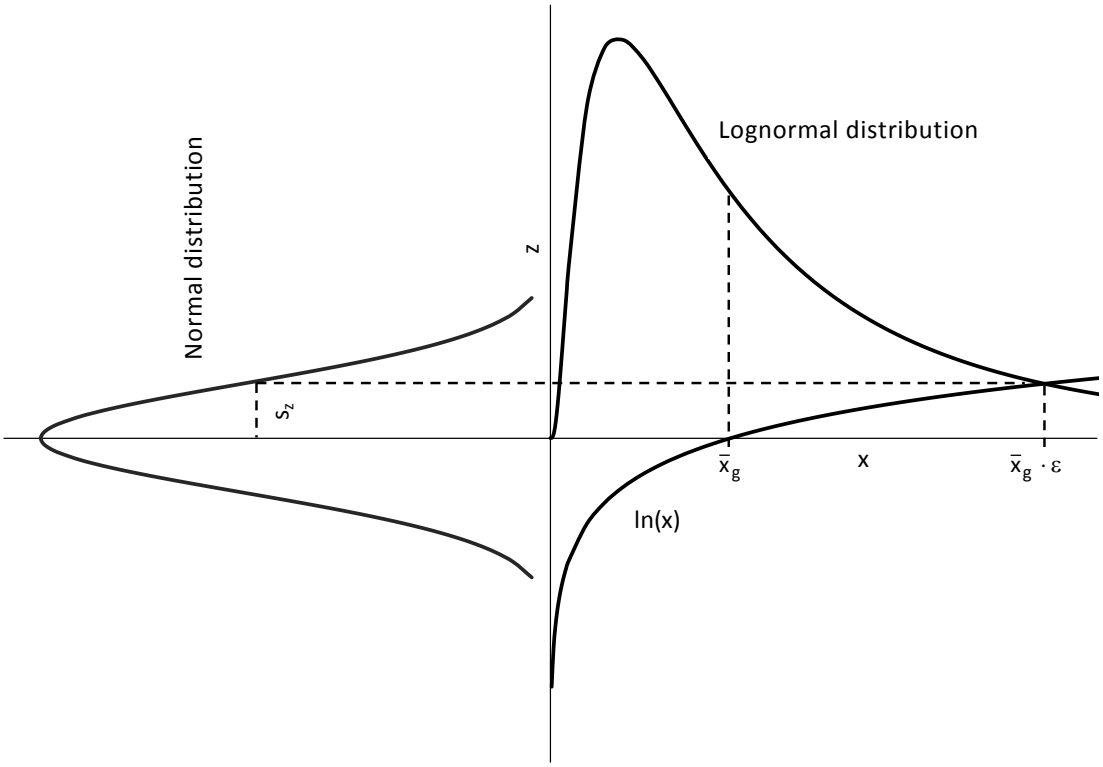



**Fig. 6.11:** Illustration of the relationship between the normal distribution and the log-normal distribution

*NOTE:*

*It may seem somewhat confusing that the preceding explanations are related to the natural log ln(x) (with base e), but the probability plot features a division based on the common logarithm log(x) (with base 10). This fact is, however, meaningless with regard to the comparability of the evaluation results, since the values are always transformed back into the original coordinate system during the determination of $\bar{x}_g$ and $\varepsilon$, and the calculation of fractions non-conforming or limits requires both of these quantities. If using log(x) as well as its inverse function $10^x$, the above expressions are replaced with $\bar{x}_g = 10^{\bar{z}}$ and $\varepsilon = 10^{s_z}$.*

## 6.4 Mixture distributions

It can happen that the dimensions of parts produced on two different machines or production lines exhibit different distributions. Normally one sees that central tendency and/or variation differ from one another although the distribution type is the same. If the parts are not separated, the resulting distribution is called in statistical terms a mixture distribution. Mixture distributions can also occur if the essential impact factors of a running production series change abruptly (tool change, different batch of materials).

The histogram of a mixture distribution normally has two or more maxima. One can also speak of a bimodal or multimodal distribution.

If the mixture distribution is caused by two superimposed normal distributions whose mean values strongly differ, a representation (of the cumulative frequency or individual values) in the probability plot shows that the resulting dot line may be approximated piecewise by two different straight lines.

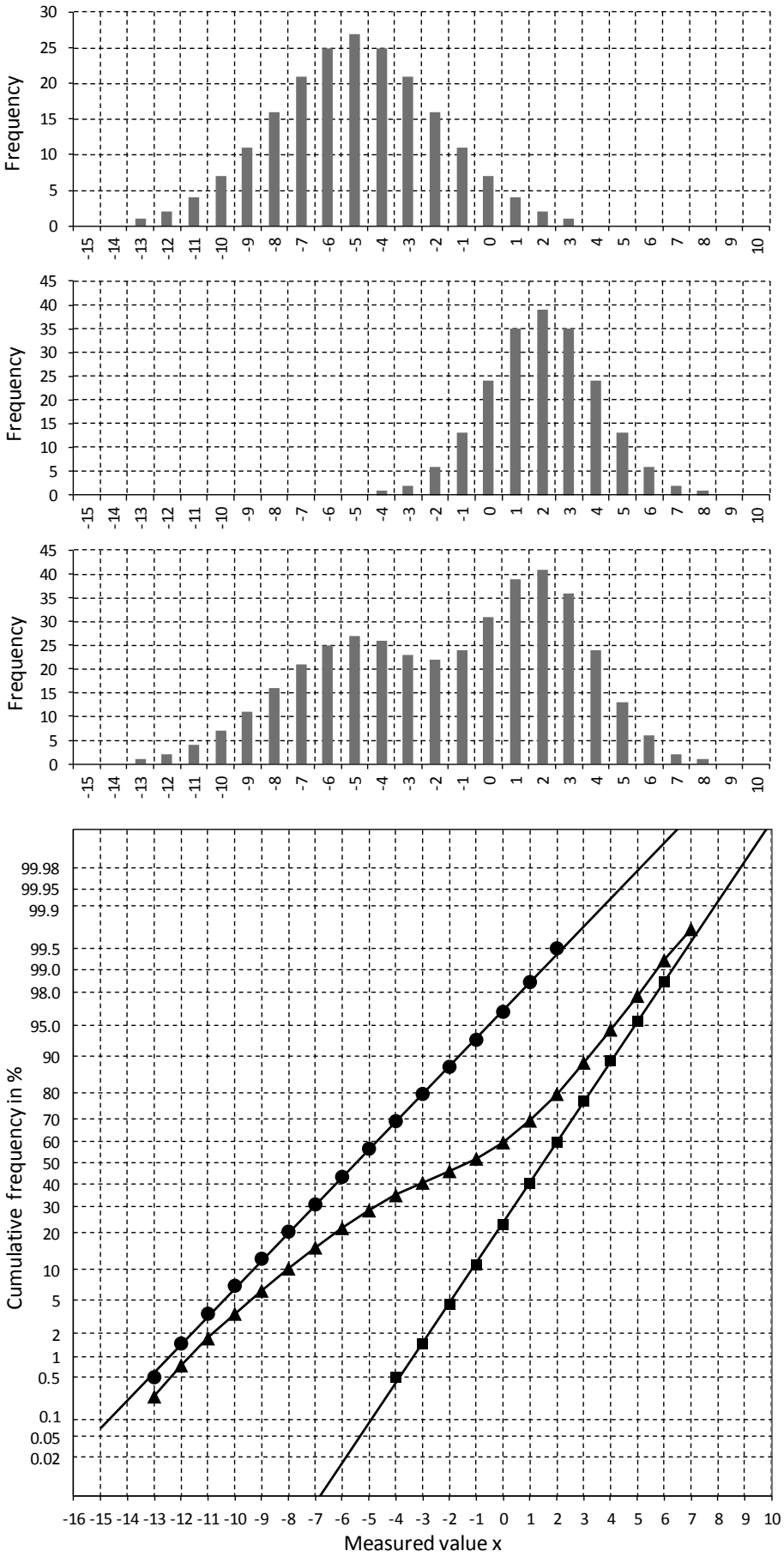



**Fig. 6.12:** Representation of a mixture distributions based on two collectives

# 7 Quality control charts

*NOTE:*

*This chapter presents only statistical fundamentals about control chart techniques. Special procedures and current control systems with regard to their practical application are described in Book 7 "Statistical Process Control (SPC)" of the Bosch Series "Quality Control in the Bosch Group, Technical Statistics".*

**S**tatistical **P**rocess **C**ontrol is best practice for controlling a production process based on statistical methods.

Thereby samples of parts are drawn from the process according to process-specific sampling rules and the values are measured and entered in so-called quality control charts. The statistical measures calculated from the values are then used to assess the current state of the process. If necessary, the process status will be corrected with appropriate measures.



The control chart technique was developed by Walter Andrew Shewhart (1891-1967) in the nineteen-twenties and described in detail in 1931 in his work "Economic Control of Quality of Manufactured Product".

SPC is a process drawn from inductive (conclusive) statistics. Not all measurement values are made available as would be the case if inspection 100%. Conclusions about the population are based on a small data set, the sample values.

The mathematical model for variable quantities is based on the idea that there are many quantities that have an influence on a process. The "5 M" — Man, Machine, Material, Milieu, Methods — are the main categories of these influencing quantities.

Each "M" can be further subdivided, e.g. Milieu (= environment) breaks down into temperature, humidity, vibration, contamination, illumination, ...

