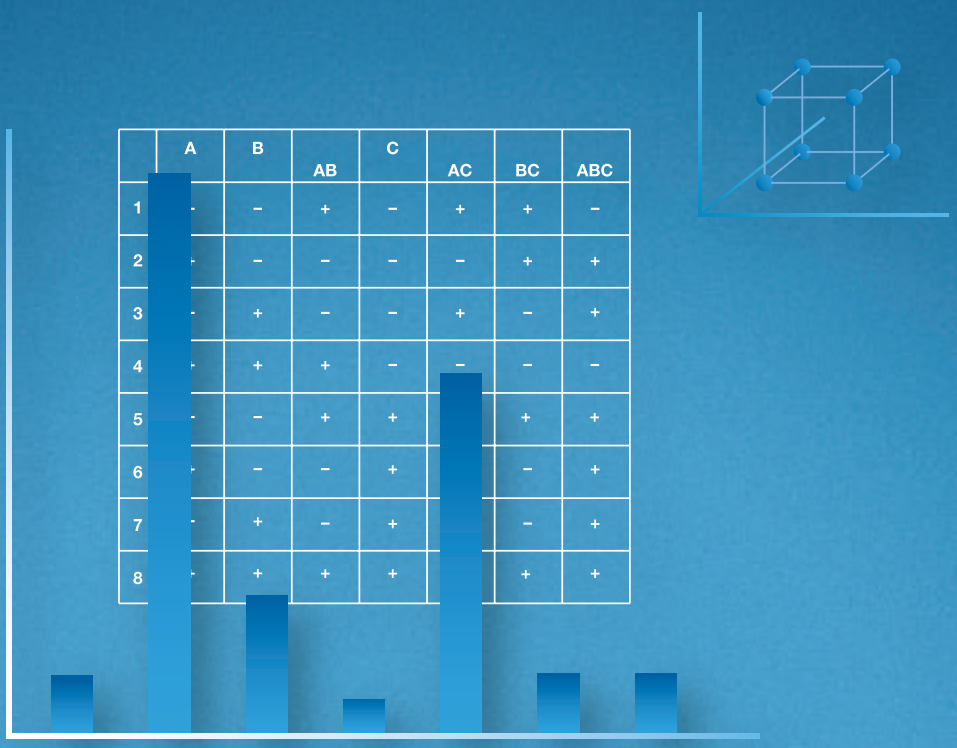
2020-04-06 - SOCOS




Quality Management in the Bosch Group | Technical Statistics

# 11. Design of **Experiments** (DoE)

BOSCH
Invented for life

**Table of contents:**

## 1. Introduction to the design of experiments

The aim of this volume is to provide associates in research, development and production with an overview of the subject of DoE at Bosch. Within this context, the design of experiments (DoE) is understood to be the systematic procedure for designing, conducting and evaluating experiments while minimizing the necessary resources. The subject of “statistical experimental design”, often equated with “DoE” in the literature, is treated in this volume as just one aspect of DoE, and as such is dealt with here only insofar as is necessary for common technical concerns. Extensive literature is available for those desiring a more detailed study. This document also largely dispenses with the mathematical explanation of relationships.

This volume covers all major aspects of the design of DoE as it pertains to engineering practice, e.g. the conducting of experiments and the role of modeling. Furthermore, it acts as a guide for the planning, preparation and evaluation of experiments. Those wishing to tackle these complex issues independently will also require adequate practical experience, however.

Layout of this volume:

Chapter 1: Overview and important points for understanding the methods, their importance and, in particular, their application at Bosch.

Chapters 2…7: Detailed description of DoE, with the individual steps in chronological order.

Chapter 8: Advanced approaches (see Appendix for details on the relevant statistical bases).



### 1.1. The importance of experiments

The objective of a test or experiment is the - mostly empirical - examination of an assumption (from the Latin experimentum: attempt to demonstrate, test, trial). A real object is examined or measured to determine whether or not the assumption is correct. Finally, every test has the objective of furthering knowledge. Conducting an experiment before posing the question that the experiment is intended to answer is not a purposeful approach.

An experiment is always an investigation of the causal relationship between cause and effect, based on the notion of a process with influences as the input and reactions as the output. Variables that describe these influences are referred to as predictor variables, those that denote the reactions, as response variables.

> An experiment is the empirical examination of the causal relationship between the predictor variables and response variables of a system, based on observation and measurement, with the aim of gaining knowledge.

We distinguish between the following experiments, depending on whether or not the predictor variables can be controlled:

a) **Experiments under controlled conditions**, e.g. laboratory tests, in which the predictor variables can be set in accordance with a defined plan.

b) **Experiments under uncontrolled conditions**, often also referred to as field tests, in which the setting of variables is either physically impossible or is deemed unviable for economic, ethical or other reasons.

Where possible, the first type of test is always preferable, as it permits targeted intervention to achieve a rapid, cost-effective acquisition of knowledge; with the second type this is significantly harder.

Furthermore, we need to differentiate between the terms **experiment and replications**. If the test encompasses variables subject to variance, several replications are generally required. These are then referred to in their entirety as an experiment.

### 1.1.1. Acquiring knowledge through experiments

In systems theory, the terms ‘system’ and ‘model’ play a key role. A ‘system’ is an initially very abstract concept denoting an extremely varied range of objects under observation, e.g. a fuel injection system, a production plant, a house or even an organism. ‘Models’, on the other hand, provide a simplified representation of real systems, in order to describe the major characteristics of those systems, such as structure and function. Experiments play a crucial role in the creation of models.

#### 1.1.1.1. The basic terms system and model

**Definition of system:**

| |
|---|
| BES-PE Glossary (issue 3-2009/11): A system is an entirety of elements that relate with each other and interact in such a way that they can be regarded as a unit with a specific task, sense or purpose, and in this way differentiate themselves from the environment that surrounds them. |

A system has a system structure and fulfils a particular function; see [13], p.35. EN ISO 9000: System = a set of elements that are interrelated or interact with one another.

Since the aspects of differentiation from the environment and function are important for DoE, however, the first of these definitions is used in this document.

**Definition of model:**

| |
|---|
| BES-PE Glossary (issue 3-2009/11): A model is a representation of reality that is reduced to relevant characteristics. |

A model is the simplified representation of a system - with limited applicability - which is intended to convey valid information about the system for a particular model purpose, see [13], p.51. Here, it is important to understand that a model is not correct or incorrect, but fit for purpose or not, depending on the task that the model is intended to accomplish. Various models of the same system may be required for answering different questions.

**System analysis** is one of the methods of systems theory that deals with the development of models. This process is often referred to as **modeling**.

**Reasons** for modeling:

- To reduce the complexity of the relationships in a system
- To increase understanding of and explain the behavior of the system through simulation, including in situations where studies of the real system are impossible, uneconomical or unviable
- To predict or optimize system behavior before the system physically exists.

**General procedure** for modeling:

1. **Creating a model:** Mostly an iterative, heuristic process whereby the model, beginning with the simplest possible notion of a model, is adapted on the basis of our knowledge of reality until it satisfies all important aspects in respect of the task to be accomplished.
2. **Validation of the model:** Here, predicted system properties and behavior are compared with observed properties and behavior.
3. **Simulation:** Acquiring knowledge of the system using the model.
4. if necessary **further development of the model**, then continuing with Step 2.

We refer to structural or behavioral models, depending on the structure and function of the system that the model is supposed to represent [13], p.61.

**Representation of the system structure** is achieved by:

- the system boundary (differentiation from the environment),
- the system elements with their characteristics, whereby the elements may in themselves be systems (subsystems),
- interactions incl. feedback between the elements and between the system and the environment through the flow of materials, energy and information.

The system structure depends to a great extent upon the depth of observation. On the one hand, it should be possible to record the major influences (e.g. interaction between the elements); on the other hand, the level of detail must not allow the model to become too complex. Thus, the **basic principle** is: as rough as possible, and as detailed as necessary.

The function of a (sub)system does not describe “what something is like”, but rather “what it does”, i.e. it is patterns of behavior, not components, that are represented [14], p.55.

**Representation of function** is achieved by:

- Describing how the system interacts with its environment through inputs and outputs.

  *REMARKS: Inputs describe the effect of the environment on the system, outputs the reaction of the system to the environment. These variables do not have to match the predictor variables and response variables of the experiment, but describe the system’s relationship with its environment. Predictor and response variables, on the other hand, describe the experiment within the framework of a process model.*

- Describing the functional relationship between inputs and outputs by means of a transfer function.

  *REMARKS: This description can apply both to the highest (super) system level and any (sub)system level.*



The structure and function of a system can be illustrated in the form of a graph.

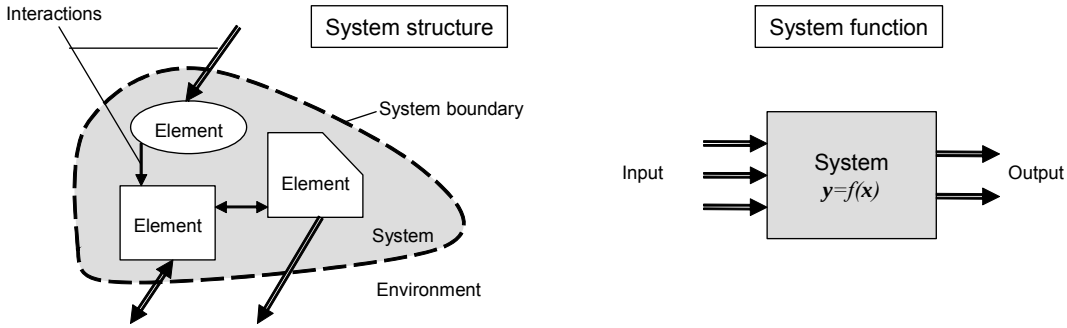



**Figure 1.1:** System structure and function

### 1.1.1.2. Methods of modeling

We distinguish between two methods of modeling, and mixed forms thereof [13], p. 52ff.:

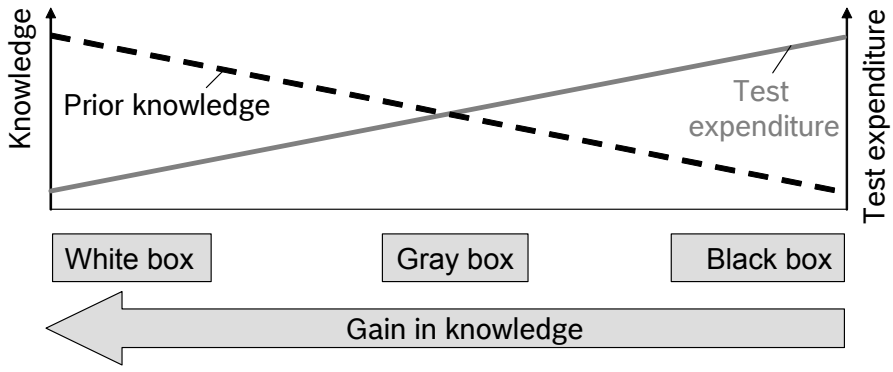



**Figure 1.2:** Types of modeling

**Physical modeling (white box)**

Physical modeling (also referred to as white-box or glass-box modeling) denotes a deductive process (from the general to the specific) on the basis of general natural laws and principles (e.g. laws of conservation, laws from chemistry, physics or thermodynamics), and structural knowledge of the system. Here, the function of the system elements at the chosen level of observation must be fully known, in order that the function of the system as a whole can be clearly deduced from the system structure. Model parameters must be defined on the basis of physical constants and the *known* characteristics of the system elements. The model is frequently represented as a set of complex, partial differential equations (PDEs).

Here, the model approach is limited only by the applicability of the laws on which it is based, which means that it may be used over a wide spectrum - a significant advantage of this method. We refer to a structural model that *explains* system behavior.

With physical modeling, too, the model must be validated by means of experiments.

> *EXAMPLE: One example of physical modeling is structural-mechanical models of components, the deformation behavior under load of which is examined by means of the finite element method. Here, the component to be modeled is regarded as a system of "finite elements". The system structure is defined by the division of the component's volume, e.g. into tetrahedral elements, and their arrangement. The behavior (function) of the individual elements is described by a system of partial differential equations; interaction between the elements is produced by global equilibrium conditions. The known laws of elasticity theory and knowledge of the system (e.g. concerning its geometry and appearance) enable global system behavior to be deduced. The loads and other boundary conditions produce the inputs. The system reacts to these with a certain deformation (outputs). The model parameters are defined by the E modulus and Poisson's ratio of the material, which must be known. The model is limited by the applicability of the fundamental laws of elasticity, but these no longer apply during plastic deformation under extreme load. For this reason, experimental validation is required, e.g. by measuring actual deformation with the aid of a strain gage and subsequently comparing this reading with the arithmetical value.*



In its purest form, white-box modeling is seldom encountered. Physical modeling may be employed for technical questions if the state of the art is adequate to the task and comprehensive knowledge of the system is available. This often does not apply to quantitative statements, because sufficient knowledge about the structure, laws or element characteristics is either not available with the required level of detail, or cannot be obtained for cost or time reasons. Variance adds to the difficulty of modeling. Gray-box modeling may provide a remedy to this problem. However, qualitative analyses of dependencies and interactions can be achieved through the study of parameters.

**Empirical modeling (black box)**

Empirical modeling, also known as system identification or black-box modeling, is an inductive approach (from the specific to the general) whereby a model is created only by recording the system inputs and outputs by way of an experiment in a finite number of discrete cases (e.g. within a finite period of time), e.g. through measurement and observation. This model then describes the relationship between inputs and outputs in the best possible way, using a transfer function. This may also be referred to as an approach that *describes* behavior. The basis for the transfer function is a general mathematical approach, which must be set a priori as default, mostly as a simple algebraic equation, e.g. as a polynomial of the first or second degree. Subsequent system identification is achieved by determining the free model parameters (coefficients in the equation), so that the transfer function best describes the system behavior, within the framework of model accuracy.

Advantage:

- Those conducting the experiment do not have to have knowledge of the structure, physical laws or element characteristics of the system (knowledge is not actually available or cannot be obtained).

Disadvantages:

- Opportunities to gain a more in-depth understanding of system behavior may not be exploited.
- The model can only be used within narrow confines, because complex, global system behavior cannot be approximated by the limited order of the model equation. Modeling is thus only suitable for interpolation.

*REMARKS: Within these narrow limits however, it is possible to show, through Taylor expansion, that the model is capable of reflecting every steady system behavior up to a higher order error, even if it differs from the (unknown) actual physical relationships. This fact also answers the question as to why a general model approach should be at all capable of describing different system behavior, and is a significant advantage of this approach.*

The verification of a black-box model is already incorporated in the modeling process. Validation (i.e. whether the model construct that was chosen beforehand is suitable for the question at hand) must be carried out separately.

Black-box modeling provides a basis for physical modeling when systems are completely unknown, by gathering facts and information from which the physical theories can be derived through hypothesis. This kind of modeling is also employed for technical questions without prior knowledge, and is the object of classic statistical experimental design, e.g. using factorial experimental designs.

*EXAMPLE: One example of empirical modeling is the corrosion behavior of stainless steel in gasoline containing ethanol, dependent upon the composition of the material and medium, the surface, temperature and time of exposure. More detailed relationships are not known here, so that black-box modeling has to be employed. In a series of experiments, the dimensions of the corrosive holes (system output) are measured to investigate the extent to which the intensity of the formation of corrosion depends upon the duration of exposure and all the other parameters mentioned above (system input).*



**Mixed modeling (gray box)**

In addition to the two modeling approaches described above, various mixed forms exist, which come under the term ‘gray-box modeling’.