Uncontrollable, random effects of many influencing quantities lead to deviations of the real values from the target value, despite careful procedures (generally the midpoint of the tolerance range).

The random interaction of many influencing quantities in general results in an approximate Gaussian normal distribution for the observed part characteristic. This fact is described in statistics by the central limit theorem. The normal distribution is therefore of fundamental importance for SPC.

## 7.1 Location control charts

One can obtain initial insight into the status of a process by taking a completed part as a random sample, measuring the desired characteristic and then comparing the obtained value with the estimation of the mean $\hat{\mu}$. If the individual value lies outside the range of, for example, $\mu \pm 3\sigma$, this result will not be a surprise, because under the requirements mentioned for purely random process behavior, approximately 99.73% of all values are supposed to lie within this range. One might already conclude from the obtained result that the current process average corresponds to the pre-production run (estimated value $\hat{\mu}$), or that there is no indication of a change to the process location.

However, confidence in such a statement would be significantly greater, if it were not based on an individual sample value , but on, e.g. $n = 5$ values. The preceding thoughts form the basis for the function of a quality control chart. A conclusion about a momentary process location is drawn from the results of a current 5-piece sample. Now, there is still the unanswered question, which measure(s) should be used as the information media regarding the process location.

Besides the possibility of considering all the five individual values separately, it is recommendable to undertake an assessment based on the location of the mean value $\bar{x}$ or the median (central value) $\tilde{x}$ of the five values. These three options correspond to three different control charts for the location: the original value chart, the mean chart and the median chart. We want to limit ourselves in this documentation to representations of the mean chart and the original value chart.



### 7.1.1 Mean chart

The mean chart is the most important and the most commonly used quality control chart in practice. The following shall explain how to use it.

At constant time intervals, samples of size $n = 5$ are drawn from a production process, the characteristic to be monitored is measured and the five individual measurements as well as their standard deviation $s$ and mean $\bar{x}$ are entered into the quality control chart.

For now we only look at the means of the sample $\bar{x}_i$ in chronological order. The standard deviations $s_i$ will be analyzed using the $s$-chart (Chapter 7.2.1).

The calculated means are entered into a diagram and connected with a line.

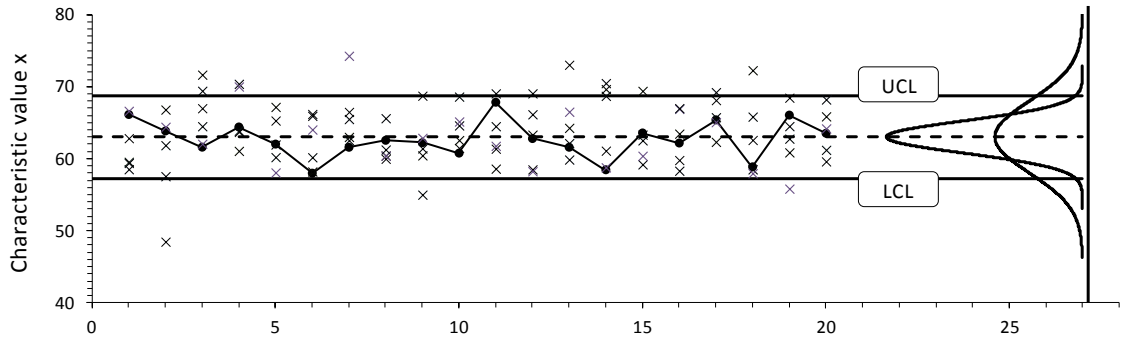



**Fig. 7.1:** Schematic about the functionality of a mean chart. In order to illustrate the relationships, the mean values (points) as well as the individual values (crosses) from each sample are represented.

The means like the individual values show a variation around the process average $\mu$; however, the variation of the means is smaller by the factor $\frac{1}{\sqrt{5}}$ than that of the individual values. The following figure illustrates this relationship.

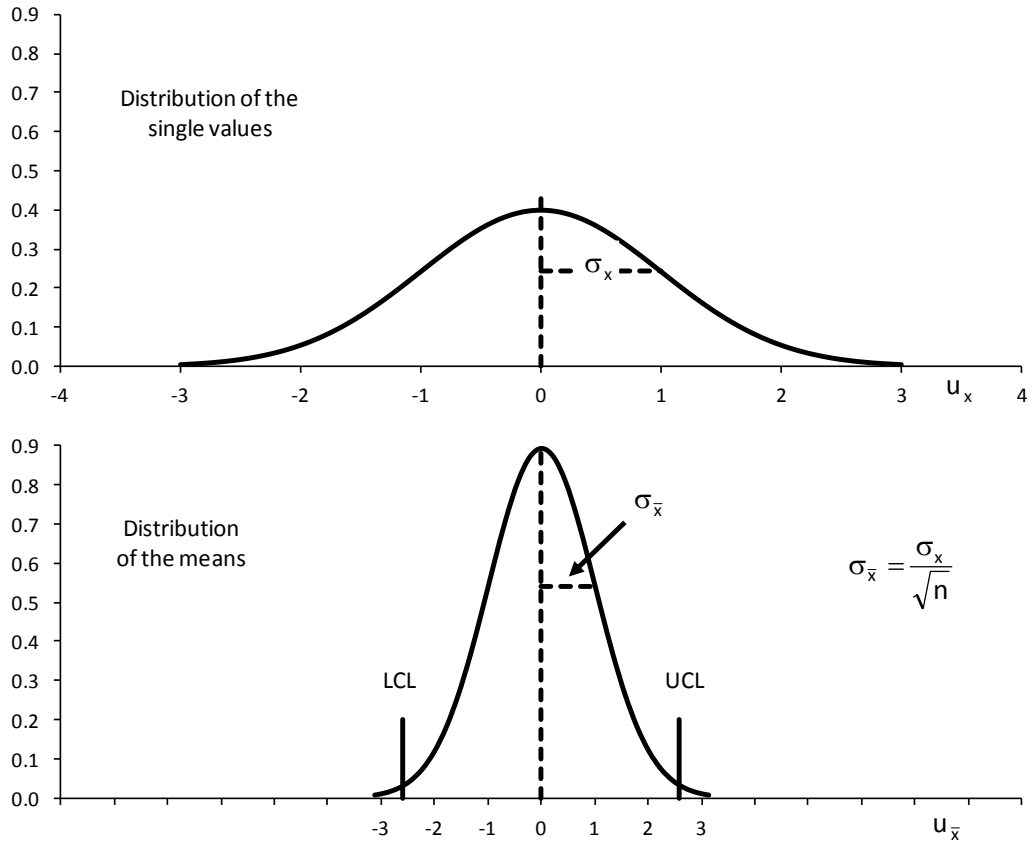



**Fig. 7.2:** Relationship between the variation of the individual values (original values) and the variation of the means.

Generally: If the individual values of a process characteristic are dispersed with the standard deviation $\sigma$ around the mean $\mu$, then the standard deviation of the means $\bar{x}$ of $n$ values is equal to $\frac{\sigma}{\sqrt{n}}$.

We are now looking at the general case of a sample with the size $n$.

One can now easily locate the random variation interval of the means by using the fact that the transformation $u = \frac{x - \mu}{\sigma}$ converts a normally distributed variable $x$ into a standard normally distributed variable $u$ (see Chapter 6.1.3), and instead of $x$ with the standard distribution $\sigma$ it actually enters the means $\bar{x}$ with the standard deviation $\frac{\sigma}{\sqrt{n}}$:

$$u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

The table of the standard normal distribution provides the limits $u_{lower} = -2.58$ and $u_{upper} = +2.58$ for the (two-sided) 99% random variation interval of the quantity $u$. Thus, with a probability of 99% $u$ lies within the limits -2.58 and +2.58:

$$-2.58 \leq u \leq +2.58$$

Substituting $u$ provides the 99% random variation interval of the mean $\bar{x}$ of a sample of size n:

$$-2.58 \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le +2.58 \quad \Leftrightarrow \quad \mu - 2.58 \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

If one also substitutes the unknown quantity $\mu$ with the nominal value $C$ and the also unknown $\sigma$ of the population with the corresponding estimation value $\hat{\sigma}$ (determined in a process capability analysis with at least 20 samples of 5 parts each), then one obtains:

$$C - 2.58 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \le \bar{x} \le C + 2.58 \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

a central relationship with regard to all of the location control charts.

The quantities

$$UCL = C + 2.58 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \quad \text{and} \quad LCL = C - 2.58 \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

are called upper and lower control limit for the mean $\bar{x}$. Since they only depend on the process variation $\sigma$ and are independent of the characteristic tolerance, one speaks of natural or process related control limits.

They limit the interval containing 99% of all means of $n$ individual values respectively. These control limits are drawn in the control chart (see Figures 7.1 and 7.2) as horizontal lines. All previous considerations required that the process conditions are stable. If an observed mean $\bar{x}_i$ exceeds the upper or falls below the lower control limits, one can conclude that the process conditions are no longer stable; the process has shifted significantly and must be corrected by the machine operator.

### 7.1.2 Original value chart (x-chart)

Practical experience shows that it is desirable in some cases to evaluate the process location directly using the individual values from the sample. This is shown by the original value chart.

The natural control limits are derived by considering the probability that all individual values from a sample of size $n$ lie within these limits.