If the system structure is known, but not the function of the system elements, the empirical modeling technique can be used at the element level to describe system behavior. Here, we refer to approaches that describe behavior in a model that explains behavior cover, [13], p.57.

On the other hand, structural knowledge and known physical relationships, if any, can be utilized to establish the model approach. A polynomial can be replaced by an exponential or logarithmic function from a physical model, for example. In the classic design of experiments using factorial experimental designs (which are based on polynomial techniques), such approaches can be achieved through the transformation of variables. This corresponds to approaches that explain behavior in a model that describes behavior.

Another type of gray-box modeling has its roots in the fact that physical modeling is mainly based on coupled systems of partial differential equations. Consequently, the characteristics of the relationship between predictor and response variables are not immediately obviously; rather, solutions only arise as the result of complex numerical calculations, such as the finite element method. Therefore, the model equation does not exist as an explicit analytical function, but implicitly as a “calculation scheme”, i.e. a computer-based calculating tool (solver) can deliver the corresponding response variable for discrete settings of predictor variables, on the basis of the complex physical model. In the context of DoE, this process is referred to as a computer experiment, as with physical tests. Here, rather than conducting experiments directly on a real system, simulations are carried out on an existing model. The resulting models are consequently referred to as meta-models, and can be designed as simply as possible (e.g. frequently as a polynomial of the first or second degree), with the aim of eliciting information about the nature of the relationship between predictor and response variables.

When determining the free parameters of a physical-based model, we refer to experimental parameter identification, or sometimes of “calibrating” the model. In the case of complex models, this is not a trivial process, as the model equation can no longer be explicitly represented, leading to so-called inverse problems. Since it is no longer possible to determine every free parameter of the model equation directly from the results of one experiment, a solution can only be obtained through optimization. With this type of modeling approach, parameter identification is simultaneously the verification of the model. Validation (whether the model is suitable for the task at hand) must be carried out separately, however.

This approach most closely approximates typical engineering procedure when dealing with complex technical questions, whereby a rough physical model is adapted through experiments on the basis of existing knowledge of the system. To a limited extent, the model can also be used for extrapolation.

*EXAMPLE: Examples of gray-box models are non-linear material models in structural mechanics, which were designed on the basis of rheological laws and possess several dozen free model parameters. These are determined by optimizing an adaptation of the model in line with complex test results.*

**Guideline:**

Wherever possible, white-box modeling is the ideal to which we should aspire. Here, experiments are only required for the purposes of verification and validation.

However, it is necessary to bear in mind that

- pure white-box modeling is only possible in exceptional cases,
- in many cases, gray-box modeling that deviates from the white-box approach to a greater or lesser extent is necessary,
- sometimes, pure black-box modeling may be required, depending on the task at hand.

In the latter two cases, insufficient theoretical knowledge is available (e.g. with new products and technologies), or cannot be acquired in time so that, as well as validation, experiments are also required for parameter and system identification.



### 1.1.2. Experiments and trials

Tests are also necessary for validating the quality of products. Virtually all relevant standards, rules, directives, etc. e.g. [6,7,8,9] require products to be subject to *trials* during their development.

Here, a trial is understood to be a test conducted on a technical product to verify and validate quality attributes.

According to DIN EN ISO 9000, verification is confirmation, through the furnishing of objective proof, that set requirements have been met.

Validation, on the other hand, is said by the above standard to be confirmation, through the furnishing of objective proof, that the requirements for a specific use or application have been met.

Trials are intended to prove that the set requirements, i.e. the functionality, reliability, robustness and safety of the product, have been satisfied. Furthermore, trials should provide knowledge about various influences (e.g. environmental) on products and processes and should, in particular, increase statement certainty.

The following **principles** illustrate the basics of trials:

- **Frontloading:**
  this means that the trial has to be started as early as possible.
- **Field relevance:**
  this means that the test specimens themselves and the trial conditions must correspond as far as possible to subsequent conditions of use.
- **Standardization:**
  an important basic principle is working with standard test procedures, which are governed by specific sets of rules, e.g. DIN IEC 68 “Basic environmental test procedure” or VDI 4005 “Influences of environmental conditions on reliability of technical products”.

Two **trial philosophies** with their respective advantages and disadvantages apply to the above principles.

1. **“Test to pass”:** The necessary product quality or reliability is demonstrated by a defined trial being passed without any adverse effect on the product. Such requirements are often based on higher-level sets of rules, so that the product can be regarded as **“fit for standard”**.

   Advantages:

   - Results are widely accepted
   - A degree of certainty, as one can always refer to the fact that a specified trial has been “passed”.

   Disadvantages:

   - The field relevance is often unclear
   - Possible field problems despite passing the trial.

2. **"Test to failure":** In (mostly) aggravated test conditions, the limits of the product are ascertained in an end-of-life test (considerable adverse effects until failure). This type of trial is an important element of development, which is intended to flag risks and open the door for improvements. The product's "rating" is not determined directly as a result of the trial, but in a separate step, which incorporates the results of the trial and field-relevant conditions of use. As field relevance has great importance in this approach, following a positive rating the product can be classed as **"fit for use"**.

   Advantage:

   When correctly applied, field problems cannot occur because the limits of the product and the field-relevant conditions of use are known, and the distance between the two has been classed as adequate.

   Difficulty:

   Determining the field-relevant conditions of use.



### 1.1.3. Experiments and quality

Today, companies face the challenge of introducing high-quality products (zero fault philosophy) of increasing complexity onto the market in the shortest possible time at competitive prices.

**Two universal approaches** can be regarded as the answer to this challenge; see [21], p.5:

- **Six Sigma** is an autonomous **quality improvement process** for reducing or eradicating existing faults/deviations on the basis of known quality and lean management principles (reactive approach)
- **Design For Six Sigma (DFSS)** is a **development methodology** for engineering fault-free, robust and reliable products from the very beginning, based on a solid understanding of the product and using suitable development methods (preventive approach), while at the same time making the most efficient possible use of resources, [21] p. 14.

**Experiments** have an important role to play in respect of both the classic Six Sigma quality philosophy and DFSS:

- **with Six Sigma** they are used primarily in the analytical phase, when it is a matter of establishing, analyzing and statistically proving the cause-effect relationships between problems and causes.
- **with DFSS** they are used as a tool for determining unknown cause-effect relationships and using them as the basis for designing, evaluating and optimizing robustness and reliability.

## 1.2. The role of statistics in experiments

A statistical approach is required when factors subject to variance occur during the testing of a technical system. Different strategies can be deployed for tackling this variance, or spread, in a sensible manner, depending on its cause.

### 1.2.1. The sources of variance

Generally speaking, three different sources of variance can be identified during a test:

- the specimens under test themselves
- the measurement of predictor and response variables
- the test arrangement.

Variance occurs with technical applications because the **material properties**, the **manufacturing conditions** and the **conditions of use** of products in quantities > 1 cannot be absolutely identical.

A measurement reading is also subject to variance. The deviation of the (corrected) measurement reading from a basically unknown true value is described as **measurement uncertainty** (see [15] Vol. 8).

Since it is impossible to guarantee, even under laboratory conditions, that only the influences under investigation are affecting the test object, and **unknown noise variables** can therefore also

occur, this variance must also be taken into consideration (influence of the test arrangement itself, including ambient conditions).

#### 1.2.1.1. Specimen variance

Two different strategies may be employed for specimen variance:

a) **Reducing variance:** The first priority is a design solution, in which the observed effects are significantly greater than the variance, e.g. by a targeted reduction of variance. Solutions such as these are always preferable, as they are extremely robust. Variance still occurs, as before, but plays a lesser role for the application. In this case, good results can be achieved in the test even without the use of statistics. However, a statistical approach will be necessary for obtaining quantitative information about the quality of the results (e.g. confidence intervals and significance).

b) **Overcoming variance**: If the technological limits are reached, with the result that any further reduction of variance is impossible or economically unviable, and it is comparable to the observed effects, steps must already be taken in the design phase to ensure that variance can be overcome (i.e. is known, stable and leads to the desired result). In this case, greater expenditure and an increased risk for the product concept can be expected. Here, the use of statistical methods is indispensable when conducting the experiment.



Basically, the first strategy is always preferable. For even if a mean effect is demonstrated through a statistical approach, the question remains as to whether this effect suffices for subsequent use, and the occurring variance leads to the desired result.

> *EXAMPLE 1: A component is to be employed in an injection component, which is to be manufactured using a cost-efficient but as yet unused shaping manufacturing process. The component's design is based completely on the use of this production method.*
>
> *At the start of development, we have no further information about the variance of the material properties of this component. However, it soon becomes clear that they are subject to considerable variation, due to the high sensitivity of the manufacturing process in the face of diverse predictor variables.*
>
> *Since no other manufacturing process appears cost-efficient for the design in question, considerable efforts have been made to reduce variance in production. The idea of simply "living with variance", e.g. through statistically well proven tests to determine the material properties, would have been extremely problematic, for the high level of variance was neither stable nor led to the desired results in terms of subsequent serial production. The aim must therefore be to find a stable manufacturing process before setting the design in stone, and to incorporate the cost risk of the less sturdy solution in the examination of costs.*
>
> *EXAMPLE 2: The fatigue strength of ultra-high-strength steel for use in high-pressure injection components is limited by fractures in non-metallic inclusions.*
>
> *The reduction in variance in the size of inclusions - and thus the variance in strength - has reached a high level and technology can currently achieve nothing further. A concept must therefore be elaborated for assessing the influence this variation in strength has on reliability. This concept includes tests with thorough statistical verification to determine the strength.*
>
> *This enables the high level of variance to be overcome, ensuring compliance with the demanding requirements for reliability to which components of this kind are subject.*

In the case of specimen variance that cannot be reduced any further but can be overcome, objective conclusions about a test can only be drawn when a statistical approach is employed. This is particularly the case when the variance is not much smaller than the effect under observation. What "much smaller" actually means and, in particular, whether the achieved effect and the variance are adequate for the technical solution, can only be decided in the light of the intended application. The following is recommended as a guide, as per [1], p.3: if the effects under observation are smaller than a fifth of the standard deviation, *or* the effect is to be determined quantitatively, a statistical approach *must* be taken.

In all cases, the confidence intervals must be determined through the use of statistics, in order to estimate the uncertainty of the conclusion.

#### 1.2.1.2. Variance in measurement results, noise variables

Where variance in measurement results and noise variables are concerned, the experiment requires objectivity: the test setup and measurement must not have any major influence on the test results. Here, two types of error may occur, see [2], p.21:

- **Systematic errors**: The measured mean values deviate from the actual mean value. These adversely affect the **accuracy** of the reading. Causes: loss of adjustment in measuring instruments, incorrectly calibrated test setups, systematic influence of the apparatus or experimenter, etc. Conclusions about accuracy can be drawn by means of comparisons with a different (reference) test method, or a standard sample. Systematic errors can be avoided by designing the experiment with care and conducting it as described in section 1.5.
- **Random errors**: Individual measurements deviate from the measured mean. These adversely affect the **precision** of the reading. Causes: unsuitable test methods, effect of noise variables, random reading and calculating mistakes, etc. Conclusions about precision are supplied by pure experimental error.



Proof of compliance with the above requirement is obtained by calibrating the test setup, analyses of measurement capability, etc. Details can be found in [15,16].

The use of a statistical approach when conducting an experiment is no remedy for inadequate measurement accuracy. Systematic errors, in particular, must be avoided. However, the use of statistics enables random noise variables to be controlled, increases the accuracy of the conclusions drawn and permits statements to be made about the significance of the results.

Statistics may not be used as a substitute for care when conducting the experiment. In this connection, a known variable, for example, which can be presumed to influence the result (e.g. tool wear) must always be recorded as an experimental factor (e.g. by blocking), and not treated as a noise variable.

### 1.2.2. The principles of statistical experimental design

Definition:

> Experimental design is understood to mean the creation of a design (including planning how the experiment is to be conducted and evaluated), within the framework of DoE.
>
> Statistical experimental design refers to a particular kind of experimental design, which uses statistical tools to permit a rational conclusion to be drawn about the matter under investigation, even if test results are subject to variance.

The following principles of statistical experimental design increase the reliability of conclusions drawn in the case of distributed variables [2]:

- **Replication** to improve accuracy (mean variance is less than individual value variance) and conclusions about the significance of the results.
- **Covering the experimental space,** e.g. by arranging the points under observation symmetrically, in order to increase effectiveness and to record interactions. The purposeful variation of several factors, which minimizes the number of tests but ensures the same data reliability, is an important approach in statistical experimental design.
- **Randomization** to neutralize unknown noise variables. If an unknown trend, e.g. in the quality of the specimens, is present due to tool wear, this can be neutralized by taking specimens at random from production for the test.
- **Blocking** for recording non-adjustable factors. For example, experiments that cannot be conducted within a narrow time frame can be divided into blocks, in order to cover any disturbing influences such as tool wear. If differences between the blocks occur, these can be recognized and eliminated.
- Confounding to achieve a limited reduction in expenditure. This method comes with disadvantages, however; see section 8.1.
- Sequential realization procedure, so that activities can either be continued or discontinued when enough results have been obtained.

The procedure described in section 4 takes most of these principles into consideration. Other approaches can be found in section 8; for basics, see [11].

### 1.3. Opportunities and risks in DoE

#### 1.3.1. Advantages, success factors and strengths

Advantages:

- **Objective procedure:** DoE enables objective, precise conclusions, including quantitative modeling and optimization on the basis of experimental data, and statements to be made about their significance, see [2], p.40.
- **Effective procedure:** Users report a shortening of project runtimes and a noticeable reduction in test expenditure [1]. Overall expenditure can be estimated early on, instead of the unplanned tactic of “just one more test”. The sample size is adapted in line with the task to be resolved, which ensures that the model is sufficiently accurate. Both insufficient accuracy (missing technical opportunities) and over-exaggerated accuracy (uneconomical procedure) are avoided. In addition, a maximum of information is obtained from the tests, e.g. in respect of interactions and the significance of the results.
- **Systematic procedure:** Users are frequently placed under the burden of extracting findings from a mass of data. DoE offers a clear, easily explained procedure for planning and evaluation. Compressing the data renders them easier to represent and communicate, enables them to be evaluated at a later point in time and to be incorporated in future tests.



Success factors:

- **Expert know-how:** The use of DoE requires sufficient knowledge and experience both in terms of the task to be resolved and the method itself, including adequate knowledge of statistics and the software used. As experience shows that one person alone can seldom satisfy all these criteria, the creation of a test team is recommended.
- **Suitable software:** If modeling is to be carried out with several factors on the basis of experimental data, the formulation of designs and most evaluations simply cannot be accomplished without the help of software. The Appendix provides information about suitable software.
- **Experimental design at an early stage:** If data (often of uncertain quality) are already available, subsequent evaluation is possible, but difficult, e.g. because of insufficient coverage of the experimental space. In this case, DoE experts face more exacting requirements, if they are to be able to estimate the significance of the results and draw conclusions. The comparability and up-to-dateness of historic data must be assured before evaluation can commence.
- **Sufficient resources:** Although DoE is a procedure that optimizes the use of resources, in certain circumstances tests may consume considerable resources. If there are insufficient resources for a DoE test, a different approach to testing is unlikely to lead to the desired goal either. The only remedy in this case is to adapt the objective aim of the experiment.

The key strength of DoE is encountered in tasks that are examining the dependence of a wide range of response variables on several predictor variables. Here, a complex system behavior can be described quantitatively with a manageable number of tests, even if physical modeling is difficult.

#### 1.3.2. Limits and dangers

What can we expect of DoE, what can we not?

DoE offers solutions to typical questions, for example (also see 2.3):

- Modeling: the purposeful acquisition of knowledge by means of cause-effect relationships in the system (product or process), in the event that physical modeling is difficult
- Prediction and optimization of product and process characteristics,
- Screening: determining what is important from a quantity of possible predictor variables.

However, the method requires **input from experts** at certain important points, which greatly determines the quality of the results.

- **Definition of factors and their ranges, and of response variables:** The omission of significant predictor variables is apparent only from the increased residual variance which, in turn, is only noticeable if empirical values are known. As a rule, no single approach enables us to determine whether a significant factor has been forgotten. If the residual variance (RMS error or coefficient of determination) appears to be unacceptable, other factors must be taken into consideration.
- **Selection of the modeling approach:** The modeling approach can only be validated to a limited extent by examining it with a higher-level approach, which means that the choice of modeling approach is extremely important.

  *REMARKS: This aspect should not be overstated, however. Through Taylor expansion, it is possible to show that each continuous function can be adequately approximated to a higher-order remainder using a polynomial approach. Furthermore, alternative model approaches that are not based on polynomials can also offer some assistance, see section 8.3.*
- **Care with experimental design, performance and evaluation:** This determines the quality of the data that are employed for modeling, also see section 1.5.
- **Plausibilization and physical interpretation of the results:** The results of modeling must be physically interpreted by experts. This physical interpretation is the point of the exercise, otherwise the tests are nothing but a collection of data and information.
- **Consequences and determination of further measures:** DoE offers an objective basis for decisions on the basis of experimental results, but cannot be a substitute for the decisions themselves, or for a discussion of the consequences for the actual engineering task.



Limits:

- **Excessive cost:** Even when using DoE, it may happen that too many or too expensive experiments are required, e.g.
  - different versions of specimens require expensive production processes,
  - insufficient number of specimens can be produced by Prototype Construction,
  - the specimens cannot be directly manufactured with the combination of properties required by the design,
  - insufficient time to complete the design, etc.
- **Limited findings:** DoE is an effective approach for extracting a maximum of information from test results. In certain cases, it may certainly make sense to attempt to reduce test expenditure to a limited extent, e.g. with screening tests; see section 8. However, this always entails a loss of information or the necessity of a priori assumptions, which lead to a risk of false interpretations, e.g. the mitigation, obliteration or even reversal of effects in low-resolution fractional factorial designs. Deciding where the **limits of sensibility** lie is always an important task for DoE experts. For example, designs with a resolution of at least V are recommended for response-surface tests, which would involve, at most, a halving of expenditure for most field-relevant cases; see section 8.1. Consequently, every attempt to save costs leads to marked losses in the findings obtained.
- **Weaknesses of this approach:** DoE is of limited suitability for experiments in which the emphasis is on the ability to extrapolate the results. Requirements of this kind can only be satisfied if at least a gray-box - or, better, still, a white-box - model exists. Here, physical approaches, e.g. using variable transformation, can be taken into consideration, see section 4.4. The same applies to systems with discontinuous or periodical behavior or a transient time response, due to the locality of the model approach.

Dangers:

- Experienced experimental experts may be tempted to undertake black-box modeling when physical relationships are known. These must, as far as possible, always be taken into consideration in modeling, see section 1.1.
- The experimental tests may deliver solutions “in the wrong place”, even though desirable solutions are required at a completely different level, e.g. in the design of the product, see section 1.2.
- DoE can encourage a false sense of safety. Where variables are subject to variation, there is always a small but finite and precisely defined residual risk that the conclusions drawn are

incorrect. Eradicating this residual risk would give rise to the need for infinitely large random samples, and cannot therefore be given serious consideration.

- Users report a certain degree of uncertainty, as the path to the result is not always obvious, particularly when complex systems are being modeled. The results are sometimes perceived as “mysterious”, although they are the result of a clear, systematic approach. The remedy here is to ensure that users have sufficient knowledge of the methodology.

### 1.4. DoE at Bosch

The advantages of employing DoE was recognized at Bosch early on. Technical experiments have been conducted with the aid of DoE (QA Information 01/1990) ever since 1988. It soon became clear that “the advantages described in the literature (e.g. minimum cost, maximum information) can be confirmed unreservedly”. For this reason, even then the recommendation was given to make increasing use of the advantages of DoE, in both engineering and production.



#### 1.4.1. The use of DoE in the product creation process (PCP)

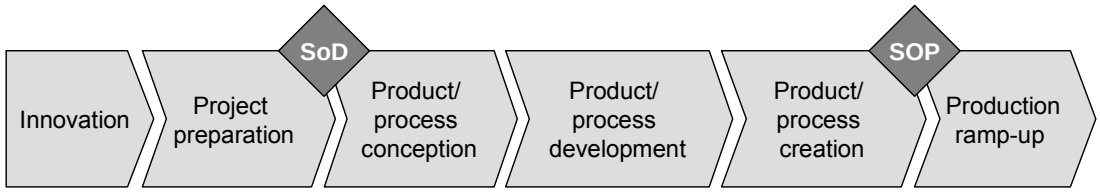



**Figure 1.3:** The product creation process

DoE is a supportive activity in engineering, and can be utilized at any time during the PCP, in every Division and for every product (and every process), if so required by the task at hand. The question as to whether to use DoE depends on the task that needs to be dealt with.

**Typical application examples:**

**Innovation phase:**

- Basic testing of novel physical effects, new technologies or product innovations by black-box modeling of the cause-effect relationships between relevant predictor and response variables, typically in research and development.

**Project preparation phase:**

- Determining important predictor variables with significant effects on response variables (screening).
- Modeling unknown cause-effect relationships for selected aspects of development in alternative product concepts.
- Testing different materials in terms of suitable properties (magnetic, electrical, strength, corrosion resistance, etc.) as part of platform engineering.

**Product and process design phase:**

- Experimental determination of complex functional relationships (characteristic maps) in the preferred design concept.
- Parameter identification of physically-based models.
- Model-based prediction of the behavior and robustness of products by means of computer experiments.

**Product and process engineering phase:**

- Predictions of behavior, robustness analyses and model-based optimizations of the design by means of computer experiments.
- Physical tests for the verification and validation of product requirements during a trial (trial design).

**Product and process implementation phase:**

- Systematic investigation of the influence of production parameters on product characteristics, with subsequent experiment or model-based optimization in terms of robustness and the minimization of variance.

**Production ramp-up and serial production phase:**

- Systematic revelation of disturbing influences in production.
- Experiment-based optimization with the aim of exploiting cost potentials.

### 1.4.2. Qualifications and points of contact

In order to use DoE, sufficient knowledge and experience is necessary, both in the specialist area of the task to be accomplished and in terms of DoE itself, including adequate background knowledge of statistics and the software used.

Necessary qualifications for DoE:

1. Study of user-oriented literature, e.g. [1] and this volume.
2. Seminars:
   - o General: Seminar TQ009 “Design of Experiments” (3 days). Prerequisite: Knowledge of basic statistics, e.g. TQ002 “Basic Technical Statistics” and TQ003 “Evaluation of Series of Measurements”.
   - o Aspects of trials: Seminar TQ023 “Testing Newly Developed Automotive Equipment Products” (3 days).
   - o Special techniques for conducting experiments: Seminars TQ007 “The Basics of Reliability”, TQ008 “The Reliability of Control Units” and TF024 “Component Fatigue Strength”. The contents extend far beyond the subject of DoE, however.
3. Those wishing to independently solve complex questions require sufficient practical experience. In this connection, we recommend involvement in DoE tests managed by an experienced colleague.
4. Questions that arise during subsequent activities should be discussed with other experts. General points of contact are, first and foremost, experts from the DoE expertise network (up-to-date list is in BGN at http://rb-knw.bosch.com/).
5. The study of more in-depth literature dealing with aspects of DoE that are of particular interest is worthwhile at any time, e.g. [4,5].

## 1.5. Elementary rules for designing and conducting experiments

The elementary rules described below are based on many years of practical experience, and are therefore of particular importance.

An experimental test generally consists of five vital steps (see Figure 1.4):

1. Task analysis
2. Systems analysis
3. Experimental design
4. Conducting the experiment, documentation
5. Evaluation of the experiment

**General principle:** Experiments should only be conducted for the purpose of achieving a quantitative determination of model parameters and for validating the model.

**Please note the following instructions** (also see Figure 1.4)**:**

**Step 1** (see section 2 for details):

- The objective of the experiment is set such that the expected cause-effect relationship can be quantitatively determined in and beyond the operating range
- The principal and the agent have clarified the achievability/feasibility of the experimental objective

**Step 2** (see section 3 for details):

- Predictor and response variables are known to the extent that the cause-effect relationships to be determined can be deduced from them qualitatively or quantitatively.
- The necessary response variables for the objective of the experiment have been defined, including the measurement principle (proof of measurement capability), measurement uncertainty, resolution, sampling rate and discontinuation criterion:
  - Select direct measurement variables during the test, if possible
  - If this is not possible: look for measurement variables with a known correlation to the response variable
  - If this is not possible (in experiments that age/damage components):
    a) Conduct the experiment in stages with defined sampling times
    b) Intermediate sampling for inspection/measurement, under consideration of the influence on the test (caution: this interrupts effective parameters!)
- Clarify and document additional active noise variables, e.g. through treatment, storage, intermediate sampling, analysis and renewed startup.

**Step 3** (see section 4 for details):

- Make sure that only the selected factors (predictor variables that are deliberately varied in the experiment) act as expected and all other predictor variables are kept as defined constants or are eliminated
- Define a suitable test apparatus, measurement procedure and measuring instrument for the requirements of the cause-effect relationship
- Define and establish the nature of test specimens
  - Carefully set design parameters, manufacturing process in conformity with test requirements, document all relevant parameters
  - Clarify manufacturability and define procedures for analyzing specimens
- Determine the number of specimens
- Plan the archiving of specimens and reference samples
- Design test run for each specimen, incl. a failure plan: event, actions to be taken (e.g. recording, deliberate power down to a safe state, and power up)
- Prediction of expected test results and comparison with the test objective (cause-effect relationship, significance)
- Plan evaluation and documentation

**Step 4** (see section 5 for details):

- Set up and inspect test apparatus, check suitability of equipment (measurement accuracy, tolerances, etc.)
- Define procedure for measuring established response variables
- Check nature and identification of specimens
- Conduct experiment in accordance with the design
  - Comply with and implement requirements for established factors, verify through suitable measurements and documentation (incl. test media)
  - Make sure that factors at the site of action conform to requirements
  - Record and log the test sequence for each specimen: all variables conform to design, incl. failure plan
- Document and draw conclusions from unforeseen failures

**Step 5** (see section 6 for details):

- Visualize actual values from the experiment
  - Test apparatus, measuring equipment used
  - Nature of specimens
  - Tests performed
  - Peculiarities during the test sequence
- Visualize the test results
- Evaluate data
- Compare with the expected result. In the event of deviations, look for other active predictor variables and assess the consequences
- Assess whether the objective of the experiment has been achieved
- Summary and final conclusions
- Transfer of knowledge

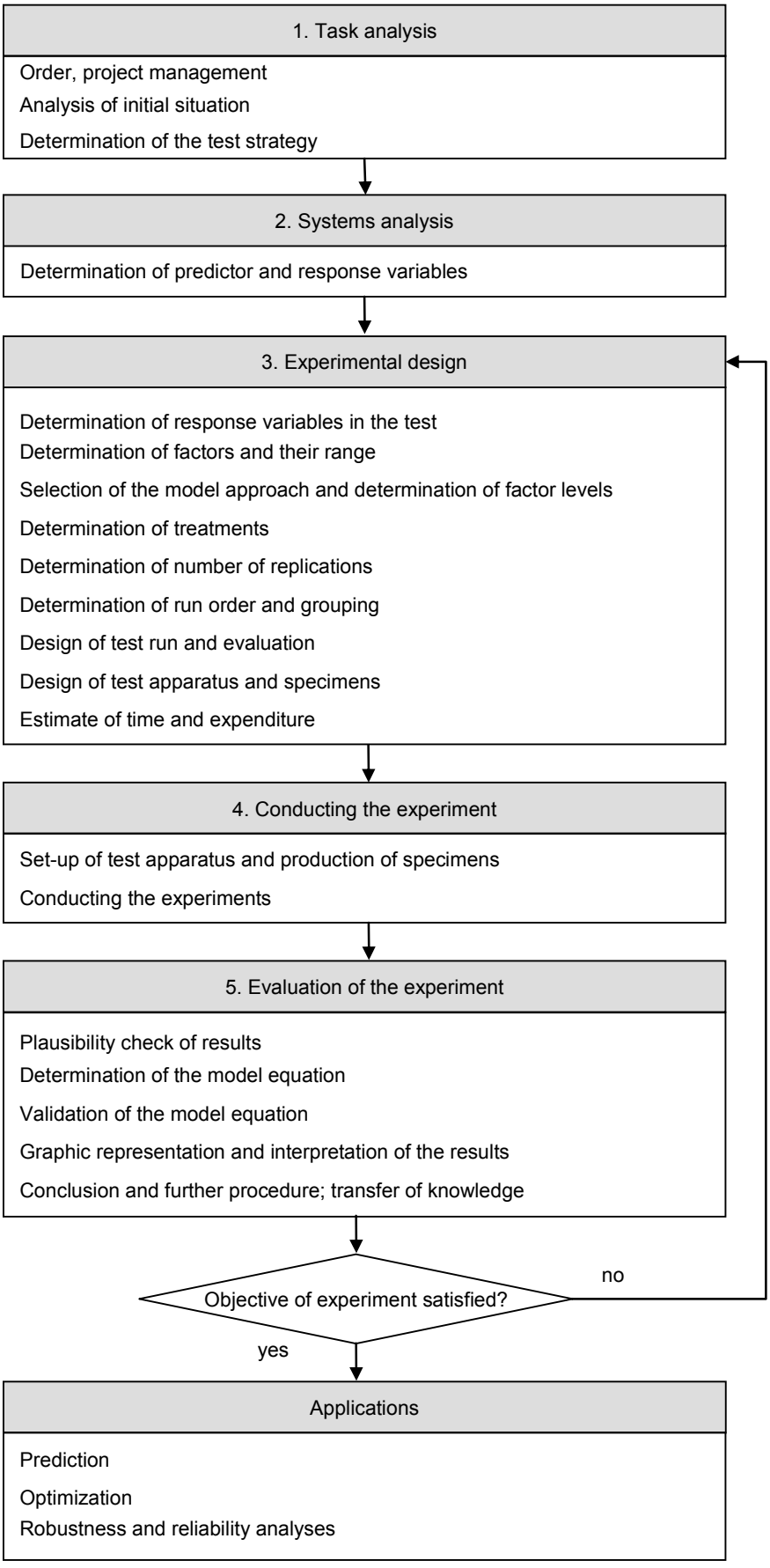



**Figure 1.4:** Sequence of events in an experimental test

## 2. Task analysis

This and the subsequent sections describe the typical steps of DoE in chronological order.

### 2.1. Order, project management

First of all, the principal appoints a coordinator with responsibility for deadlines, costs and the quality of results. The coordinator must receive a clear and precise order for the test, stating

- the task that is to be solved
- the anticipated available resources, such as time, money, machine and personnel capacity.

This is based on initial rough estimates. It will only rarely be possible to avoid the need for adjustments both to the task at hand and the approaches to a solution.

As a rule, it is also necessary to appoint a test team comprising experts, specialists in DoE and test experts, as the use of DoE requires sufficient knowledge and experience both in the specialist area and in DoE itself, including statistics and the software used. Responsibility in the team must be regulated clearly and without ambiguity.

The test team may also be convened later on, depending on the task at hand. In any case, however, the staff involved in the experiment must be clarified at an early stage. Their representatives must be involved in or at least informed of the solution.