It was shown in the chapter about the mean chart that, under certain conditions (normally distributed characteristic, stable process), there is a 99% chance that an individual value lies within the interval between

$$C - 2.58 \cdot \sigma \quad \text{and} \quad C + 2.58 \cdot \sigma$$

Since, for example, 5 sample values are random results that are independent from one another, the probability that all 5 values lie within this interval is equal to the product of the individual probabilities, thus

$$P_{total} = 0.99 \cdot 0.99 \cdot 0.99 \cdot 0.99 \cdot 0.99$$

$$P_{total} = 0.99^5 \approx 0.95.$$

In order to calculate the natural control limits for the original value chart with $n = 5$, one merely has to set the total probability $P_{ges} = 0{,}99$ and transform the equation

$$P_{total} = (P_{individual})^5 = 0.99 \quad \text{into}$$

$$P_{individual} = \sqrt[5]{0.99} = 0.998.$$

They can, therefore, be obtained from the standard normal distribution table as (two-sided) 99.8% random variation limits. For the example $n = 5$, this is the value $u = 3.09$, and the control limits are

$$UCL = C + 3.09 \cdot \hat{\sigma}$$

$$(n = 5)$$

$$LCL = C - 3.09 \cdot \hat{\sigma}.$$

If generalized to samples of size $n$, the natural control limits for the original value chart can be calculated as follows

$$UCL = C + u_{\sqrt[n]{0.99}} \cdot \hat{\sigma}$$

$$LCL = C - u_{\sqrt[n]{0.99}} \cdot \hat{\sigma}.$$



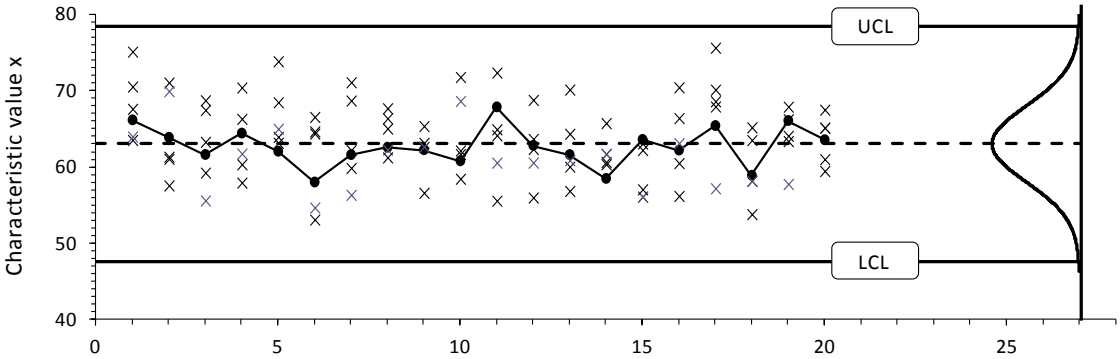



**Fig. 7.3:** Schematic of the functionality of an original value chart. For illustrative purposes, both the largest and smallest individual value from each sample are connected by a line (same data as in Figure 7.1).

## 7.2 Variation control charts

Just like the process average, the process variation, in general synonymous with the standard deviation of a part characteristic, is a key measure to evaluate production quality. Hereby, recognizing increases in variation is just as important as decreases in variation. The latter provides the chance to find out the reason for the short-term improvement and to maintain the favorable process conditions permanently.

Quality control charts for variation are suitable tools to determine such changes. For this purpose, the samples of size $n$ that were used in association with the mean chart are going to be used. In addition to the information about the current process location, each group of individual sample values also contains information about the current process variation.

One can consider the standard deviation $s$ or the range $R$ of each sample as the measure of this information.

### 7.2.1 s-chart

As already indicated, it is convenient not to regard the $s$-chart as a separate chart, but rather to display it as a second diagram on the form sheet for the mean chart. One calls this combination the $\bar{x}$-$s$-chart.

In parallel to representing the mean of each $i$-th sample of 5 (or generally samples of size $n$), here its standard deviation $s_i$ is recorded.

The size of the current value of this standard deviation $s_i$ naturally depends on the mean process variation estimated by $\hat{\sigma}$ and can by chance in a special case be somewhat greater or smaller than a long-term mean $s$. In order to decide when to interpret such fluctuations of $s$ as indications of an actual change to the process variation, and not as a random occurrence, one needs limits for $s$ that contain, e.g. the 99% random variation range of $s$. If the actual value of $s$ is greater than the upper limit or smaller than the lower limit, it is an indication of significant changes to the process variation. Such limit values can be determined by considering the fact that the quantity $f \cdot \frac{s^2}{\sigma^2}$ is subject to a $\chi^2$-distribution (called: chi-squared).

From $f \cdot \frac{s^2}{\sigma^2} = \chi^2$ with

$$s_{up} = \sqrt{\frac{\chi^2_{f,1-\alpha/2}}{f}} \cdot \sigma \quad \text{and} \quad s_{lo} = \sqrt{\frac{\chi^2_{f,\alpha/2}}{f}} \cdot \sigma$$

we obtain the desired control limits for $s$.

The $\chi^2$-distribution with $f$ degrees of freedom ($f$ corresponds to the sample size minus 1, thus $n-1$) is tabulated like the standard normal distribution. However, the $\chi^2$-distribution is not symmetrical.

This results in different factors to calculate the upper and lower control limits. The square-root terms in the above expressions are, to simply matters, designated with $B'_{Eup}$ and $B'_{Elo}$ and given directly with the sample size $n$ in the table below. The equations for the control limits of the $s$-chart are thus:

$$UCL = B'_{Eup} \cdot \hat{\sigma}$$

$$LCL = B'_{Elo} \cdot \hat{\sigma}.$$

| n | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $B'_{Eup}$ | 2.807 | 2.302 | 2.069 | 1.927 | 1.830 |
| $B'_{Elo}$ | 0.006 | 0.071 | 0.155 | 0.227 | 0.287 |

**Table 7.1:** Factors to calculate the control limits of the $s$-chart

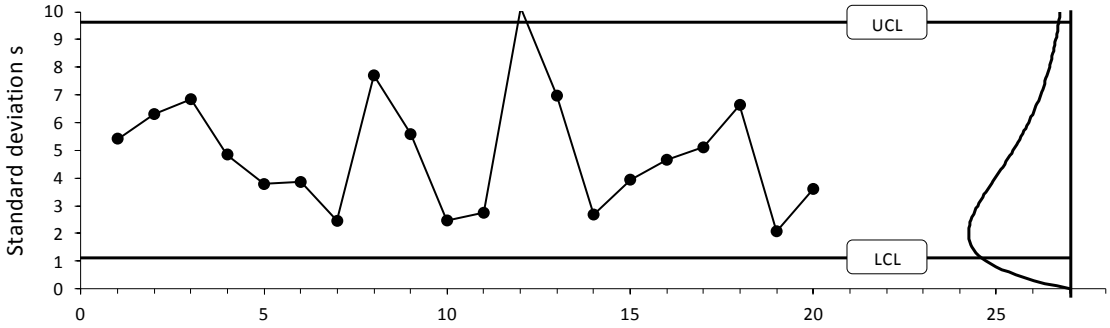



**Fig. 7.4:** Schematic of an s-chart (same data as in Figures 7.1 and 7.3). The standard deviations obey a skewed distribution. Since the control limits for $s$ are calculated using those for $s^2$, the surface areas above UCL or below LCL in the illustration are depicted too large.



### 7.2.2 R-chart

With R-charts, one uses the range $R$, therefore the difference between the largest and smallest individual values of a 5-piece group (or a sample of size $n$) as a measure for the momentary process variation.

The basis for calculating the variation limits of $R$ is given by the distribution of the standardized range $w_n$.

This distribution can, for example, be simulated using a computer by repeatedly "drawing" a random sample of size $n$ from a population of standard normally distributed values and determining its range $R$. The 99% random variation range of this quantity $R$ is then given by the upper limit $w_{n;0.995}$ and the lower limit $w_{n;0.005}$.

Since, when using the range chart, it is appropriate to estimate the process standard deviation $\sigma$ via the average range $\overline{R}$ (see Chapter 3.9)

$$\hat{\sigma} = \frac{\overline{R}}{d_n},$$

it is beneficial to combine the factors $w_{n;0.005}$ or $w_{n;0.995}$ and $\frac{1}{d_n}$ and denote the resulting factors $D_{Eup}$ or $D_{Elo}$. These quantities are tabulated below and allow one to calculate the control limits for the range chart according to

$$UCL = D_{Eup} \cdot \overline{R}$$

$$LCL = D_{Elo} \cdot \overline{R}.$$

| n | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $D_{Eup}$ | 3.518 | 2.614 | 2.280 | 2.100 | 1.986 |
| $D_{Elo}$ | 0.008 | 0.080 | 0.166 | 0.239 | 0.296 |

**Table 7.2:** Factors to calculate the control limits of the $R$-chart

# 8 Evaluating frequency distributions in connection with a tolerance

The following illustrations show schematics of examples of possible frequency distributions, which may occur when examining a production process; short explanations are included.



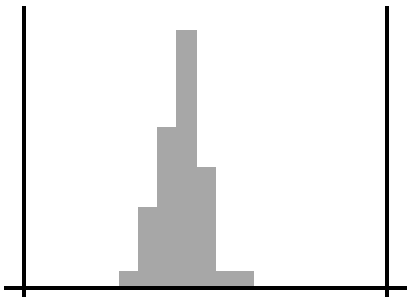


**Fig. 8.1:**

Variation range considerably smaller than tolerance. Mean corresponds well with the tolerance average.

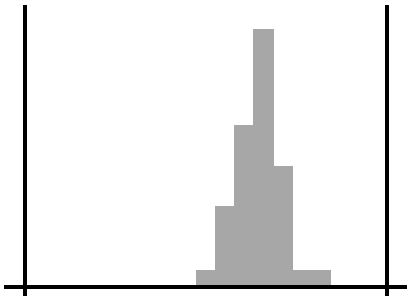


**Fig. 8.2:**

Variation range considerably smaller than tolerance. However, mean lies outside the tolerance average. Rejection rate may go up.

Center process!

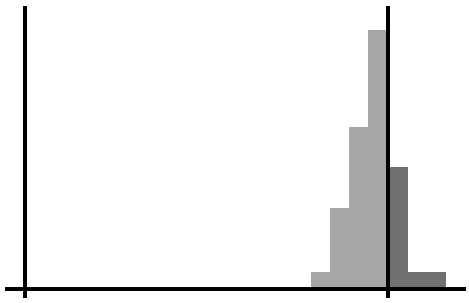


**Fig. 8.3:**

Variation range considerably smaller than tolerance. However, mean of the distribution is far removed from the tolerance average.

Center process!

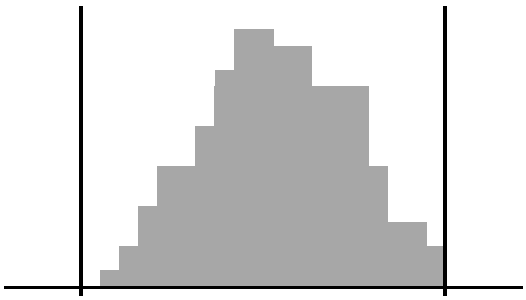


**Fig. 8.4:**

Variation range is about the same as the tolerance.

Mean of the distribution corresponds well with the tolerance average. However, a systematic shift of the mean can lead to deficient product.

Reduce variation!



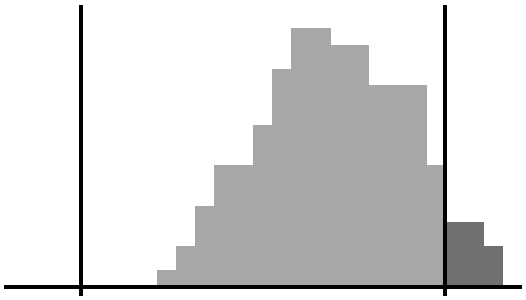


**Fig. 8.5:**

Variation range is about the same as the tolerance.

However, mean of the distribution does not coincide with tolerance average. Deficient product on the upper tolerance limit.

Center process, reduce variation!

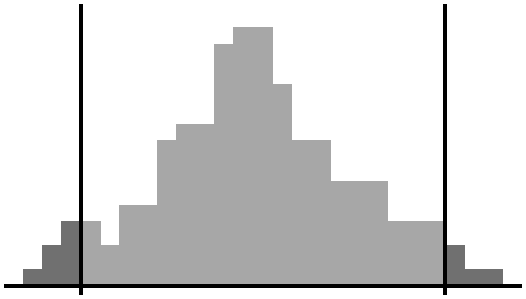


**Fig. 8.6:**

Mean of the distribution corresponds well with the tolerance average. Variation range is too large. Both tolerance limits exceeded.

Reduce variation!

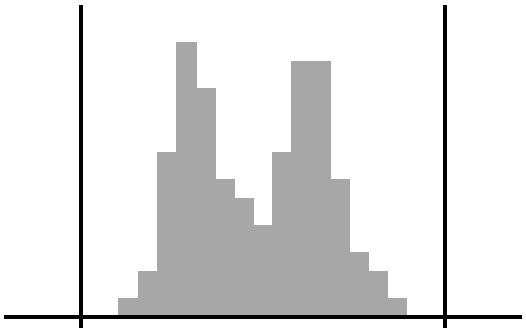


**Fig. 8.7:**

Superposition of two distributions.

Possibly caused by systematic process changes (e.g. tool, material).

After eliminating the cause, the tolerance is easy to maintain because the variation range of both distributions is comparably small.

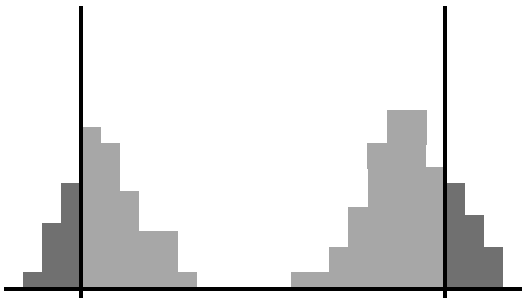


**Fig. 8.8:**

Similar situations like in Figure 8.7, however, the means of both distributions are placed so far from one another resulting in deficient product at both tolerance limits.

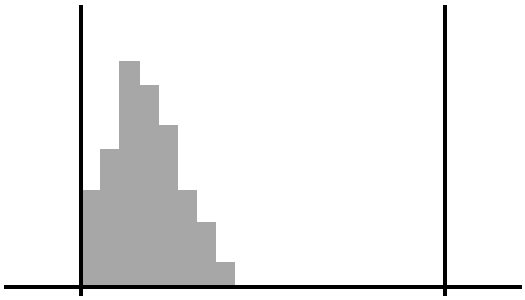


**Fig. 8.9:**

Mean of the distribution is shifted toward the lower tolerance limit. Apparently the lot was sorted 100%. If the process can be centered, sorting will no longer be necessary.

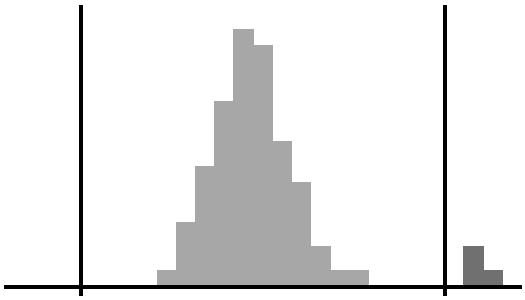


**Fig. 8.10:**

The main distribution has a low variation range. Mean corresponds well with the tolerance average. A small portion lies beyond the upper tolerance limit. We might be dealing with deficient product that occurred while setting up the machine and that was not sorted out.

# 9 Accuracy of estimating mean and standard deviation

Statements about a population that were derived from a sample are always associated with statistical uncertainty. This uncertainty generally increases, the smaller the underlying data basis, meaning the size $n$ of the sample.

In the following, we require that the population is normally distributed and its mean $\mu$ and standard deviation $\sigma$ are not known.

The (empirical) quantities $\bar{x}$ and $s$ calculated from a sample are estimations of the unknown quantities $\mu$ or $\sigma$. This is expressed with

$\hat{\mu} = \bar{x}$ ($\bar{x}$ is an estimation for $\mu$)

$\hat{\sigma} = s$ ($s$ is an estimation for $\sigma$)

One can give an interval around $\bar{x}$ or $s$, where the unknown quantities $\mu$ and $\sigma$, respectively, lie with a great probability. The width of this so-called confidence level depends, on the one hand, on the sample size $n$, and on the other hand, on a specified confidence level $P_A$. The quantity $1 - P_A$ is the associated significance level.

In Figures 9.1 and 9.2, the confidence levels of $\mu$ and $\sigma$ are represented as a function of the sample size. The curves are valid for a confidence level of 95%. This means that, on average in 95 of 100 cases, in which $\mu$ is estimated by $\bar{x}$ or $\sigma$ through $s$, $\mu$ or $\sigma$ lie within the confidence limits derived from the curves.

a) Confidence level for $\sigma$:

$$\frac{s_R}{D_{up}} \le \sigma \le \frac{s_R}{D_{lo}}$$

$s_R$ is the standard deviation of the sample determined using the range method (see Chapter 3.9). One finds the factors $\frac{1}{D_{lo}}$ and $\frac{1}{D_{up}}$ by looking for the sample size $n$ on the x-axis (which is divided logarithmically), going up perpendicularly until reaching both of the curves and then going from here horizontally to the left until the y-axis. One then reads the corresponding value for $\frac{1}{D_{lo}}$ or $\frac{1}{D_{up}}$ (Figure 9.1).

b) Confidence level for $\mu$:

$$\bar{x} - \frac{t'}{\sqrt{n}} \cdot s_R \le \mu \le \bar{x} + \frac{t'}{\sqrt{n}} \cdot s_R$$

One can find the factor $\frac{t'}{\sqrt{n}}$ similar to the procedure described in a) (Figure 9.2). Since the confidence level for $\mu$ is symmetrical, it is enough to read $\frac{t'}{\sqrt{n}}$ from the upper curve.

The larger the sample, the more the factors $\frac{1}{D_{lo}}$, $\frac{1}{D_{up}}$ approach the value 1, or $\frac{t'}{\sqrt{n}}$ converges to 0.

The curve progressions also indicate that the confidence levels become insignificantly smaller for sample sizes beyond $n = 50$ (despite increasing testing effort). In contrast, one should carefully interpret statements about mean and standard deviation for sample sizes $n < 25$, since the associated confidence levels become very large.