### 2.2. Analysis of the initial situation

At the beginning, the test team must get to grips with the following questions:

- What is the aim of the experiment, i.e. what question is this experiment intended to answer? The importance of this point must not be underestimated, because clarity regarding the task at hand is a prerequisite for a successful solution and the principal's satisfaction. Here, detailed agreement is required.
- What is already known, and how can this prior knowledge be incorporated in the experiments? Completely new problems, for which no prior knowledge exists, are confronted only rarely. Moreover, all known physical relationships must be used for subsequent modeling.
- What might the results of the experiment look like, and what effect will the new findings have on the original question? Here, the team has to grapple seriously with the question as to whether the test is at all capable of delivering a desirable solution to the original task. It is seldom worthwhile to perform extensive tests in order to prove that a current concept is inadequate.
- How does the test fit in the surrounding test landscape, or in long-term overall strategy? Sometimes, a limited amount of additional expenditure leads to a long-term minimization of expenditure, or is the only means of enabling the meaningful analysis of results. A typical practical example of this is prematurely discontinued tests that may have consumed "only" 80% of the budget, but only achieved a 20 % gain in knowledge.

### 2.3. Determining the test strategy

It is necessary to define a suitable test strategy, with which a satisfactory solution to the task can be found. DoE offers suitable approaches for an array of typical questions, which are explained in more detail in the sections that follow:

- The purposeful acquisition of knowledge by means of cause-effect relationships in the system (product or process), if complete physical modeling is not possible.

  *REMARKS: This task is frequently a central issue in product development. Designs for one or more factors may be employed, depending on the number of factors.*

- Prediction of product and process characteristics in early stages of development.

  *REMARKS: DoE approaches can be utilized both for empirical modeling on the basis of physical tests and for the creation of meta-models with the aid of computer experiments, in the event that no physical prototypes exist.*



- Optimization of product and process parameters.

  *REMARKS: Here, a suitable approach is the creation of meta-models with the aid of computer experiments, followed by model-based optimization (e.g. using the gradient method), or alternatively using experimental parameter or system identification followed by model-based optimization, or the use of test-based optimization strategies such as the simplex algorithm.*

- Robustness analyses.

  *REMARKS: If no explicit models exist, the preferred method here is the Monte Carlo approach in combination with computer experiments and stochastic sampling techniques. Alternatively, explicit meta-models can initially be derived from computer experiments, with subsequent robustness analyses.*

- Robustness optimization.

  *REMARKS: The Taguchi or related methods constitute a suitable approach for reducing variance with the aim of achieving a robust product or process.*

- Determining what is important from a quantity of possible predictor variables and interactions.

  *REMARKS: This is also referred to as screening or sensitivity tests. Fractional factorial designs are a suitable solution.*

- Rapid detection and elimination of significant noise variables in production.

  *REMARKS: The D. Shainin method is suitable for this purpose.*

- Verification and validation of quality attributes of technical products, e.g. by demonstrating functionality, durability, robustness or safety.

  *REMARKS: Here, special experimental techniques for trials of components are necessary.*

Once the test strategy has been selected, final agreement must be reached with all involved parties and the principal regarding the objective of the experiment and of the test strategy.

## 3. Systems analysis

The first step towards conducting the experiment involves an analysis of the system with regard to its predictor and response variables, whereby both products and processes may be the subject of observation.

The experiment is an empirical study of the causal relationship between the influences and behavior of a system. First of all, it is necessary to identify the variables that describe the system behavior (response variables), and the variables that exert an influence on it (predictor variables). Depending on the question being posed, this can be a time-consuming step, but one that is extremely important. If it is skipped, subsequent problems are inevitable, e.g. because significant predictor variables have not been taken into consideration. Which predictor variables are actually to be varied during a test depends to a great extent on the question being asked, and is therefore decided later on.

The analysis of the system must, as far as possible, be based on physical modeling, see section 1.1.1. Existing knowledge of the system and physical laws must be used to identify response and predictor variables and to assess their relevance. At the same time, all known relationships provide valuable input for establishing the model approach to be used, see section 4.4.



The first step in identifying response and predictor variables is to make use of a functional model of the system under observation. This relationship is multi-faceted, however, and general rules can scarcely apply.

Here are two possible tasks:

- If the objective of the experiment is the creation of the model itself, e.g. as the basis for subsequent simulation, the predictor variables of the experiment will typically be the system inputs, and the response variables the system outputs.

  *EXAMPLE: An emission map of a combustion engine has to be defined for more in-depth analysis. To this aim, the engine is modeled as a system that "produces" emissions with a certain composition (system output) by means of a flow of fuel/air under a defined load (system input). The test enables us to determine the concentration of carbon monoxide and nitrous oxides in the exhaust gas (response variables) as a function of the defined load and the fuel/air mixture ratio (predictor variables).*

- Another relationship will typically arise when a system is being optimized. Here, predictor variables may be system parameters, which can be changed. Response variables may be both system inputs and outputs, e.g. if they must remain unchanged as an interface with the environment, or if they require optimization.

  *EXAMPLE: In the above example of an engine, one could be asked, for example, to attempt to improve efficiency by optimizing combustion without increasing the concentration of nitrous oxides in the exhaust gas. For this purpose, pistons and injection nozzles in various forms (system parameters = predictor variables in the test) are drafted, so that their influence on efficiency and the concentration of nitrous oxides (system outputs = response variable in the test) can be examined.*

If not enough is known about the cause-effect relationships, and a detailed functional model is therefore not available, or must be produced on the basis of the test, the following two-stage process is recommended for determining the relevant predictor and response variables:

- Firstly, a team of experts must determine the system variables and categorize them as predictor or response variables, depending on the matter at hand. Here, specialist knowledge and experience are indispensable. This first step serves as the basis for the subsequent assessment of the relevance of the variables in question, and can be backed up methodologically by various approaches (e.g. Ishikawa diagram). Variables that are ignored at this point cannot later be classed as relevant.

- The second important step consists in assessing the importance of the predictor variables in terms of their possible influence on the response variables. This process must be based, first and foremost, on known physical relationships. In the event that such relationships are not known, the technique of pairwise comparison, or the cause & effect (C&E) matrix can be of assistance.

  In the C&E matrix, the intensity of the influence is typically rated with the values 0, 1, 3 and 9 for "non-existent, weak, some and strong". Reasons must be stated for these ratings. The sum total of ratings of a particular predictor variable is referred to as its active total. Predictor variables can be ranked on the basis of their active totals. The sum total of ratings of a response variable is referred to as its passive total. This provides information about those response variables that may be relevant for the current study. Below is an example C&E matrix, which reveals that it is likely that

only the action of predictor variables 1 and 3 on response variable 1 has to be investigated in an experiment.

| | Response variable 1 | Response variable 2 | Active total |
|---|---|---|---|
| Predictor variable 1 | 9 | 1 | 10 |
| Predictor variable 2 | 1 | 0 | 1 |
| Predictor variable 3 | 3 | 1 | 4 |
| Passive total | 13 | 2 | |

**Table 3.1:** Example cause & effect matrix

The result is often also represented graphically in the form of a P (process) diagram, as a general description of the causal relationship between predictor and response variables. The representation also covers the (unknown) measurement errors that influence the response variables. An additional problem is generally posed by variance in response variables, because the causality between predictor and response variables cannot simply be maintained indefinitely. Here, help can be supplied by the introduction of additional, unknown predictor variables, referred to as noise variables, which also influence the response variable in a similar fashion to the unknown measurement errors. Thus, the variance in response variables can be explained by the occurrence of unknown noise variables and measurement errors, without losing the principle of causality and within the framework of a deterministic approach.



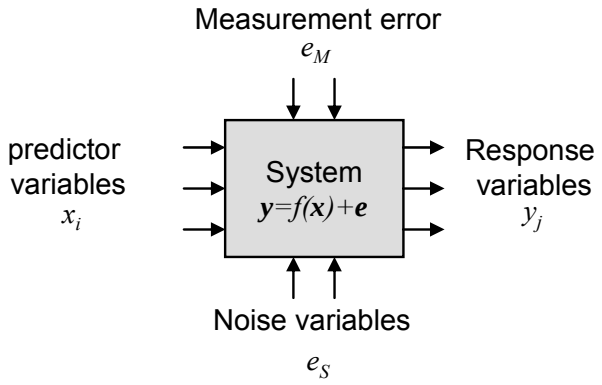



**Figure 3.1:** P-diagram of an experiment

## 4. Experimental design

Experimental design forms the heart of DoE. It enables historic data to be analyzed as well (in this case, please turn to chapter 6). However, only systematically created designs adequately cover the experimental area and withstand misconstruction. The comparability and up-to-dateness of historic data must be assured before evaluation can commence. The steps below are described in this chapter in more detail:

1. Determination of response variables
2. Determination of factors
3. Determination of the factor range
4. Selection of the model approach
5. Determination of factor levels
6. Determination of treatments
7. Determination of the number of replications
8. Determination of run order and grouping
9. Design of test run and evaluation
10. Design of test equipment and specimens
11. Estimate of time and expenditure



### 4.1. Determination of response variables

Firstly, from all response variables it is necessary to determine those that should be observed in the experiment. During this process, please note the following:

- Their number must be kept as low as possible, to reduce expenditure. Redundant response variables unnecessarily complicate the evaluation.
- All response variables that are relevant to the object under test must be recorded. The passive totals from the C&E matrix described in chapter 3 help to show which response variables may be relevant.
- It is not just the variables that are currently “problematic” that should be taken into account, since during optimization other response variables may also change - an effect that must be limited by constraints.
- The response variables under observation must be quantitative in nature, otherwise they cannot be evaluated in a meaningful way. Response variables of the “good-poor” type can be quantified, e.g. by the introduction of a rating scale.
- The variables must be directly measurable, if at all possible. If a response variable is not directly measurable:
  - a correlated variable can be sought, which can be measured directly, or
  - the experiment can be interrupted by a non-destructive intermediate test.

    *REMARKS: This is mostly not possible without repercussions, e.g. due to intermediate handling, storage, analysis or restarting the test apparatus.*

  - The experiment may also be conducted in stages with a destructive intermediate test. In this case, the variance of different specimens must be noted.

If several response variables exist, these can be evaluated separately. Difficulties occur here during optimization, because different (and possibly contradictory) response variables must be taken into consideration. So-called Pareto optimality may offer a fitting solution, see section 7.2.

The time at which the response variable is recorded must be clearly defined. The response variable must often be recorded on completion of the respective treatment. In this case, it is necessary to establish when the individual treatment can be considered completed. This is important for tests in which the end of the test may depend on a parameter, e.g. when determining the criterion to abort endurance tests. If necessary, a periodic sampling rate must be established for the response variable.

Last but not least, for both response variables and factors we must determine the measurement procedure, the measurement equipment and the measurement resolution with which the individual response variables are to be recorded. We must also clarify whether the accuracy and precision of the measuring equipment and measurement procedure are also adequate for the objectives of the experiment (proof of measurement capability).

### 4.2. Determination of factors

Factors are predictor variables that are to be changed in a targeted manner during the course of the experiment.

Factors must satisfy the following points:

- they must have a suspected significant influence on the response variables

  *REMARKS: Information to this effect is supplied by known physical relationships or the active totals from the C&E matrix presented in chapter 3. All factors that are suspected to exert an influence must be considered. The test itself provides clear proof later on as to whether or not a factor is actually relevant, but only if this factor was in fact taken into consideration in the test. Factors can be subsequently excluded from the model without difficulty (the tests in question are not "wasted", but can be used for more thorough statistical verification), but this is not the case for the inclusion of additional predictor variables. In cases of doubt, therefore, it is best to decide in favor of including a predictor variable as a factor in the test. Omitted significant predictor variables are only apparent from the increased variance in residuals which, in turn, is only noticeable if empirical values are known. Thus, as a rule, no single approach enables us to determine whether a significant factor has been forgotten. If residual variance appears to be unacceptable, further factors must be taken into consideration.*

  *According to [1], 3-6 factors are typically considered in tests, and seldom more than 10.*

- they must be directly or indirectly measurable
- they must be controllable

  *REMARKS: Non-controllable predictor variables (e.g. outside temperature, air pressure, but also production dimensions, fits) must be mastered, i.e. either recorded or neutralized through either experimental or statistic measures.*

- it must be possible to set them to be as precise and reproducible as possible

  *REMARKS: Otherwise, special approaches such as D-optimal designs are required, for which existing treatments in the experimental area can be preset as a boundary condition for producing the design.*

- expenditure for setting factors must be kept within reasonable limits
- they must be mutually independent, and they can be independently set (orthogonality condition)
- if possible, they should be quantitative in character

  *REMARKS: Qualitative factors give rise to difficulties (the model equation is discontinuous, interpolation between levels is meaningless, evaluation is conducted by analyzing variance, instead of through regression), but may nevertheless be observed.*

All other predictor variables that are not factors must be observed, documented and, as far as possible, kept constant.

### 4.3. Determination of the factor range

The operating range of the factors that is of interest must be determined, and the range for the test decided on this basis. The range must at least cover the operating range. However, we must beware of the problem of a "small level - small effect": the effect can sometimes be unclear, because the variation level is too small. The chosen level must therefore be sufficiently large to allow the effect to be visible. As a rule, the factor range extends noticeably beyond the operating range, particularly where increased variance is present. On the other hand, the factor range must not be excessively large either, for then local effects may influence the result.

Possible restrictions in the experimental area must be noted, e.g. areas that are pointless, unfeasible, dangerous or otherwise inadmissible. If such areas arise, special experimental designs, such as D-optimal designs, are required, see section 8.

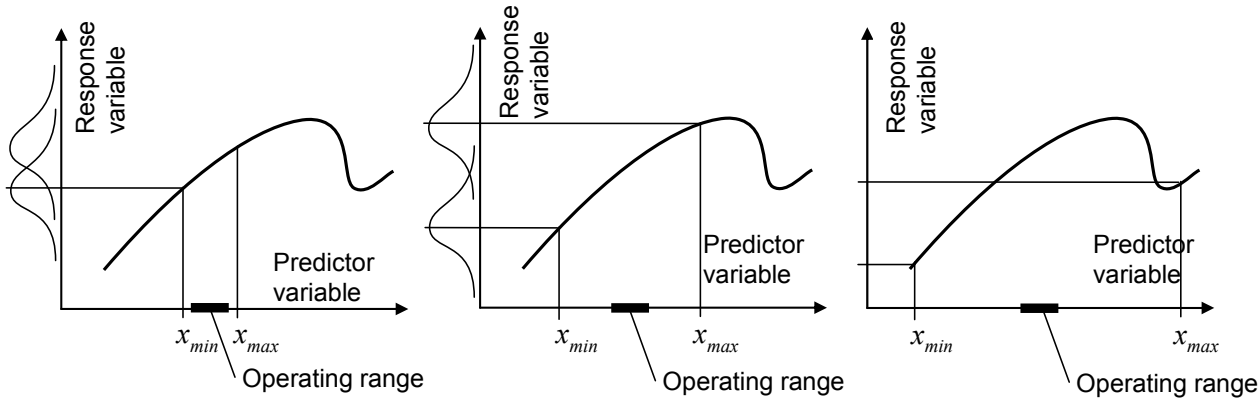



**Figure 4.1:** Operating range and factor range $(x_{min}, x_{max})$ of a factor (schematic). Left: Factor range is too small (effect too unclear with this degree of variance). Center: Factor range is correct. Right: Factor range is too large (local effects at the right-hand margin are not of interest for the range under observation).



### 4.4. Selection of the model approach

The model approach selected must be such that it is capable of describing the cause-effect relationships with sufficient accuracy. This constitutes an important input in experimental design. The assertion of the model approach is an important step, which cannot be skipped for reasons of time, for example, because errors made during this phase cannot be remedied later when the experiment is conducted. A deficient model will result in a **model error**, in addition to the types of errors mentioned in section 1.2.1.

To prevent misunderstandings: it is purely the model *approach*, i.e. the type of mathematical transfer function, that must be established here, not the transfer function itself. The complete function is determined only upon evaluation of the test. If a simple polynomial approach is selected as a transfer function, the only question that arises concerns the polynomial order. The free coefficients of the polynomial are determined by evaluating the subsequent tests, e.g. by means of regression.

When establishing the model approach, a physical model must be created, as far as possible, see section 1.1.1. Existing knowledge about the system and physical laws must be utilized by classing all known relationships in the model approach. As a minimum requirement, the model must not contradict empirical knowledge. If, for example, an exponential dependence, such as in population growth, is to be expected, a logarithmic approach would be inappropriate. The following example is intended to clarify the relationship between physical and empirical modeling.

*EXAMPLE: The power loss of a furnace is to be determined as a function of the furnace temperature at a given ambient temperature. Figure 4.2 shows a schematic diagram of the furnace wall and adjacent areas, together with the temperature profile from the interior of the furnace ($T_i$) to the outside environment ($T_a$).*

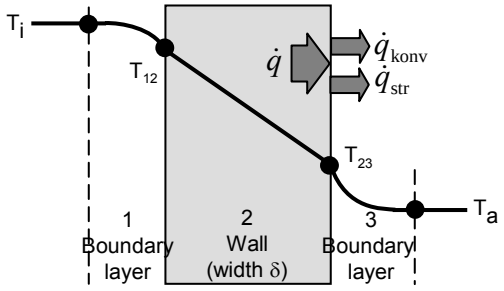



**Figure 4.2:** Temperature profile in the furnace wall

*The power loss of the furnace takes place through convection and radiation. These physical effects can be described by means of heat flow densities on the basis of thermodynamic relationships, as follows:*

$$\dot{q}_{konv} = \alpha_3 \cdot (T_{23} - T_a); \quad \dot{q}_{str} = \sigma\varepsilon \cdot (T_{23}^4 - T_a^4)\,. \qquad (4.1)$$

*$\alpha_3$ denotes the heat transfer coefficient in the boundary layer 3, $\sigma$ is a known natural constant (Stefan-Boltzmann constant), $\varepsilon$ refers to the emissivity of the wall. An observation of heat flow on the outer surface 2-3 of the furnace wall is expressed as:*

$$\dot{q} = \dot{q}_{\text{konv}} + \dot{q}_{\text{str}} = \alpha_3 \cdot (T_{23} - T_a) + \sigma\varepsilon \cdot (T_{23}^4 - T_a^4), \quad (4.2)$$

*whereby $\dot{q}$ denotes the heat flow density from the interior of the furnace. This must be conveyed via the furnace wall through heat conduction and via the boundary layer 1 in the interior via convection:*

$$\dot{q} = \frac{\lambda}{\delta} \cdot (T_{12} - T_{23}), \quad (4.3)$$

$$\dot{q} = \alpha_1 \cdot (T_i - T_{12}). \quad (4.4)$$

*$\lambda$ describes the heat conductivity of the wall (a material constant). Applying equation 4.4 in equation 4.3 results in*

$$T_{23} = T_i - \dot{q} \cdot \left( \frac{1}{\alpha_1} + \frac{\delta}{\lambda} \right), \quad (4.5)$$

*which allows the heat flow density to be expressed thus:*

$$\dot{q} = \alpha_3 \cdot \left( T_i - \dot{q} \cdot \left( \frac{1}{\alpha_1} + \frac{\delta}{\lambda} \right) - T_a \right) + \sigma\varepsilon \cdot \left( \left( T_i - \dot{q} \cdot \left( \frac{1}{\alpha_1} + \frac{\delta}{\lambda} \right) \right)^4 - T_a^4 \right). \quad (4.6)$$

*When solving this problem on the basis of the above equation, we now encounter the following difficulties:*

*- The heat flow density we are looking for is both on the left and right-hand sides of the equation (and in the 4th power in this case), which renders symbolic resolution impossible. Therefore, the desired relationship can only be stated implicitly:*

$$F(\dot{q}, T_i, \ldots) = 0 \quad (4.7)$$

*- The equation contains unknown constants such as $\alpha_1$, $\alpha_3$ and $\varepsilon$, which must be determined through experiments. The first two are dependent on the convection flow ratios in the boundary layer, the third on the nature (e.g. roughness and color) of the wall's surface.*

*At this point, empirical modeling is employed, under consideration of the known physical relationships. The following model approach would be suitable, because in equation 4.6 the inside temperature is contained in the 4th power:*

$$\dot{q} = a_0 + a_1 T_i + a_2 T_i^2 + a_3 T_i^3 + a_4 T_i^4 \quad (4.8)$$

*The coefficients $a_i$ are determined by measuring the heat flow at different inside temperatures in stationary condition. They are implicitly dependent on the above-mentioned constants and on the ambient temperature.*

If white-box modeling is not possible, a polynomial of the nth degree can be used as a general approach to empirical modeling. This is due to the fact that thanks to Taylor expansion, in the area under observation it is possible to show that every constant behavior can be approximated locally by a polynomial, with the exception of a higher-order remainder.

A multilinear approach takes account of linear effects, but also of interactions due to the mixed terms of the factors. For 1, 2 or 3 factors, a multilinear approach would look like this:

$$y = f(x_1) = a_0 + a_1 x_1 \quad (4.9)$$

$$y = f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 \quad (4.10)$$

$$y = f(x_1, x_2, x_3) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_3 x_3 + a_{13} x_1 x_3 + a_{23} x_2 x_3 + a_{123} x_1 x_2 x_2 \quad (4.11)$$

Higher-order approaches additionally take into consideration quadratic, cubic and other effects of the factors. One example here is a 2nd-degree polynomial:

$$y = f(x_1) = a_0 + a_1 x_1 + a_{11} x_1^2 \quad (4.12)$$

$$y = f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_{11} x_1^2 + a_{22} x_2^2 \quad (4.13)$$

$$y = f(x_1, x_2, x_3) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_3 x_3 + a_{13} x_1 x_3 + a_{23} x_2 x_3 + a_{123} x_1 x_2 x_2 + \ldots + a_{11} x_1^2 + a_{22} x_2^2 + a_{33} x_3^2 \quad (4.14)$$

It may be necessary to further simplify the approach. This is achieved, for example, by linearization, or by variable transformation if the model is non-linear in variables, but linear in coefficients. Physical considerations may have led to the deduction that the response variable is exponentially dependent upon a predictor variable, for example. The simplest procedure now consists in building a linear model with the predictor variable and the logarithm of the response variable, or to introduce a new predictor

variable through substitution, which is exponentially dependent upon the original predictor variable. See [2], p.34 for further details.

It is essential to check whether the model can meet expectations as regards conclusions and accuracy. Thus, a linear model would generally be disadvantageous in an optimization process, because curvature cannot be ignored in the optimum environment, see illustration below. The model approach must therefore reflect the environment of the area under examination (e.g. in the environment of extreme values of the response variables) as well and in as detailed a manner as possible. On the other hand, the model approach is generally related to the scope of the experiment, since the calculation of model parameters must be possible with the planned tests. A high-order approach can increase the scope of the experiment such that practical implementation is no longer possible. The aim is therefore to select the simplest model from all possible variants.

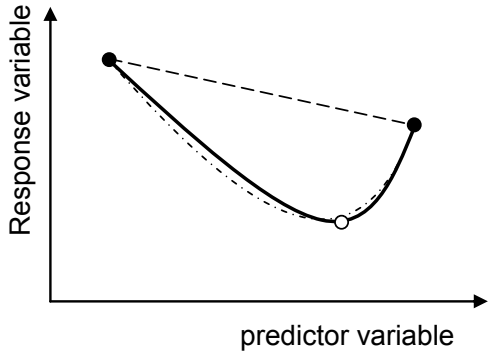



**Figure 4.3:** Schematic profile of a response variable in the area of the optimum (continuous line). A linear approach (dashed line) would not be able to find the optimum. A quadratic approach (dashed and dotted line) is more suitable.

### 4.5. Determination of factor levels

The quantity and value of the factor levels now have to be determined on the basis of the model approach. Here, most of the work has already been done in determining the model approach (see section 4.4), for every approach requires a certain minimum number of factor levels. A linear approach, for example, requires at least 2 factor levels, a quadratic approach 3 factor levels, etc. This is because it must be possible to determine the number of free parameters in the model with the available number of experiments.

So, although the problem cannot be resolved independently from the model approach, a certain degree of freedom exists in the selection of factor levels. First of all, we can go beyond the required minimum number of levels in order to check the expediency of the chosen approach: with a linear approach, for example, a treatment can additionally be placed at the center of the factor range, in order to monitor the deviation of actual behavior from linearity. In the case of quadratic approaches, for example, we can decide whether to employ central composite or face-centered designs, see section 8.2.

Furthermore, it is necessary to bear in mind that qualitative factors can only take on certain of the values defined by the task at hand. With the factor “material”, for example, they cannot assume any intermediate values between metal and plastic. The same applies to the factors supplier, operator or inspector.

## 4.6. Determination of treatments

This section discusses just 3 basic procedures; further approaches can be found in chapter 8.

### 4.6.1. Trial-and-error method

This frequently describes the simplest possible procedure, whereby testers do not proceed in a targeted manner, but simply “try things out” until a satisfactory solution is found. Tempting though this method may seem, we have to bear in mind that experimental research is generally very expensive, and the waste of resources has to be avoided.

I would also like to make it clear, at this point, that the argument that this method enables us to learn a lot about the problem at hand does not really hold water. The amount of knowledge gained is limited if no systematic approach is employed during the test. One can never be sure, for example, whether an optimum has been found, even if this is hit on by chance during an experiment (which is extremely improbable). In contrast, with model-based optimization, the optimum can also be predicted with some certainty even if it did not form part of the experiment.

It therefore does not make sense to pursue a non-targeted approach. Due to the disadvantages described here, this approach cannot really be regarded as an element of DoE, and is therefore not dealt with any further in this volume.



### 4.6.2. One-factor-at-a-time (OFAT) method

OFAT is used to describe experiments whereby the factors all initially have a basic setting, and then just one factor at a time is adjusted while all others retain this basic setting. When the next factor is adjusted, the preceding one is returned to its basic setting, and so on. In this way, all factors are adjusted in succession. The argument often deployed in favor of this method asserts that it enables the effect of a single cause to be assigned to precisely the factor that has currently been adjusted.

Firstly, it is true that if only a single factor occurs, this is certainly the most natural method.

Furthermore, this procedure can also be employed if several factors occur, but the effects and interactions of all factors except one can be predicted on the basis of physical considerations. A test using the remaining factor would be, in effect, an OFAT test.

*EXAMPLE: We will take a look at the “catapult” system, which can hurl balls over a certain distance.*

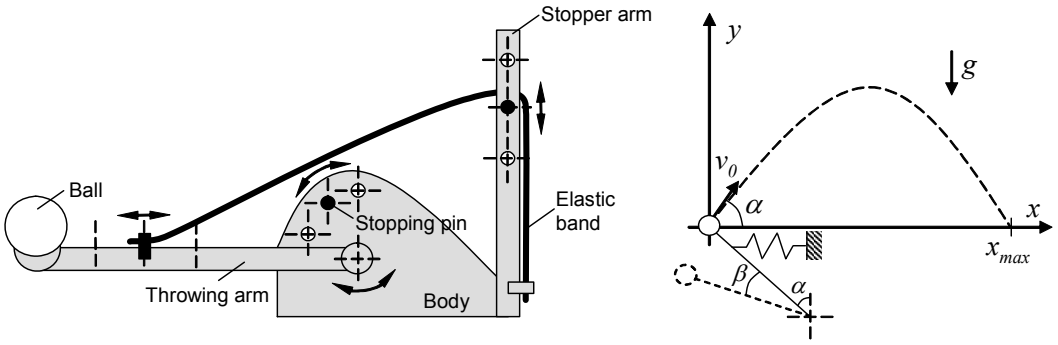



***Figure 4.4:*** *The catapult system*

*The catapult consists of a frame, a swiveling throwing arm, a fixed stopper arm and an elastic band, which connects the throwing arm to the front of the catapult frame via the stopper arm. The top end of the throwing arm contains a hollow that can hold a ball. Its function is extremely simple: the swiveling throwing arm is moved out of its resting position by tensioning the elastic band. When released, the throwing arm is pulled back to its resting position by the elastic band, causing the ball to accelerate. The throwing arm hits a pin when it reaches the resting position, which brings it to a stop. At this moment, the ball is projected forwards. The following settings influence the range:*

*- Deflection angle $\beta$*

*- Stop position (resting position) of the throwing arm, denoted by the angle $\alpha$*

*- Tensioning point on the throwing arm*

*- Tensioning point on the stopper arm*

*Our task is to create a model of the catapult to enable us to ascertain the settings required for a certain range in the future.*

*First of all, known physical relationships must be determined and taken into consideration in the modeling process. In the example here, this is possible without further ado: the tension of the elastic band upon the deflection of the throwing arm can be modeled by a spring, which is tensioned and generates a retractive force and thus also the accelerated movement of the ball until the moment it is fired. The trajectory of the ball once it is fired can be described by the laws of projectile motion:*

$$x(t) = v_0 \cos\alpha \cdot t \qquad (4.15)$$

$$y(t) = -g\frac{t^2}{2} + v_0 \sin\alpha \cdot t \qquad (4.16)$$

*The range $x_{max}$ can be deduced from the fact that $y(t_{max}) = 0$ at the moment $t_{max}$ the ball hits the ground. Consequently, equation 4.16 shows that*

$$t_{max} = 2\frac{v_0}{g}\sin\alpha \qquad (4.17)$$

*and, when included in equation 4.15*

$$x_{max} = \frac{v_0^2}{g} 2\sin\alpha\cos\alpha = \frac{v_0^2}{g}\sin(2\alpha) \qquad (4.18)$$

*However, some information is still missing for complete white-box modeling. Characteristics of system elements are not known. e.g. the elasticity constant of the elastic band, which means that $v_0$ is also unknown. However, the known physical relationships can be incorporated in a gray-box model and added to during tests to identify parameters.*

*The question now arises as to which of the four predictor variables mentioned above must be adjusted as a factor in a test. Firstly, it is clear that if the deflection angle $\beta$ is zero, the range will also be zero, irrespective of all other settings. With maximum deflection $\beta$= 90°-$\alpha$, the range is maximized also, but the achievable range depends upon the other settings. Physical considerations tell us, for example, that both tensioning points significantly influence the retractive force of the "elastic band" spring, and thus the acceleration and hurling speed of the ball. Maximum pre-tension means maximum hurling speed $v_0$ and according to equation 4.18 this also means maximum range $x_{max}$. All that is left is the stop position of the throwing arm. Here, we can deduce from the above equation that the range is maximized when the stop position is set to $\alpha = 45°$, whereby $\sin(2\alpha) = 1$.*

*In summary therefore, we can conclude — without having conducted a test — that with the tensioning points set to maximum pre-tension and a stop position of $\alpha = 45°$*

*- the maximum achievable range is produced with a deflection of $\beta = 45°$,*

*- a range of 0 is achieved with a deflection of $\beta = 0°$, and*

*- any range in between can be achieved only by adjusting the deflection angle $\beta$.*

*Consequently, the task now consists of examining $x_{max}$ solely as a function of the deflection angle $\beta$ factor. However, we can only establish which deflection angle would be required for which range through experiments, since important knowledge about the properties of the catapult system is lacking, as mentioned above. This can, of course, take the form of an OFAT test.*

*The test may be conducted such that the deflection angle $\beta$ is adjusted between 0 and 45° in several stages, and the associated range $x_{max}$ is recorded. The use of 2 stages would be preferable, namely the maximum setting and half the maximum, for in concrete cases $\beta$ results trivially in $x_{max} = 0$. An examination of the maximum setting alone would not suffice, because this would entail a linear dependence between the range and the deflection angle. It would be reckless to make such an assumption on the basis of existing knowledge. However, if the tests would be very expensive, it may be worthwhile to consider whether such an assumption might fit.*

*With the treatments set at half and at maximum deflection (and with knowledge about zero deflection), we would be in a position to create a quadratic model of the relationship between deflection and range:*

$$x_{max} = a_0 + a_1\beta + a_2\beta^2 \qquad (4.19)$$

*Higher-level approaches would be unlikely to reach the desired goal, for this would necessitate interest in a point of inflection, which in the present monotonous relationship (greater deflection results in greater ranges) can probably not be determined due to the limited accuracy of the tests. The number of replications in both*

*treatments must also be considered, however, depending on the variance of the range; this is described in more detail in section 4.7.*

Although OFAT tests constitute the most obvious method in the presence of just one (or quasi just one) factor, this type of test suffers major disadvantages when several factors are present:

1) Interactions between the factors are not registered. Here, 'interactions' signify that the effect of a factor on the response variable can depend to a great extent on the setting of the remaining factors. For example, if the remaining factors are set in a certain way, an increased factor may cause the response variable to increase also. When set in a different way, however, this effect may be intensified, weakened or even reversed, so that as the factor increases the response variable now suddenly drops. Therefore, to record this effect it does not suffice to examine the factor only with the remaining factors at one basic setting, as in an OFAT test.
2) If treatments have to be replicated several times in order to achieve the required accuracy in the face of variance in response variables, OFAT tests entail unnecessarily high expenditure.