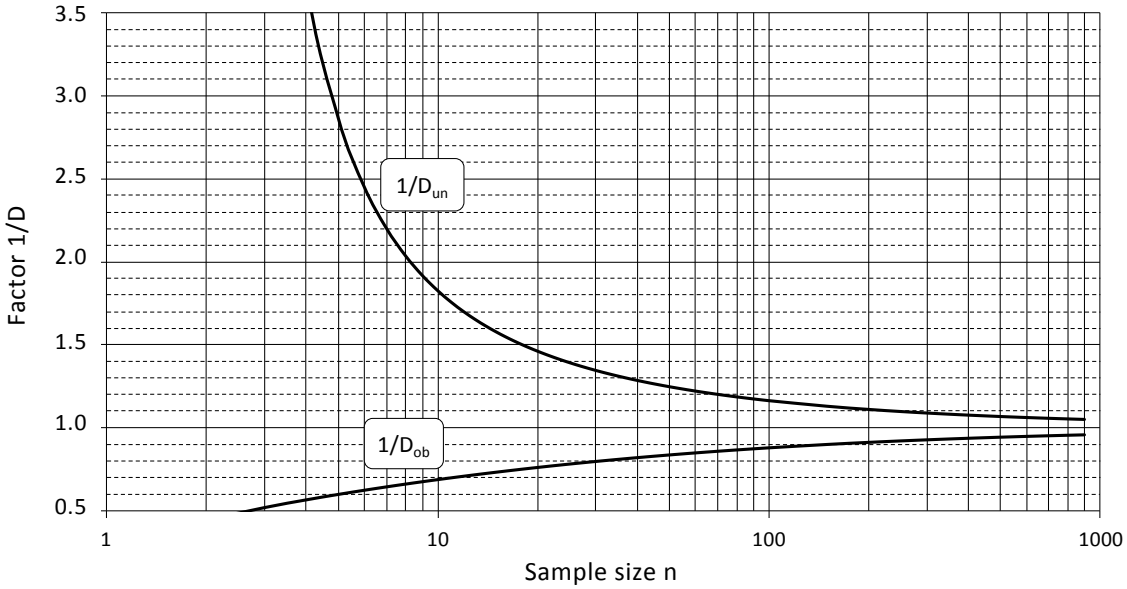



**Fig. 9.1:** Diagram to determine the confidence level for $\sigma$ ($P_A = 95\%$)

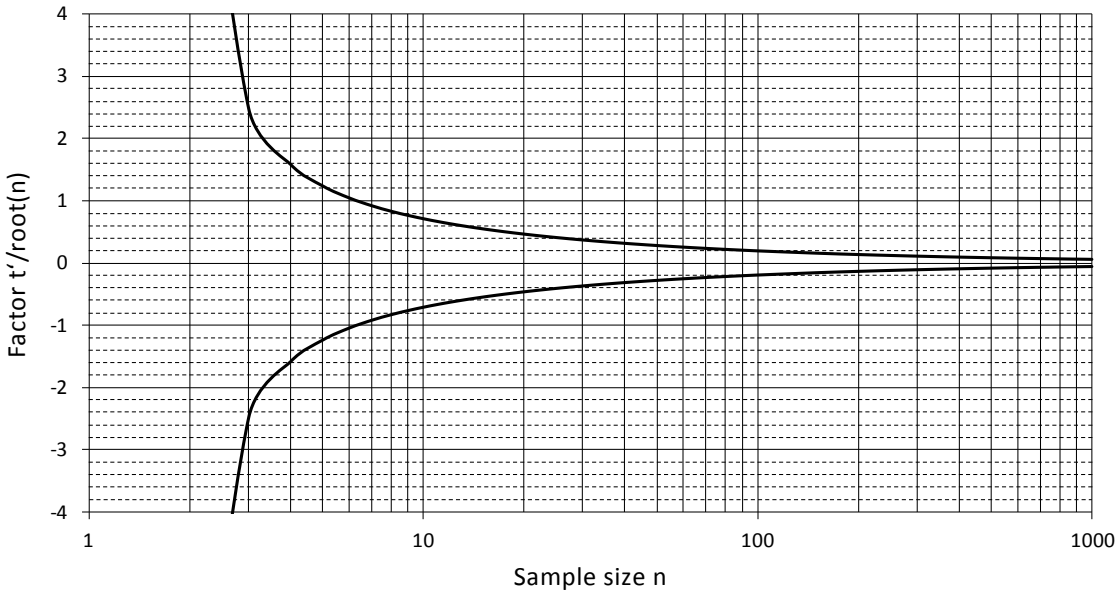



**Fig. 9.2:** Diagram to determine the confidence level for $\mu$ ($P_A = 95\%$)

# 10 Standard normal distribution $\Phi(-u) = 1 - \Phi(u)$ $D(u) = \Phi(u) - \Phi(-u)$

| u | Φ (-u) | Φ (u) | D(u) |
|---|---|---|---|
| 0.01 | 0.496011 | 0.503989 | 0.007979 |
| 0.02 | 0.492022 | 0.507978 | 0.015957 |
| 0.03 | 0.488033 | 0.511967 | 0.023933 |
| 0.04 | 0.484047 | 0.515953 | 0.031907 |
| 0.05 | 0.480061 | 0.519939 | 0.039878 |
| 0.06 | 0.476078 | 0.523922 | 0.047845 |
| 0.07 | 0.472097 | 0.527903 | 0.055806 |
| 0.08 | 0.468119 | 0.531881 | 0.063763 |
| 0.09 | 0.464144 | 0.535856 | 0.071713 |
| 0.10 | 0.460172 | 0.539828 | 0.079656 |
| 0.11 | 0.456205 | 0.543795 | 0.087591 |
| 0.12 | 0.452242 | 0.547758 | 0.095517 |
| 0.13 | 0.448283 | 0.551717 | 0.103434 |
| 0.14 | 0.444330 | 0.555670 | 0.111340 |
| 0.15 | 0.440382 | 0.559618 | 0.119235 |
| 0.16 | 0.436441 | 0.563559 | 0.127119 |
| 0.17 | 0.432505 | 0.567495 | 0.134990 |
| 0.18 | 0.428576 | 0.571424 | 0.142847 |
| 0.19 | 0.424655 | 0.575345 | 0.150691 |
| 0.20 | 0.420740 | 0.579260 | 0.158519 |
| 0.21 | 0.416834 | 0.583166 | 0.166332 |
| 0.22 | 0.412936 | 0.587064 | 0.174129 |
| 0.23 | 0.409046 | 0.590954 | 0.181908 |
| 0.24 | 0.405165 | 0.594835 | 0.189670 |
| 0.25 | 0.401294 | 0.598706 | 0.197413 |
| 0.26 | 0.397432 | 0.602568 | 0.205136 |
| 0.27 | 0.393580 | 0.606420 | 0.212840 |
| 0.28 | 0.389739 | 0.610261 | 0.220522 |
| 0.29 | 0.385908 | 0.614092 | 0.228184 |
| 0.30 | 0.382089 | 0.617911 | 0.235823 |
| 0.31 | 0.378281 | 0.621719 | 0.243439 |
| 0.32 | 0.374484 | 0.625516 | 0.251032 |
| 0.33 | 0.370700 | 0.629300 | 0.258600 |
| 0.34 | 0.366928 | 0.633072 | 0.266143 |
| 0.35 | 0.363169 | 0.636831 | 0.273661 |
| 0.36 | 0.359424 | 0.640576 | 0.281153 |
| 0.37 | 0.355691 | 0.644309 | 0.288617 |
| 0.38 | 0.351973 | 0.648027 | 0.296054 |
| 0.39 | 0.348268 | 0.651732 | 0.303463 |
| 0.40 | 0.344578 | 0.655422 | 0.310843 |
| 0.41 | 0.340903 | 0.659097 | 0.318194 |
| 0.42 | 0.337243 | 0.662757 | 0.325514 |
| 0.43 | 0.333598 | 0.666402 | 0.332804 |
| 0.44 | 0.329969 | 0.670031 | 0.340063 |
| 0.45 | 0.326355 | 0.673645 | 0.347290 |
| 0.46 | 0.322758 | 0.677242 | 0.354484 |
| 0.47 | 0.319178 | 0.680822 | 0.361645 |
| 0.48 | 0.315614 | 0.684386 | 0.368773 |
| 0.49 | 0.312067 | 0.687933 | 0.375866 |
| 0.50 | 0.308538 | 0.691462 | 0.382925 |