To illustrate both of the above, we will take a look at a problem that examines the effect of two factors $A$ and $B$ – each between a minimum and a maximum value – on a response variable.



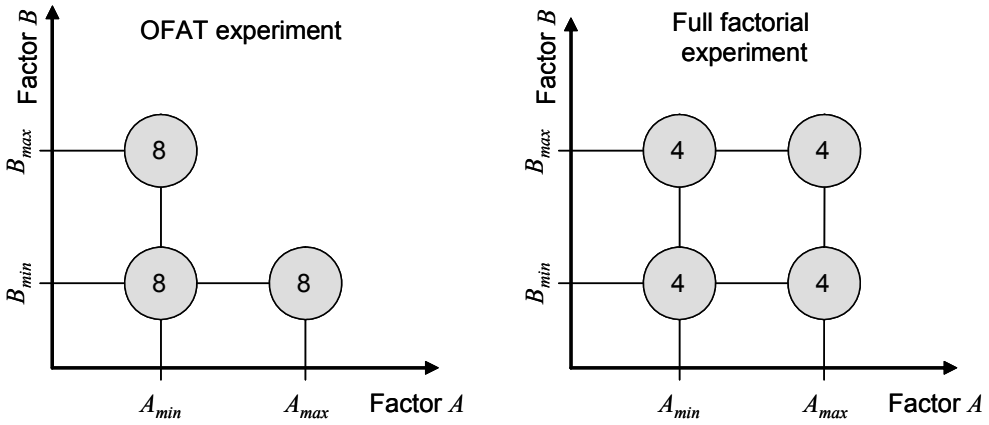



**Figure 4.5:** OFAT tests compared with a full factorial design

Experience tells us that the response variable is subject to variance, and for this reason each test variant requires 8 replications in order to record the effects correctly (details will be discussed at a later point). In an OFAT test with the basic setting ($A_{min}$, $B_{min}$), first the factor $A$ and then the factor $B$ must each be set to their maximum. Consequently, together with the basic variant, 3 treatments with 8 replications each - i.e. 24 individual tests - would be required. With each additional factor, 8 more tests would be needed.

Despite the relatively high test expenditure, however, important questions remain unanswered. The effect of factor $B$ on the response variable in the overall area of interest is known if factor $A$ is set to $A_{min}$, but with setting $A_{max}$ we only have information on a boundary point of the experimental area, that is, $B_{min}$. It therefore becomes clear that in an OFAT test, on the one hand, the basic combination is lent exceptional importance. On the other hand, the choice of this combination often appears so random that it has no objective basis. Why should ($A_{min}$, $B_{min}$) be more suitable than ($A_{min}$, $B_{max}$) or ($A_{max}$, $B_{min}$)?

An optimized approach is presented in the following section.

### 4.6.3. Factorial designs

In order to record interactions, an additional point is required at $A_{max}$ and $B_{max}$. This enables the effect of $B$ to be described in its entirety, both at $A_{min}$ and $A_{max}$. This method also has further advantages, particularly in respect of the required test expenditure. If this variant undergoes 4 replications, the same statistical certainty is achieved for each factor level as with the OFAT test. At $A_{min}$ a total of 8 test runs exist, namely 4 at ($A_{min}$,$B_{min}$) and 4 at ($A_{min}$,$B_{max}$). The same also applies to $A_{max}$. Nevertheless, significantly fewer runs are required: 16 (4 treatments times 4 replications) instead of 24. The additional treatment at ($A_{max}$,$B_{max}$) is used several times in a certain respect. As the number of factors grows, so does this effect. Thus, while (4+1)*8 = 40 runs would be required for an OFAT test with 4 factors, with the method described above $2^4$ = 16 runs would still suffice, with one run in each treatment. Asymmetric treatments hamper the comparability of results. Symmetrical coverage of the experimental area, on the other hands, brings nothing but advantages.

Conclusion:

The method described here requires significantly fewer tests, while simultaneously enabling the recognition of interactions between factors. Designs in which all factor combinations are varied are known as full factorial designs. These form the basis of statistical experimental design.



| $2^4$ design | | | | |
|---|---|---|---|---|
| $2^3$ design | | | | |
| $2^2$ design | | | | |
| Test | A | B | C | D |
| 1 | - | - | - | - |
| 2 | + | - | - | - |
| 3 | - | + | - | - |
| 4 | + | + | - | - |
| 5 | - | - | + | - |
| 6 | + | - | + | - |
| 7 | - | + | + | - |
| 8 | + | + | + | - |
| 9 | - | - | - | + |
| 10 | + | - | - | + |
| 11 | - | + | - | + |
| 12 | + | + | - | + |
| 13 | - | - | + | + |
| 14 | + | - | + | + |
| 15 | - | + | + | + |
| 16 | + | + | + | + |

**Table 4.1:** Full factorial design with 4 factors on two levels

In the case of a full factorial design with $k$ factors on two levels, $m = 2^k$ combinations arise. These designs are therefore referred to as $2^k$ designs. Table 4.1 shows the combinations for 2 to 4 factors on two levels. The designations “+/-” that are usually used indicate the upper or lower level of the factor in question. This plan can easily be expanded to any desired number of factors, by repeating the design for the lower and upper levels of the next factor.

As is evident from the above formula, the test expenditure increases sharply as the number of factors grows. Even with 6 factors, 64 treatments are required, which may also have to undergo several runs. At the same time, we have only 2 levels on which each factor must be varied, and have therefore selected only one linear model approach to describe its effect on the response variable. If we wish to take quadratic effects into account, for example, in a full factorial design 3 levels would be required for each factor, with expenditure rising to $3^k$ treatments. The most sensible way of tackling these issues

is illustrated by further approaches in chapter 8. These can, to a limited extent, reduce test expenditure.

### 4.7. Determination of the number of replications

The total number $N$ of test runs must now be established on the basis of the expected variance of the experiment and the required accuracy with which the effect under investigation must be determined.

According to [1], p. 78

$$N = 60\left(\frac{\sigma}{\Delta}\right)^2, \tag{4.20}$$

whereby the standard deviation $\sigma$ is a measure of experimental variance and $\Delta$ denotes the effect that is to be demonstrated with certainty. The probability of an effect that is actually present not being recognized is assumed to be 10% (type II error). The probability of an effect being recognized that is not actually present is 1% (type I error). The problem to be solved has two sides, as the aim is to recognize both positive and negative effects, see section 9.1.4.

If the test covers $m$ treatments, $n = N / m$ replications can be performed on each treatment. Here, the advantage of using factorial designs is that several factors can be examined simultaneously, almost without increasing the scope of the experiment.

This is further illustrated by Table 4.2. If an effect to the magnitude of $1\sigma$ is to be detected, $N = 60$ test runs will be required. Depending on the number of factors, this will give us

| No. of factors | Treatments | Replications per treatment $n$ |
|---|---|---|
| 1 | $2^1 = 2$ | 30 |
| 2 | $2^2 = 4$ | 15 |
| 3 | $2^3 = 8$ | 8 |
| 4 | $2^4 = 16$ | 4 |

**Table 4.2:** Possible replications with a constant number of runs

If the number of factors does not permit multiple replications ($n=1$), difficulties will arise during the subsequent analysis of significance. The following possibilities exist:

- The scope of the experiment is increased beyond what is strictly necessary. This is the ideal solution, as additional degrees of freedom are “always good for the statistics”.
- Variance is known or is ascertained separately through multiple replications of a single treatment, e.g. at the centre point. This alternative is also perfectly feasible.
- Fractional factorial designs are used, in which not all $2^k$ treatments are examined. This approach only makes sense if the resolution of the fractional factorial design is sufficient for the task at hand, see section 8.1.
- Variance is estimated by pooling during evaluation. This describes a procedure whereby insignificant, mostly higher order interactions in the normal distribution plot of the effects are recognized, eliminated from the model and therefore employed as additional degrees of freedom to provide additional statistical certainty. This tactic must always be the last resort, and must be deployed with the utmost caution, as the risk of subjective influence and the possibility of misinterpretation cannot be excluded, see section 6.3.1.

Conclusion:

The most sensible procedure is to cover just as many factors as necessary for conducting the required number of test runs, with further degrees of freedom retained in order to determine significance.

### 4.8. Determination of run order and grouping

At this point, the aim is to overcome the noise variables. This can be achieved by:

- Neutralization, either experimentally by keeping them constant, or statistically by means of a random sequence, or
- Recording them as well, either experimentally by blocking, or statistically by an analysis of covariance.

Experiments can be divided into blocks, within which treatments occur with the same frequency, as far as possible, and random fluctuations are as small as possible. If differences between the blocks occur, these can be recognized and eliminated. For example, experiments that cannot be conducted within a narrow time frame can be divided into blocks, in order to record any disturbing influences.

In order to prevent an unknown trend (e.g. systematically changing specimen dimensions due to tool wear) from falsifying the results, the treatments are performed in a block in random order. This process is known as randomization.



However, in certain cases changing the level of a factor can be extremely time-consuming, e.g. when a system is converted or in the event of long waiting times. Then, it may be necessary to sort the tests according to this level. Here, we have to reflect on how trends can be recognized and avoided.

### 4.9. Design of test run and evaluation

Expensive tests sometimes founder purely because a requirement that is not obvious to the operator of the test equipment is not explicitly defined or is not communicated with sufficient clarity. For example, the order of runs in a DoE design (incl. blocking and random sequence) can sometimes be complicated to put into practice. All those involved must clearly be made to understand that the correct order must be observed.

With this step, therefore, it is important to:

- **design the course of the experiment for each specimen**,

  *REMARKS: This includes the creation of a failure plan, which specifies each event and suitable action for its elimination. Such action may include recording, deliberately powering down the test equipment to a safe state and then powering up again, as far as possible in such a way so as not to impair the conclusiveness of the test.*

- **determine the necessary documentation** for conducting the tests,

  *REMARKS: This step includes recording (possibly through simultaneous documentation) and logging*

  - *the test sequence incl. any special events that have occurred,*
  - *the respective setting of factors, e.g. through the designation of specimens and other boundary conditions of the test, and*
  - *the results.*

- **plan the archiving of specimens and reference samples**, which will entail a certain amount of logistics in the case of extensive tests,
- **document** everything and **communicate** it to those entrusted with conducting the tests.

Now, at the latest, it is time to answer the question as to how the results of the experiment are to be evaluated, and whether the design is compatible with the planned evaluation. Statistical experimental design has a major advantage at this point, because a standardized procedure can be utilized for evaluation. The planned evaluation must be documented, particularly if a different person in the company is assigned this task, or if evaluation is to take place much later. In these cases, it must be assumed that at the time of evaluation, background information will already be lost, which can result in errors.

The expediency of the design in terms of achieving the intended objective must also be checked. For this purpose, the expected test results, incl. conclusions about their significance, must be compared with the objective of the experiment.

### 4.10. Design of test equipment and specimens

This step concerns the selection of suitable test equipment or, if this is not available, designing and drafting said equipment. During this process, the main focus must be on the aptitude of such equipment for the test, which must be assured (e.g. by selecting suitable tolerances).

In addition, the specimens must be designed. The attributes and number of specimens are initially based on the experimental design.

The following must be taken into consideration:

- A sufficient number of specimens must be available for the design and for reference samples, as well as a certain reserve. If future tests are likely, pre-emptive manufacture from the same material and production batch may make sense, to prevent noise variables from falsifying the result.
- Manufacturability must be clarified, especially the question as to whether all combinations of attributes specified in the design can be manufactured. This is often difficult, because in a certain sense we are dealing with boundary samples, which may be either physically impossible, irrelevant or unfeasible.
- Specimen characteristics that are not being examined as factors must also be established and documented. It may become clear later that these characteristics exert a noticeable influence and must therefore be considered as factors. For example, it may happen that, contrary to original assumptions, the surface roughness of a product plays a role in the experiment.
- The repercussions of the manufacturing process, e.g. on surface or residual stress, must be taken into consideration.
- Specimens must be labeled, to ensure correct documentation of results. Small components can be stored in labeled containers, for example.

### 4.11. Estimate of time and expenditure

The necessary resources such as time, money, machine and personnel capacity for the planned experiment must be estimated and compared with existing resources and requirements. It will often be necessary to adapt the experimental design and seek a healthy compromise between expenditure and the acquisition of knowledge.

However, we must constantly make sure that the conclusiveness of the test conforms to requirements. Sometimes, a critical minimum number of runs is required first, before any kind of conclusive result can be obtained.

## 5. Conducting the experiment, documentation

In conducting and documenting experiments, two different methods are employed, depending on whether real physical tests or computer experiments are planned to ascertain the results. The latter of these is increasingly gaining in importance when a validated model of the system under test is available that fully describes the behavior of that system on a deeper level, but in which the effect of various influences is not obvious due to the model's complexity, and must first be determined through simulation.

### 5.1. Physical tests

With real physical tests, firstly the test equipment must be designed, built and then prepared for conducting the tests, as described in section 4.10. This process has often turned out to be particularly time-consuming in practice, and must therefore be started as early as possible.

Preparation covers, above all, a check of the aptitude of the equipment for the test, and its calibration. Here, calibration entails comparing the results of a test conducted using the equipment with a known reference. In this step, it is necessary to make sure that the deviation from the reference lies within acceptable limits in respect of the test objective.

Checks must also be carried out to ensure that settings function correctly, particularly in the case of factors that are set directly on the test equipment. This is especially critical if the setting cannot be monitored later on in the test by online measurement of the actual values. If a measurement of this kind is conducted, the aptitude of the measurement process must be checked and/or verified. The same applies to measuring instruments that are used to determine the response variables. More on this subject can be found e.g. in [15,16].