| u | Φ (-u) | Φ (u) | D(u) |
|---|---|---|---|
| 0.51 | 0.305026 | 0.694974 | 0.389949 |
| 0.52 | 0.301532 | 0.698468 | 0.396936 |
| 0.53 | 0.298056 | 0.701944 | 0.403888 |
| 0.54 | 0.294598 | 0.705402 | 0.410803 |
| 0.55 | 0.291160 | 0.708840 | 0.417681 |
| 0.56 | 0.287740 | 0.712260 | 0.424521 |
| 0.57 | 0.284339 | 0.715661 | 0.431322 |
| 0.58 | 0.280957 | 0.719043 | 0.438085 |
| 0.59 | 0.277595 | 0.722405 | 0.444809 |
| 0.60 | 0.274253 | 0.725747 | 0.451494 |
| 0.61 | 0.270931 | 0.729069 | 0.458138 |
| 0.62 | 0.267629 | 0.732371 | 0.464742 |
| 0.63 | 0.264347 | 0.735653 | 0.471306 |
| 0.64 | 0.261086 | 0.738914 | 0.477828 |
| 0.65 | 0.257846 | 0.742154 | 0.484308 |
| 0.66 | 0.254627 | 0.745373 | 0.490746 |
| 0.67 | 0.251429 | 0.748571 | 0.497142 |
| 0.68 | 0.248252 | 0.751748 | 0.503496 |
| 0.69 | 0.245097 | 0.754903 | 0.509806 |
| 0.70 | 0.241964 | 0.758036 | 0.516073 |
| 0.71 | 0.238852 | 0.761148 | 0.522296 |
| 0.72 | 0.235762 | 0.764238 | 0.528475 |
| 0.73 | 0.232695 | 0.767305 | 0.534610 |
| 0.74 | 0.229650 | 0.770350 | 0.540700 |
| 0.75 | 0.226627 | 0.773373 | 0.546745 |
| 0.76 | 0.223627 | 0.776373 | 0.552746 |
| 0.77 | 0.220650 | 0.779350 | 0.558700 |
| 0.78 | 0.217695 | 0.782305 | 0.564609 |
| 0.79 | 0.214764 | 0.785236 | 0.570472 |
| 0.80 | 0.211855 | 0.788145 | 0.576289 |
| 0.81 | 0.208970 | 0.791030 | 0.582060 |
| 0.82 | 0.206108 | 0.793892 | 0.587784 |
| 0.83 | 0.203269 | 0.796731 | 0.593461 |
| 0.84 | 0.200454 | 0.799546 | 0.599092 |
| 0.85 | 0.197662 | 0.802338 | 0.604675 |
| 0.86 | 0.194894 | 0.805106 | 0.610211 |
| 0.87 | 0.192150 | 0.807850 | 0.615700 |
| 0.88 | 0.189430 | 0.810570 | 0.621141 |
| 0.89 | 0.186733 | 0.813267 | 0.626534 |
| 0.90 | 0.184060 | 0.815940 | 0.631880 |
| 0.91 | 0.181411 | 0.818589 | 0.637178 |
| 0.92 | 0.178786 | 0.821214 | 0.642427 |
| 0.93 | 0.176186 | 0.823814 | 0.647629 |
| 0.94 | 0.173609 | 0.826391 | 0.652782 |
| 0.95 | 0.171056 | 0.828944 | 0.657888 |
| 0.96 | 0.168528 | 0.831472 | 0.662945 |
| 0.97 | 0.166023 | 0.833977 | 0.667954 |
| 0.98 | 0.163543 | 0.836457 | 0.672914 |
| 0.99 | 0.161087 | 0.838913 | 0.677826 |
| 1.00 | 0.158655 | 0.841345 | 0.682689 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 1.01 | 0.156248 | 0.843752 | 0.687505 |
| 1.02 | 0.153864 | 0.846136 | 0.692272 |
| 1.03 | 0.151505 | 0.848495 | 0.696990 |
| 1.04 | 0.149170 | 0.850830 | 0.701660 |
| 1.05 | 0.146859 | 0.853141 | 0.706282 |
| 1.06 | 0.144572 | 0.855428 | 0.710855 |
| 1.07 | 0.142310 | 0.857690 | 0.715381 |
| 1.08 | 0.140071 | 0.859929 | 0.719858 |
| 1.09 | 0.137857 | 0.862143 | 0.724287 |
| 1.10 | 0.135666 | 0.864334 | 0.728668 |
| 1.11 | 0.133500 | 0.866500 | 0.733001 |
| 1.12 | 0.131357 | 0.868643 | 0.737286 |
| 1.13 | 0.129238 | 0.870762 | 0.741524 |
| 1.14 | 0.127143 | 0.872857 | 0.745714 |
| 1.15 | 0.125072 | 0.874928 | 0.749856 |
| 1.16 | 0.123024 | 0.876976 | 0.753951 |
| 1.17 | 0.121001 | 0.878999 | 0.757999 |
| 1.18 | 0.119000 | 0.881000 | 0.762000 |
| 1.19 | 0.117023 | 0.882977 | 0.765953 |
| 1.20 | 0.115070 | 0.884930 | 0.769861 |
| 1.21 | 0.113140 | 0.886860 | 0.773721 |
| 1.22 | 0.111233 | 0.888767 | 0.777535 |
| 1.23 | 0.109349 | 0.890651 | 0.781303 |
| 1.24 | 0.107488 | 0.892512 | 0.785024 |
| 1.25 | 0.105650 | 0.894350 | 0.788700 |
| 1.26 | 0.103835 | 0.896165 | 0.792331 |
| 1.27 | 0.102042 | 0.897958 | 0.795915 |
| 1.28 | 0.100273 | 0.899727 | 0.799455 |
| 1.29 | 0.098525 | 0.901475 | 0.802949 |
| 1.30 | 0.096801 | 0.903199 | 0.806399 |
| 1.31 | 0.095098 | 0.904902 | 0.809804 |
| 1.32 | 0.093418 | 0.906582 | 0.813165 |
| 1.33 | 0.091759 | 0.908241 | 0.816482 |
| 1.34 | 0.090123 | 0.909877 | 0.819755 |
| 1.35 | 0.088508 | 0.911492 | 0.822984 |
| 1.36 | 0.086915 | 0.913085 | 0.826170 |
| 1.37 | 0.085344 | 0.914656 | 0.829313 |
| 1.38 | 0.083793 | 0.916207 | 0.832413 |
| 1.39 | 0.082264 | 0.917736 | 0.835471 |
| 1.40 | 0.080757 | 0.919243 | 0.838487 |
| 1.41 | 0.079270 | 0.920730 | 0.841460 |
| 1.42 | 0.077804 | 0.922196 | 0.844392 |
| 1.43 | 0.076359 | 0.923641 | 0.847283 |
| 1.44 | 0.074934 | 0.925066 | 0.850133 |
| 1.45 | 0.073529 | 0.926471 | 0.852941 |
| 1.46 | 0.072145 | 0.927855 | 0.855710 |
| 1.47 | 0.070781 | 0.929219 | 0.858438 |
| 1.48 | 0.069437 | 0.930563 | 0.861127 |
| 1.49 | 0.068112 | 0.931888 | 0.863776 |
| 1.50 | 0.066807 | 0.933193 | 0.866386 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 1.51 | 0.065522 | 0.934478 | 0.868957 |
| 1.52 | 0.064256 | 0.935744 | 0.871489 |
| 1.53 | 0.063008 | 0.936992 | 0.873983 |
| 1.54 | 0.061780 | 0.938220 | 0.876440 |
| 1.55 | 0.060571 | 0.939429 | 0.878858 |
| 1.56 | 0.059380 | 0.940620 | 0.881240 |
| 1.57 | 0.058208 | 0.941792 | 0.883585 |
| 1.58 | 0.057053 | 0.942947 | 0.885893 |
| 1.59 | 0.055917 | 0.944083 | 0.888165 |
| 1.60 | 0.054799 | 0.945201 | 0.890401 |
| 1.61 | 0.053699 | 0.946301 | 0.892602 |
| 1.62 | 0.052616 | 0.947384 | 0.894768 |
| 1.63 | 0.051551 | 0.948449 | 0.896899 |
| 1.64 | 0.050503 | 0.949497 | 0.898995 |
| 1.65 | 0.049471 | 0.950529 | 0.901057 |
| 1.66 | 0.048457 | 0.951543 | 0.903086 |
| 1.67 | 0.047460 | 0.952540 | 0.905081 |
| 1.68 | 0.046479 | 0.953521 | 0.907043 |
| 1.69 | 0.045514 | 0.954486 | 0.908972 |
| 1.70 | 0.044565 | 0.955435 | 0.910869 |
| 1.71 | 0.043633 | 0.956367 | 0.912734 |
| 1.72 | 0.042716 | 0.957284 | 0.914568 |
| 1.73 | 0.041815 | 0.958185 | 0.916370 |
| 1.74 | 0.040929 | 0.959071 | 0.918141 |
| 1.75 | 0.040059 | 0.959941 | 0.919882 |
| 1.76 | 0.039204 | 0.960796 | 0.921592 |
| 1.77 | 0.038364 | 0.961636 | 0.923273 |
| 1.78 | 0.037538 | 0.962462 | 0.924924 |
| 1.79 | 0.036727 | 0.963273 | 0.926546 |
| 1.80 | 0.035930 | 0.964070 | 0.928139 |
| 1.81 | 0.035148 | 0.964852 | 0.929704 |
| 1.82 | 0.034379 | 0.965621 | 0.931241 |
| 1.83 | 0.033625 | 0.966375 | 0.932750 |
| 1.84 | 0.032884 | 0.967116 | 0.934232 |
| 1.85 | 0.032157 | 0.967843 | 0.935687 |
| 1.86 | 0.031443 | 0.968557 | 0.937115 |
| 1.87 | 0.030742 | 0.969258 | 0.938516 |
| 1.88 | 0.030054 | 0.969946 | 0.939892 |
| 1.89 | 0.029379 | 0.970621 | 0.941242 |
| 1.90 | 0.028716 | 0.971284 | 0.942567 |
| 1.91 | 0.028067 | 0.971933 | 0.943867 |
| 1.92 | 0.027429 | 0.972571 | 0.945142 |
| 1.93 | 0.026803 | 0.973197 | 0.946393 |
| 1.94 | 0.026190 | 0.973810 | 0.947620 |
| 1.95 | 0.025588 | 0.974412 | 0.948824 |
| 1.96 | 0.024998 | 0.975002 | 0.950004 |
| 1.97 | 0.024419 | 0.975581 | 0.951162 |
| 1.98 | 0.023852 | 0.976148 | 0.952297 |
| 1.99 | 0.023295 | 0.976705 | 0.953409 |
| 2.00 | 0.022750 | 0.977250 | 0.954500 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 2.01 | 0.022216 | 0.977784 | 0.955569 |
| 2.02 | 0.021692 | 0.978308 | 0.956617 |
| 2.03 | 0.021178 | 0.978822 | 0.957644 |
| 2.04 | 0.020675 | 0.979325 | 0.958650 |
| 2.05 | 0.020182 | 0.979818 | 0.959636 |
| 2.06 | 0.019699 | 0.980301 | 0.960602 |
| 2.07 | 0.019226 | 0.980774 | 0.961548 |
| 2.08 | 0.018763 | 0.981237 | 0.962475 |
| 2.09 | 0.018309 | 0.981691 | 0.963382 |
| 2.10 | 0.017864 | 0.982136 | 0.964271 |
| 2.11 | 0.017429 | 0.982571 | 0.965142 |
| 2.12 | 0.017003 | 0.982997 | 0.965994 |
| 2.13 | 0.016586 | 0.983414 | 0.966829 |
| 2.14 | 0.016177 | 0.983823 | 0.967645 |
| 2.15 | 0.015778 | 0.984222 | 0.968445 |
| 2.16 | 0.015386 | 0.984614 | 0.969227 |
| 2.17 | 0.015003 | 0.984997 | 0.969993 |
| 2.18 | 0.014629 | 0.985371 | 0.970743 |
| 2.19 | 0.014262 | 0.985738 | 0.971476 |
| 2.20 | 0.013903 | 0.986097 | 0.972193 |
| 2.21 | 0.013553 | 0.986447 | 0.972895 |
| 2.22 | 0.013209 | 0.986791 | 0.973581 |
| 2.23 | 0.012874 | 0.987126 | 0.974253 |
| 2.24 | 0.012545 | 0.987455 | 0.974909 |
| 2.25 | 0.012224 | 0.987776 | 0.975551 |
| 2.26 | 0.011911 | 0.988089 | 0.976179 |
| 2.27 | 0.011604 | 0.988396 | 0.976792 |
| 2.28 | 0.011304 | 0.988696 | 0.977392 |
| 2.29 | 0.011011 | 0.988989 | 0.977979 |
| 2.30 | 0.010724 | 0.989276 | 0.978552 |
| 2.31 | 0.010444 | 0.989556 | 0.979112 |
| 2.32 | 0.010170 | 0.989830 | 0.979659 |
| 2.33 | 0.009903 | 0.990097 | 0.980194 |
| 2.34 | 0.009642 | 0.990358 | 0.980716 |
| 2.35 | 0.009387 | 0.990613 | 0.981227 |
| 2.36 | 0.009137 | 0.990863 | 0.981725 |
| 2.37 | 0.008894 | 0.991106 | 0.982212 |
| 2.38 | 0.008656 | 0.991344 | 0.982687 |
| 2.39 | 0.008424 | 0.991576 | 0.983152 |
| 2.40 | 0.008198 | 0.991802 | 0.983605 |
| 2.41 | 0.007976 | 0.992024 | 0.984047 |
| 2.42 | 0.007760 | 0.992240 | 0.984479 |
| 2.43 | 0.007549 | 0.992451 | 0.984901 |
| 2.44 | 0.007344 | 0.992656 | 0.985313 |
| 2.45 | 0.007143 | 0.992857 | 0.985714 |
| 2.46 | 0.006947 | 0.993053 | 0.986106 |
| 2.47 | 0.006756 | 0.993244 | 0.986489 |
| 2.48 | 0.006569 | 0.993431 | 0.986862 |
| 2.49 | 0.006387 | 0.993613 | 0.987226 |
| 2.50 | 0.006210 | 0.993790 | 0.987581 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 2.51 | 0.006037 | 0.993963 | 0.987927 |
| 2.52 | 0.005868 | 0.994132 | 0.988264 |
| 2.53 | 0.005703 | 0.994297 | 0.988594 |
| 2.54 | 0.005543 | 0.994457 | 0.988915 |
| 2.55 | 0.005386 | 0.994614 | 0.989228 |
| 2.56 | 0.005234 | 0.994766 | 0.989533 |
| 2.57 | 0.005085 | 0.994915 | 0.989830 |
| 2.58 | 0.004940 | 0.995060 | 0.990120 |
| 2.59 | 0.004799 | 0.995201 | 0.990402 |
| 2.60 | 0.004661 | 0.995339 | 0.990678 |
| 2.61 | 0.004527 | 0.995473 | 0.990946 |
| 2.62 | 0.004397 | 0.995603 | 0.991207 |
| 2.63 | 0.004269 | 0.995731 | 0.991461 |
| 2.64 | 0.004145 | 0.995855 | 0.991709 |
| 2.65 | 0.004025 | 0.995975 | 0.991951 |
| 2.66 | 0.003907 | 0.996093 | 0.992186 |
| 2.67 | 0.003793 | 0.996207 | 0.992415 |
| 2.68 | 0.003681 | 0.996319 | 0.992638 |
| 2.69 | 0.003573 | 0.996427 | 0.992855 |
| 2.70 | 0.003467 | 0.996533 | 0.993066 |
| 2.71 | 0.003364 | 0.996636 | 0.993272 |
| 2.72 | 0.003264 | 0.996736 | 0.993472 |
| 2.73 | 0.003167 | 0.996833 | 0.993666 |
| 2.74 | 0.003072 | 0.996928 | 0.993856 |
| 2.75 | 0.002980 | 0.997020 | 0.994040 |
| 2.76 | 0.002890 | 0.997110 | 0.994220 |
| 2.77 | 0.002803 | 0.997197 | 0.994394 |
| 2.78 | 0.002718 | 0.997282 | 0.994564 |
| 2.79 | 0.002635 | 0.997365 | 0.994729 |
| 2.80 | 0.002555 | 0.997445 | 0.994890 |
| 2.81 | 0.002477 | 0.997523 | 0.995046 |
| 2.82 | 0.002401 | 0.997599 | 0.995198 |
| 2.83 | 0.002327 | 0.997673 | 0.995345 |
| 2.84 | 0.002256 | 0.997744 | 0.995489 |
| 2.85 | 0.002186 | 0.997814 | 0.995628 |
| 2.86 | 0.002118 | 0.997882 | 0.995763 |
| 2.87 | 0.002052 | 0.997948 | 0.995895 |
| 2.88 | 0.001988 | 0.998012 | 0.996023 |
| 2.89 | 0.001926 | 0.998074 | 0.996147 |
| 2.90 | 0.001866 | 0.998134 | 0.996268 |
| 2.91 | 0.001807 | 0.998193 | 0.996386 |
| 2.92 | 0.001750 | 0.998250 | 0.996500 |
| 2.93 | 0.001695 | 0.998305 | 0.996610 |
| 2.94 | 0.001641 | 0.998359 | 0.996718 |
| 2.95 | 0.001589 | 0.998411 | 0.996822 |
| 2.96 | 0.001538 | 0.998462 | 0.996923 |
| 2.97 | 0.001489 | 0.998511 | 0.997022 |
| 2.98 | 0.001441 | 0.998559 | 0.997117 |
| 2.99 | 0.001395 | 0.998605 | 0.997210 |
| 3.00 | 0.001350 | 0.998650 | 0.997300 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 3.01 | 0.001306 | 0.998694 | 0.997387 |
| 3.02 | 0.001264 | 0.998736 | 0.997472 |
| 3.03 | 0.001223 | 0.998777 | 0.997554 |
| 3.04 | 0.001183 | 0.998817 | 0.997634 |
| 3.05 | 0.001144 | 0.998856 | 0.997711 |
| 3.06 | 0.001107 | 0.998893 | 0.997786 |
| 3.07 | 0.001070 | 0.998930 | 0.997859 |
| 3.08 | 0.001035 | 0.998965 | 0.997930 |
| 3.09 | 0.001001 | 0.998999 | 0.997998 |
| 3.10 | 0.000968 | 0.999032 | 0.998065 |
| 3.11 | 0.000936 | 0.999064 | 0.998129 |
| 3.12 | 0.000904 | 0.999096 | 0.998191 |
| 3.13 | 0.000874 | 0.999126 | 0.998252 |
| 3.14 | 0.000845 | 0.999155 | 0.998310 |
| 3.15 | 0.000816 | 0.999184 | 0.998367 |
| 3.16 | 0.000789 | 0.999211 | 0.998422 |
| 3.17 | 0.000762 | 0.999238 | 0.998475 |
| 3.18 | 0.000736 | 0.999264 | 0.998527 |
| 3.19 | 0.000711 | 0.999289 | 0.998577 |
| 3.20 | 0.000687 | 0.999313 | 0.998626 |
| 3.21 | 0.000664 | 0.999336 | 0.998673 |
| 3.22 | 0.000641 | 0.999359 | 0.998718 |
| 3.23 | 0.000619 | 0.999381 | 0.998762 |
| 3.24 | 0.000598 | 0.999402 | 0.998805 |
| 3.25 | 0.000577 | 0.999423 | 0.998846 |
| 3.26 | 0.000557 | 0.999443 | 0.998886 |
| 3.27 | 0.000538 | 0.999462 | 0.998924 |
| 3.28 | 0.000519 | 0.999481 | 0.998962 |
| 3.29 | 0.000501 | 0.999499 | 0.998998 |
| 3.30 | 0.000483 | 0.999517 | 0.999033 |
| 3.31 | 0.000467 | 0.999533 | 0.999067 |
| 3.32 | 0.000450 | 0.999550 | 0.999100 |
| 3.33 | 0.000434 | 0.999566 | 0.999131 |
| 3.34 | 0.000419 | 0.999581 | 0.999162 |
| 3.35 | 0.000404 | 0.999596 | 0.999192 |
| 3.36 | 0.000390 | 0.999610 | 0.999220 |
| 3.37 | 0.000376 | 0.999624 | 0.999248 |
| 3.38 | 0.000362 | 0.999638 | 0.999275 |
| 3.39 | 0.000350 | 0.999650 | 0.999301 |
| 3.40 | 0.000337 | 0.999663 | 0.999326 |
| 3.41 | 0.000325 | 0.999675 | 0.999350 |
| 3.42 | 0.000313 | 0.999687 | 0.999374 |
| 3.43 | 0.000302 | 0.999698 | 0.999396 |
| 3.44 | 0.000291 | 0.999709 | 0.999418 |
| 3.45 | 0.000280 | 0.999720 | 0.999439 |
| 3.46 | 0.000270 | 0.999730 | 0.999460 |
| 3.47 | 0.000260 | 0.999740 | 0.999479 |
| 3.48 | 0.000251 | 0.999749 | 0.999498 |
| 3.49 | 0.000242 | 0.999758 | 0.999517 |
| 3.50 | 0.000233 | 0.999767 | 0.999535 |