The production of specimens can commence parallel to the preparation of the test equipment. Most treatments of the design are realized by appropriate specimen attributes, which is why particular care is essential in their production. This must also be discussed with the manufacturer of the specimens. Attributes such as dimensions, surface and mechanical properties must be examined judiciously after production. This will prevent tests being conducted with specimens that subsequently reveal themselves to be unsuitable for the test.

After the equipment and specimens have been produced and checked, the tests, designed as described in section 4, can be conducted. Specific procedure, incl. a failure plan and the necessary documentation, has already been established, as described in section 4.9.

Conclusions must be drawn for future tests in respect of any unforeseen events such as failures.

### 5.2. Computer experiments

As an alternative to the above procedure, if a validated model of the system under test is available (e.g. a structural-mechanical FE model of a complex product), "experiments" may also be conducted on the model. These are commonly referred to as computer experiments. In this case, the individual treatments are simulations, the findings from which can be transferred to the real system.

The advantage of computer experiments lies in the fact that, generally speaking, more simulations can be carried out than would be the case with real tests. However, simulations on especially complex models are also costly and time-consuming. It therefore makes sense to save resources using the approaches of DoE for these experiments, too.

The general procedure does not differ from that of a real physical experiment: the designs are drawn up as explained in section 4 and evaluated as described in section 6. The results do not supply a physical experiment but a simulation tool, known as a solver.

One important difference from a real physical test is that computer experiments do not produce any statistical error; this must be taken into consideration when selecting the type of experimental design. However, we can employ not just deterministic but also stochastic models, in which the occurrence of confounders can be simulated by superimposed random values, meaning that the use of statistical approaches does indeed make sense. Furthermore, distributed factors in analyses of robustness and reliability can be taken into account by stochastic designs such as the plain Monte Carlo.

Typical applications comprise three steps:

1. The relevant predictor variables are singled out from many through sensitivity tests.
2. The response variables undergo model-based optimization, taking into account the relevant predictor variables and other constraints.
3. The optimum design is subject to a robustness analysis. This answers the question as to what effect small, mostly random and unavoidable deviations from the optimum values of the factors have on the response variables, and what proportion of all possible combinations of attributes still satisfies the requirements of the design. In addition, a reliability analysis can be carried out, in which the proportion of reliable combinations of attributes can be determined from all the possible combinations. The two analyses differ solely in that the states being examined are very probable in the case of the robustness analysis, and very improbable in the case of the reliability analysis.

These methods enable us to investigate problems that cannot be overcome by real experiments. As one example, users report the weight optimization of an entire ship with 30,000 variables [18].

## 6. Evaluation

The aim of evaluation is to quantitatively determine the model equation, which describes the relationship between factors and response variables, and to validate it.

### 6.1. Plausibility check

For an initial plausibility check, it is a good idea to draw up a simple presentation of the important test results, i.e. response variables as a function of factors. This presentation should include the test equipment used, the measuring technology and any peculiarities arising during the course of the test. In this way, discrepancies will quickly become apparent when the results are discussed by the team, which must perhaps be examined more closely. This is difficult, however, when more than 2 factors are involved. In this case, a parametric representation or a representation solely of the treatments in the experimental area (without response variables) may be of assistance.

At any rate, this initial visualization must reveal the coverage of the experimental area. This is important because only systematically created designs cover the experimental area adequately and are robust against misconstruction, as opposed to designs that are simply wedged into place in existing data of dubious quality. Moreover, the comparability and up-to-dateness of historic data must be assured before evaluation can commence.



An analysis of plausibility must determine whether anything unusual, e.g. response variable with implausible values, are evident. The cause of any such unusual occurrences must be sought in order that the result can be corrected if necessary. In the simplest of cases, the cause is transmission or typing errors. Simple reading errors can also sometimes be corrected. If the cause remains unclear or a factor was incorrectly set, the run in question must be repeated. If this is not possible, this run is excluded from further analysis, for outliers can significantly distort the result of the test, and must not be used. The utmost caution must be exercised when proceeding thus, however, especially if there is only one run. In this case, a suspected outlier may signify an unsuspected effect. Statistical outlier tests can help with this decision.

### 6.2. Determining the model equation

Firstly, two simple examples provide an introduction to the subject. Next, the general procedure is explained through regression. The mathematical bases are contained in the Appendix.

#### 6.2.1. Example of single-factor test on 2 levels

Task: The thermal expansion of an alloy is to be determined through experiments. Several replications are conducted at 2 temperatures. Table 6.1 shows the mean values of the results of the replications.

| Test | *T, °C* | *L, cm* |
|---|---|---|
| 1 | 25 | 100.04 |
| 2 | 100 | 100.16 |

**Table 6.1:** Thermal expansion of an alloy

We are assuming that a linear relationship exists between the change in length and the temperature, and wish to determine the linear equation that will enable us to calculate any intermediate values. This equation is:

$$L = A_0 + A_1 T \tag{6.1}$$

Through the process of coordinate transformation, the pair of variates $(T_1, T_2)$ is formally translated as $(-1, +1)$. The transformation equation is:

$$x = \frac{2}{T_2 - T_1}(T - T_2) + 1. \tag{6.2}$$

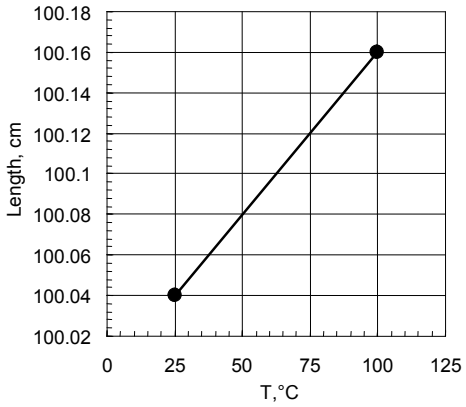



**Figure 6.1:** Linear expansion of an alloy



Here, it is easy to check that $x(T=T_1)=-1$ and $x(T=T_2)=+1$. In the transformed coordinate system, the linear equation is:

$$L = a_0 + a_1 x \,. \tag{6.3}$$

The equation (6.3) and the two test results give rise to the following:

for $x = +1$: $\quad 100.16 = a_0 + a_1$, (6.4)

for $x = -1$: $\quad 100.04 = a_0 - a_1$. (6.5)

At this point, the reason for the coordinate transformation becomes obvious; namely, the coefficients $a_0$ and $a_1$ can be calculated with great ease by adding or subtracting the equations (6.4) and (6.5):

$$a_0 = \frac{L_2 + L_1}{2} = \frac{100.16 + 100.04}{2} = 100.1 \qquad a_1 = \frac{L_2 - L_1}{2} = \frac{100.16 - 100.04}{2} = 0.06 \,. \tag{6.6}$$

In general, therefore the coefficient $a_0$ is the mean of the two lengths, while coefficient $a_1$ is the half difference between the lengths.

The difference in length at the lower and upper temperature is referred to as the *effect* of the temperature factor. Thus, we can also calculate the coefficient $a_1$ from the half effect of the temperature. This constitutes a vital relationship between the effect of the factor and the coefficients of the model equation; thus by calculating the effects, the coefficients of the model equation are also automatically determined, and vice versa.

In the transformed system, the linear equation (6.3) is:

$$L = 100.1 + 0.06x \,. \tag{6.7}$$

The linear equation (6.1) in the original system can be obtained through inverse transformation (6.2)

$$L = 100.1 + 0.06\frac{T - 62.5}{37.5} = 100 + 0.0016 \cdot T \,. \tag{6.8}$$

### 6.2.2. Example of dual-factor test on 2 levels

This example is intended to elucidate the procedure for evaluating experiments on the basis of Ohm's Law. We will therefore put ourselves in the shoes of an experimenter who wishes to determine the relationship between voltage, current and resistance with the aid of a simple experiment. He is looking for

$$U = f(R,I) \,. \tag{6.9}$$

Let us assume that he has conducted test runs with several replications, and has measured the following mean voltage:

| Test | *R, ohm* | *I, A* | *U, V* |
|---|---|---|---|
| 1 | 20 | 4 | 80 |
| 2 | 60 | 4 | 240 |
| 3 | 20 | 12 | 240 |
| 4 | 60 | 12 | 720 |

**Table 6.2:** Ohm's Law

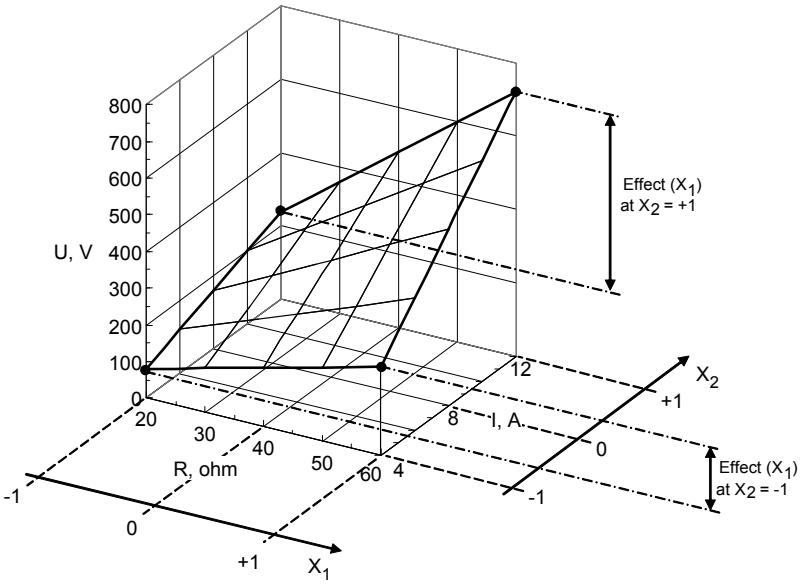



**Figure 6.2:** Ohm's Law

In order to determine the relationship, first of all a multilinear approach is employed for two factors:

$$U = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 \,. \tag{6.10}$$

The following transformation can further simplify the evaluation:

$$x_1 = \frac{R - \frac{R_2 + R_1}{2}}{\frac{R_2 - R_1}{2}} = \frac{R - 40}{20}\,,\quad x_2 = \frac{I - \frac{I_2 + I_1}{2}}{\frac{I_2 - I_1}{2}} = \frac{I - 8}{4}\,. \tag{6.11}$$

The result is a system of equations from which the free coefficients of the approach can easily be determined (the voltages $U$ arising during the individual test combinations are on the right):

$$\begin{aligned}
&x_1 = -1\,,\ x_2 = -1\,,\ a_0 - a_1 - a_2 + a_{12} = 80\,,\\
&x_1 = +1\,,\ x_2 = -1\,,\ a_0 + a_1 - a_2 - a_{12} = 240\,,\\
&x_1 = -1\,,\ x_2 = +1\,,\ a_0 - a_1 + a_2 - a_{12} = 240\,,\\
&x_1 = +1\,,\ x_2 = +1\,,\ a_0 + a_1 + a_2 + a_{12} = 720\,.
\end{aligned} \tag{6.12}$$

Consequently:

$$a_0 = \frac{80 + 240 + 240 + 720}{4} = 320\,,$$

$$a_1 = \frac{720 - 240}{4} + \frac{240 - 80}{4} = 160\,,$$

$$a_2 = \frac{720-240}{4} + \frac{240-80}{4} = 160\,, \tag{6.13}$$

$$a_{12} = \frac{720-240}{4} + \frac{80-240}{4} = 80\,.$$

Applying (6.13) in the approach (6.10) gives rise to:

$$U = 320 + 160x_1 + 160x_2 + 80x_1x_2\,. \tag{6.14}$$

Through inverse transformation of (6.14) on the basis of equation (6.11), we obtain the model equation in the untransformed coordinates:

$$U = 320 + 160\frac{R-40}{20} + 160\frac{I-8}{4} + 80\left(\frac{R-40}{20}\right)\left(\frac{I-8}{4}\right) = RI\,. \tag{6.15}$$

In this structured example, the correct solution necessarily arises (Ohm's Law) because, with the multilinear form, exactly the right approach was selected. A more complicated functional relationship involving quotients or powers of the predictor variables could only roughly be described with this approach.



### 6.2.3. Tests with k factors on 2 levels

If a full factorial test with $k$ factors each on 2 levels with a total of $N$ runs and $m=2^k$ treatments was conducted, the effects and interactions and the coefficients of the model equation can easily be calculated using the evaluation matrix of the experimental design.

In this connection, take a look at the design below:

| Test no. | A | B | … | Results $y_{ij}$ | Mean $\bar{y}_i$ |
|---|---|---|---|---|---|
| 1 | - | - | … | $y_{11}, y_{12}, \ldots, y_{1n}$ | $\bar{y}_1$ |
| 2 | + | - | | $y_{21}, y_{22}, \ldots, y_{2n}$ | $\bar{y}_2$ |
| … | | | | | |
| m | + | + | | $y_{m1}, y_{m2}, \ldots, y_{mn}$ | $\bar{y}_m$ |

**Table 6.3:** Full factorial experimental design with $k$ factors each on 2 levels

When each treatment is examined several times, first of all the evaluation can help us to determine the mean of the runs on each treatment; the respective column has been added to in the above matrix.

In this case, the multilinear model equation is:

$$y = a_0 + \sum_{i=1}^{k} a_i x_i + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} a_{ij} x_i x_j + \ldots + a_{12\ldots k}(x_1 x_2 \ldots x_k)\,. \tag{6.16}$$

In total, this contains $2^k$ coefficients and just as many terms: one free term, $k$ terms with the factors and mixed terms with 2, 3, etc. factors, up to a mixed term with all $k$ factors. All mixed terms represent dual, triple etc. interactions.

The effect of factor j can simply be calculated as

$$\Delta_j = \frac{2}{m}\sum_{i=1}^{m} \mathrm{sgn}_{ij} \cdot \bar{y}_i\,, \tag{6.17}$$

whereby "sgn" denotes the corresponding factor in the evaluation matrix. Interactions can be handled in the same way; their +/- column is calculated as a product of the corresponding columns of the involved factors. For example, with a dual interaction $x_1x_2$ an additional column is formed, in which columns A and B are multiplied together. Next, this +/- column is used to calculate the interaction as described above.

The coefficients in the multilinear form are then revealed as:

$$a_0 = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i\ ,\ a_j = \frac{\Delta_j}{2}\ ,\ j = 1...2^k - 1\,. \tag{6.18}$$

### 6.2.4. Regression analysis

In the calculation of the model equation as described above, as many treatments were present as there were unknown coefficients. In this way, we were able to produce a system of equations with $m$ equations (from the test results of the $m$ treatments) and $m$ unknowns (coefficients), from which these coefficients could then be determined. If an experiment incorporates more treatments (but not necessarily replications) than coefficients, more equations than unknowns would exist, with the result that the system of equations would be over-determined and the above procedure would no longer be possible. In this general case, a regression analysis is necessary. Performing more tests than is strictly necessary for determining the unknown coefficients of the model equation has considerable advantages. For one, it enables us to check whether or not the chosen model approach was suitable for the task at hand.

In a regression, the parameters of an already defined mathematical approach are adapted so as to permit the best possible description of the experimental data using the chosen model. This is achieved by means of the least-squares method, whereby the coefficients of the model are determined in such a way as to minimize the sum of deviations of treatments from the model curve. We do not go into the formal mathematical basis at this point; more information on this subject can be found in the Appendix.