| u | Φ(-u) | Φ(u) | D(u) |
|---|---|---|---|
| 3.51 | 0.000224 | 0.999776 | 0.999552 |
| 3.52 | 0.000216 | 0.999784 | 0.999568 |
| 3.53 | 0.000208 | 0.999792 | 0.999584 |
| 3.54 | 0.000200 | 0.999800 | 0.999600 |
| 3.55 | 0.000193 | 0.999807 | 0.999615 |
| 3.56 | 0.000185 | 0.999815 | 0.999629 |
| 3.57 | 0.000179 | 0.999821 | 0.999643 |
| 3.58 | 0.000172 | 0.999828 | 0.999656 |
| 3.59 | 0.000165 | 0.999835 | 0.999669 |
| 3.60 | 0.000159 | 0.999841 | 0.999682 |
| 3.61 | 0.000153 | 0.999847 | 0.999694 |
| 3.62 | 0.000147 | 0.999853 | 0.999705 |
| 3.63 | 0.000142 | 0.999858 | 0.999717 |
| 3.64 | 0.000136 | 0.999864 | 0.999727 |
| 3.65 | 0.000131 | 0.999869 | 0.999738 |
| 3.66 | 0.000126 | 0.999874 | 0.999748 |
| 3.67 | 0.000121 | 0.999879 | 0.999757 |
| 3.68 | 0.000117 | 0.999883 | 0.999767 |
| 3.69 | 0.000112 | 0.999888 | 0.999776 |
| 3.70 | 0.000108 | 0.999892 | 0.999784 |
| 3.71 | 0.000104 | 0.999896 | 0.999793 |
| 3.72 | 0.000100 | 0.999900 | 0.999801 |
| 3.73 | 0.000096 | 0.999904 | 0.999808 |
| 3.74 | 0.000092 | 0.999908 | 0.999816 |
| 3.75 | 0.000088 | 0.999912 | 0.999823 |
| 3.76 | 0.000085 | 0.999915 | 0.999830 |
| 3.77 | 0.000082 | 0.999918 | 0.999837 |
| 3.78 | 0.000078 | 0.999922 | 0.999843 |
| 3.79 | 0.000075 | 0.999925 | 0.999849 |
| 3.80 | 0.000072 | 0.999928 | 0.999855 |
| 3.81 | 0.000070 | 0.999930 | 0.999861 |
| 3.82 | 0.000067 | 0.999933 | 0.999867 |
| 3.83 | 0.000064 | 0.999936 | 0.999872 |
| 3.84 | 0.000062 | 0.999938 | 0.999877 |
| 3.85 | 0.000059 | 0.999941 | 0.999882 |
| 3.86 | 0.000057 | 0.999943 | 0.999887 |
| 3.87 | 0.000054 | 0.999946 | 0.999891 |
| 3.88 | 0.000052 | 0.999948 | 0.999896 |
| 3.89 | 0.000050 | 0.999950 | 0.999900 |
| 3.90 | 0.000048 | 0.999952 | 0.999904 |
| 3.91 | 0.000046 | 0.999954 | 0.999908 |
| 3.92 | 0.000044 | 0.999956 | 0.999911 |
| 3.93 | 0.000042 | 0.999958 | 0.999915 |
| 3.94 | 0.000041 | 0.999959 | 0.999918 |
| 3.95 | 0.000039 | 0.999961 | 0.999922 |
| 3.96 | 0.000037 | 0.999963 | 0.999925 |
| 3.97 | 0.000036 | 0.999964 | 0.999928 |
| 3.98 | 0.000034 | 0.999966 | 0.999931 |
| 3.99 | 0.000033 | 0.999967 | 0.999934 |
| 4.00 | 0.000032 | 0.999968 | 0.999937 |

# 11 References

[1] Brochure: Elementary Quality Assurance Tools (C/QMM)

[2] M. Sadowy: Industrielle Statistik, Vogel-Verlag, Würzburg, 1970

[3] Erwin Kreyszig: Statistische Methoden und ihre Anwendungen, Vandenhoeck u. Ruprecht, Göttingen, 1988

[4] H. Weber: Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Ingenieure, Teubner, Stuttgart, 1988

[5] Karl Bosch: Elementare Einführung in die Wahrscheinlichkeitsrechnung, Vieweg, Braunschweig, 1989

[6] Karl Bosch: Elementare Einführung in die angewandte Statistik, Vieweg, Braunschweig, 1989

[7] F. Barth, H. Berghold, R. Haller: Stochastik (Grundkurs, Leistungskurs), Ehrenwirth, München, 1992

[8] G. Wagner und R. Lang: Statistische Auswertung von Meß- und Prüfergebnissen, Hrsg.: Deutsche Gesellschaft für Qualität e. V., DGQ 14, 1976

[9] Graf, Henning, Stange, Wilrich: Formeln und Tabellen der angewandten mathematischen Statistik, Springer-Verlag, Berlin, 1987

[10] Lothar Sachs: Angewandte Statistik, Springer-Verlag, Berlin, 1992

[11] Hartung: Statistik, Oldenbourg, München, 1989

# 12 Symbols and terms

$\int_{+\infty}^{-\infty}$ = Integral from negative infinity to positive infinity

$\sqrt{}$ = Root sign

$\sum$ = Summation symbol

$\prod$ = Product symbol

$\leq$ = less than or equal to

$\geq$ = Greater than or equal to

$\neq$ = Does not equal

$|x|$ = Absolute value of x (positive value of x)

$B_{lo}$, $B_{up}$ = Factors to calculate variation limits for s

C = Midpoint of the tolerance range or nominal value

$D_{lo}$, $D_{up}$ = Factors to calculate the confidence level for $\sigma$

e = Base of the natural log ($\approx$ 2.71828)

f = Degrees of freedom

f(x) = Probability density function

$G_j$ = Cumulative frequency

$h_j$ = Relative frequency

$H_j$ = Relative cumulative frequency

i, j = Count indices

k = Number of classes

ln = Natural logarithm

m = Number of value sets

n = Sample size

$n_j$ = Absolute frequency in the j-th class

P = Probability

$P_A$ = Confidence level

R = Range

s = Standard deviation of the sample

$s_R$ = Standard deviation determined using the range method

t' = Factor to calculate the confidence level for $\mu$ with unknown standard deviation of the population

| | |
|---|---|
| $u$ | = Standardized parameter of the normal distribution $N(\mu = 0; \sigma^2 = 1)$ |
| $v$ | = Variation coefficient |
| $w$ | = Class width |
| $x$ | = Continuous values |
| $x_i$ , $y_i$ | = Values of a measurement series |
| $x_{(1)}, \ldots, x_{(n)}$ | = Measurement series arranged according to the size of the values |
| $\bar{x}_g$ | = Geometric mean of a sample |
| $\bar{x}_H$ | = Harmonic mean |
| $x_{max}$ | = Largest value in a sample |
| $x_{min}$ | = Smallest value in a sample |
| $\tilde{x}$ | = Median of a sample |
| $\bar{x}$ | = Arithmetic mean of a sample |
| $\Delta$ | = Interval width |
| $\varepsilon$ | = Variation factor of lognormal distribution, geometric standard deviation |
| $\mu$ | = Mean of a population |
| $\mu_g$ | = Geometric mean of a population |
| $\sigma$ | = Standard deviation of the population |
| $\sigma^2$ | = Variance of a population |
| $\pi$ | = Number pi ($\approx 3.1416$) |

# Index

2020-04-06 - SOCOS

**Robert Bosch GmbH**
C/QMM
Postfach 30 02 20
D-70442 Stuttgart
Germany
Phone +49 711 811 - 0
**www.bosch.com**

2020-04-06 - SOCOS

**Robert Bosch GmbH**
C/QMM
Postfach 30 02 20
D-70442 Stuttgart
Germany
Phone +49 711 811-0
**www.bosch.com**