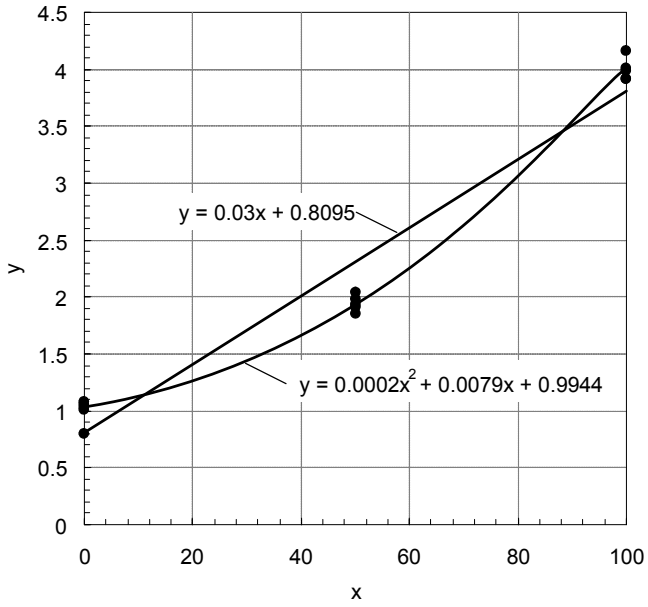



**Figure 6.3:** Linear and quadratic regression of a single-factor test at 3 levels with 5 replications per level

In the event of a non-linear relationship between response and predictor variables, in most cases the attempt can be made to remedy the problem through quasi-linear regression, by the suitable transformation of variables. If, for example, we are expecting a power dependence of the type

$$y = a_0 + a_1 t^2\ , \tag{6.19}$$

the initial non-linear problem can be translated as a linear relationship through substitution $x = t^2$:

$$y = a_0 + a_1 x\ , \tag{6.20}$$

which can be tackled with linear regression.

If an approach is to be adapted in line with test results through regression, whereby the response variable depends upon more than one predictor variable, this is referred to as multiple regression. Evaluation takes place according to the same principles as with simple regression.

Conducting a regression analysis is a complex proposition, especially when there is more than one predictor variable. At this point I would therefore like to recommend the use of suitable software tools. More information on this subject can be found in the Appendix.

### 6.3. Validation of the model equation

Generally speaking, when a test is replicated (repeated), the response variables will vary to a greater or lesser extent as the result of disturbing influences. The effects and coefficients of the model equation calculated on the basis of the test results are also subject to certain fluctuations, and generally always differ from zero, even if the true but unknown effect is equal to zero. The question therefore arises as to whether the calculated effects differ from zero purely by chance, or not. If they differ by chance, this would be insignificant; if however, this is not the case, they must be regarded as significant.

The aim of this section is to show how an effect (or coefficient of the model equation) can be classed as significant or not.



#### 6.3.1. Simple significance rating with $2^k$ designs

If a full factorial test with $k$ factors each on 2 levels with a total of $N$ test runs and $m = 2^k$ treatments is conducted, the process of mean comparison can be utilized to investigate the significance of effects and interactions, see [1], p.99. This method is based on the fact that the effects can be calculated as the difference between two means of a factor.

In this connection, we examine the following evaluation matrix of a design:

| Test no. | A | B | … | Results $y_{ij}$ | Mean $\bar{y}_i$ | Variance $s_i^2$ |
|---|---|---|---|---|---|---|
| 1 | - | - | | $y_{11}, y_{12}, \ldots, y_{1n}$ | $\bar{y}_1$ | $s_1^2$ |
| 2 | + | - | | $y_{21}, y_{22}, \ldots, y_{2n}$ | $\bar{y}_2$ | $s_2^2$ |
| … | | | | | | |
| m | + | + | | $y_{m1}, y_{m2}, \ldots, y_{mn}$ | $\bar{y}_m$ | $s_m^2$ |

**Table 6.4:** Evaluation matrix of a design

When each treatment is examined several times, the evaluation can help us to determine the mean and the variance of the individual runs on each treatment; the respective columns were added to in the above matrix.

The effect of factor j can be calculated as

$$\Delta_j = \frac{2}{m} \sum_{i=1}^{m} \mathrm{sgn}_{ij} \cdot \bar{y}_i \,, \tag{6.21}$$

whereby “sgn” denotes the corresponding factor in the evaluation matrix, see section 6.2.3. The variance

$$\overline{s_y^2} = \frac{1}{m} \sum_{i=1}^{m} s_i^2 \,, \tag{6.22}$$

which can be calculated as a mean of the individual variances, constitutes an estimate of experimental variance. This is sometimes also referred to as pure error.

Since each effect forms the difference between two means each of $N/2$ individual values, its standard deviation can be estimated by means of

$$s_\Delta^2 = \frac{4}{N} \overline{s_y^2} \,. \tag{6.23}$$

The significance of each effect can now be stated through the process of mean comparison based on the t-test, as described in section 9.1.3. The associated test statistic can be calculated thus

$$t_j = \frac{\Delta_j}{s_\Delta}. \tag{6.24}$$

If $t_j > t(f, 99\%)$, the null hypothesis must be rejected, and the associated effect regarded as significant (the statistic $t(f, 99\%)$ is set out in a table in the Appendix). The degree of freedom is $f = N - 2^k$ ($k$ - number of factors, $N$ – total number of runs). A graduated rating of the significance of the effect can be carried out with the aid of Table 6.5:

| Lower limit | Parameter | Upper limit | Assessment |
|---|---|---|---|
| | $t \leq$ | $t_{f;99\%}$ | not significant |
| $t_{f;99\%}$ | $< t \leq$ | $t_{f;99.9\%}$ | significant |
| $t_{f;99.9\%}$ | $< t$ | | highly significant |

**Table 6.5:** Significance assessment using the t-test

**Significance in tests with one run**

If a test with a single run $n=1$ has been carried out, difficulties will arise when estimating the variance of the individual values and therefore the variance of the effects and their significance. This situation is best overcome if the variation of other tests is already known, or can be determined separately by performing several replications of a single traetment (often at the centre point).

If the above strategies are not possible, variance can be estimated by a process known as pooling, whereby random effects and interactions are removed from the model. Background: If the actual effects and interactions are zero, the calculated random effects and interactions are normally distributed with the mean value zero. Here, all outliers have a high probability of being actually occurring effects or interactions. All random effects can now be eliminated from the model and used for greater statistical certainty. “Randomness” is recognized in that these effects and interactions are positioned on a straight line in the probability plot.

This tactic must always be the last resort, and must be deployed with the utmost caution, as the risk of subjective influence and the possibility of misinterpretation cannot be excluded. We recommend employing this method only with more than 5 factors, and even then only for higher order interactions, see [1], p.107.

### 6.3.2. Validation of regression

It is generally necessary to ascertain the coefficients of the model equation by means of regression analysis. Certain evaluations and characteristic values can be used to analyze the quality of the regression analysis, and these are explained below.

#### 6.3.2.1. Assessing the goodness of fit

First of all, a visual inspection needs to be carried out to determine how well the calculated model equation concurs with the test results. To achieve this, more treatments (not more replications) must be available than there are free coefficients in the model equation. For example, in a single-factor test on two levels, the middle of the interval can also be recorded as an additional treatment. This enables us to check the “deviation” of the test result from the assumption of a linear relationship. If only two treatments are present a check of this kind is not possible, because 2 points always define one straight line.

The advantage of this additional treatment is also evident in that when there are 3 points, evaluation is possible through linear regression, whereby the goodness of the linear fit can also be quantified with the aid of the characteristic values described below. This would not be possible with only 2 treatments, see section 9.1.6.

Furthermore, it makes sense to state and quantitatively assess the **quality of the regression** using the following characteristic values:

- One common figure is **RMS error** $S_R$ (root mean square error), also known as standard error or standard deviation of the regression. This is calculated as the root of the mean of the residual squares. Residuals are deviations of the response variable in the experiment from the prediction of the model equation

  $$r_i = y_i - \hat{y}_i \,, \tag{6.25}$$

  whereby $\hat{y}_i = f(x_i)$ denotes the value of the function calculated through regression at the point $x_i$, in contrast to the test value $y_i$ at this point. The RMS error is therefore a scalar variable, which characterizes the totality of all residuals and can be regarded as the mean deviation of the response variable readings from the regression function. A low value indicates that the regression function is very capable of describing the experimental data. The standard error of regression is an absolute variable, which cannot be interpreted directly, particularly if no earlier comparable tests exist. In that case, another important characteristic - the so-called coefficient of determination - can provide a remedy.
- The **coefficient of determination** $R^2$ can be interpreted as the proportion of the variance in response variables that is "explained" by the regression, see section 9.1.6 in the Appendix. This is due to the fact that relationships that are not present in the model (e.g. unknown predictor variables) cannot be explained by the model either. The coefficient of determination assumes values between 0.0 and +1.0. Thus, an $R^2$ of or close to 1.0 is a sign that the regression approach was suitable for explaining the experimental data.
- Alternatively, the **correlation coefficient** R can be used. R assumes values between -1.0 and +1.0, whereby both limit values only occur if the data can be completely explained by the regression approach. In a linear regression, for example, all treatments would have to be positioned on a straight line. If deviations from this occur, we have to assess how great these are and what causes they might have. For a correlation coefficient of or close to 0.0 does not automatically mean that high variation is present, but rather that the regression approach was not able to explain the data with sufficient accuracy. This may be due to high variation, or to the approach itself. Thus, the attempt to bring a parabola positioned symmetrically in the experimental area (i.e. a $2^{nd}$ order approach) closer to a straight line through linear regression is confirmed with an R of 0.0, even when no variation is present, see section 9.1.6 in the Appendix.

#### 6.3.2.2. Residual analysis

The residuals supply important information about the goodness of fit of the model equation in relation to the test results. They characterize the deviation between the results measured in the test and those predicted on the basis of the model equation. If the model approach is suitable for describing the test results, this deviation may only be *random*. This must be checked after every regression analysis, whereby a graphic representation is always advantageous, because it provides direct, incisive information about the relationships under discussion. Such representation can be generated automatically by good software, see section 9.2. The following evaluations are recommended:

- **Residuals in the normal distribution plot:** The purpose of this representation is to ensure that the residuals are normally distributed, in this case they will lie approximately on a straight line in the probability plot. If definite curves are observed, a transformation of variables may be necessary. Normal distribution can also be backed up by suitable statistical tests.
- **Residuals as a function of the test number:** The aim of this evaluation is to answer the question as to whether the test results feature a trend or jump. Here, too, no systematic relationship is required; the residuals must be randomly distributed within a band. If a trend does exist, the residuals may systematically increase or decrease, for example. This would indicate the existence of unrecorded noise variables (e.g. tool wear), which are influencing the result. What is important at this stage is understanding that a trend of this kind is only visible in the residuals if a random sequence (randomization) was used when conducting the experiment. Otherwise, the trend is ascribed to the dependence of the response variable on the factors, and is no longer visible in the residuals.

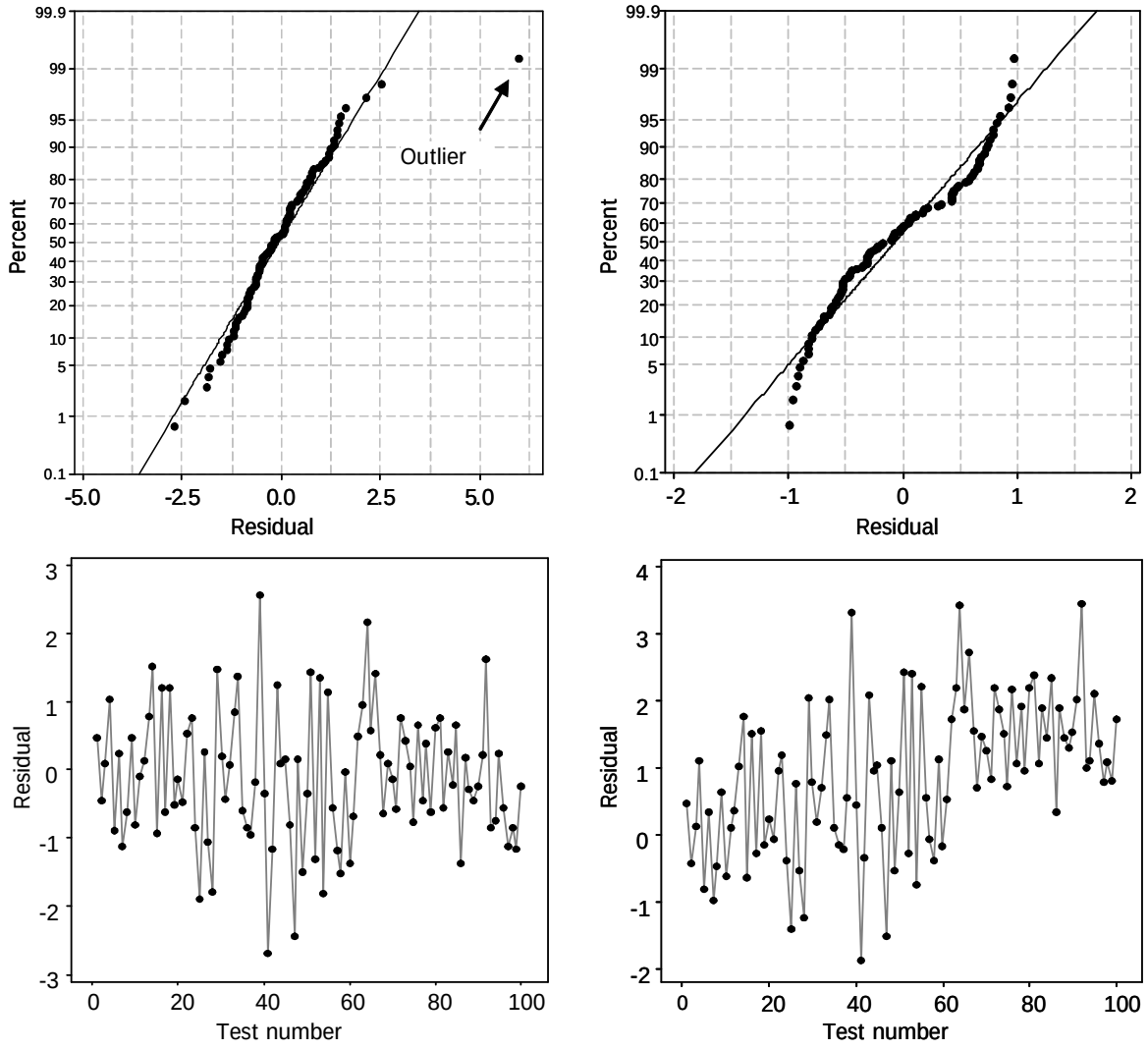



**Figure 6.4:** Top: Residuals in the normal distribution plot with normally distributed data with outlier (left), and with equally distributed data in zone [-1,+1] (right). Bottom: Residuals as a function of the test number without trend (left), and with trend (right).

- **Residuals as a function of factors:** This evaluation is used to examine whether the model approach used is consistent with the test results. This analysis can basically also be conducted by representing the response variable as a function of the factors; for example, a linear model approach can be validated by comparing it with test results at the center of the zone. However, the residuals are normally smaller than the response variable, so that representation of residuals as a function of the factors is more suited to this task. At the same time, care must be taken to ensure that no systematic relationship exists, but rather the residuals are randomly distributed in a band around the calculated response variable.
  The distribution of residuals on the $x$ axis, on the other hand, is irrelevant. In the case of systematically designed tests, the residuals will only be present on the factor levels used in the test; in the case of tests that have not been designed (or the evaluation of historic data), any factor values may be present.
- **Residuals as a function of the calculated response variable**: For many statistical analyses, it is a prerequisite that the standard deviation is not dependent upon the response variable. Compliance with this condition can be checked by monitoring whether the residuals are spread opposite the calculated response variable. If they are not lying at random in a band, but rather a trend is visible (often a funnel-shaped increase or decrease, for example), variable transformation (e.g. logarithmic) must be risked.

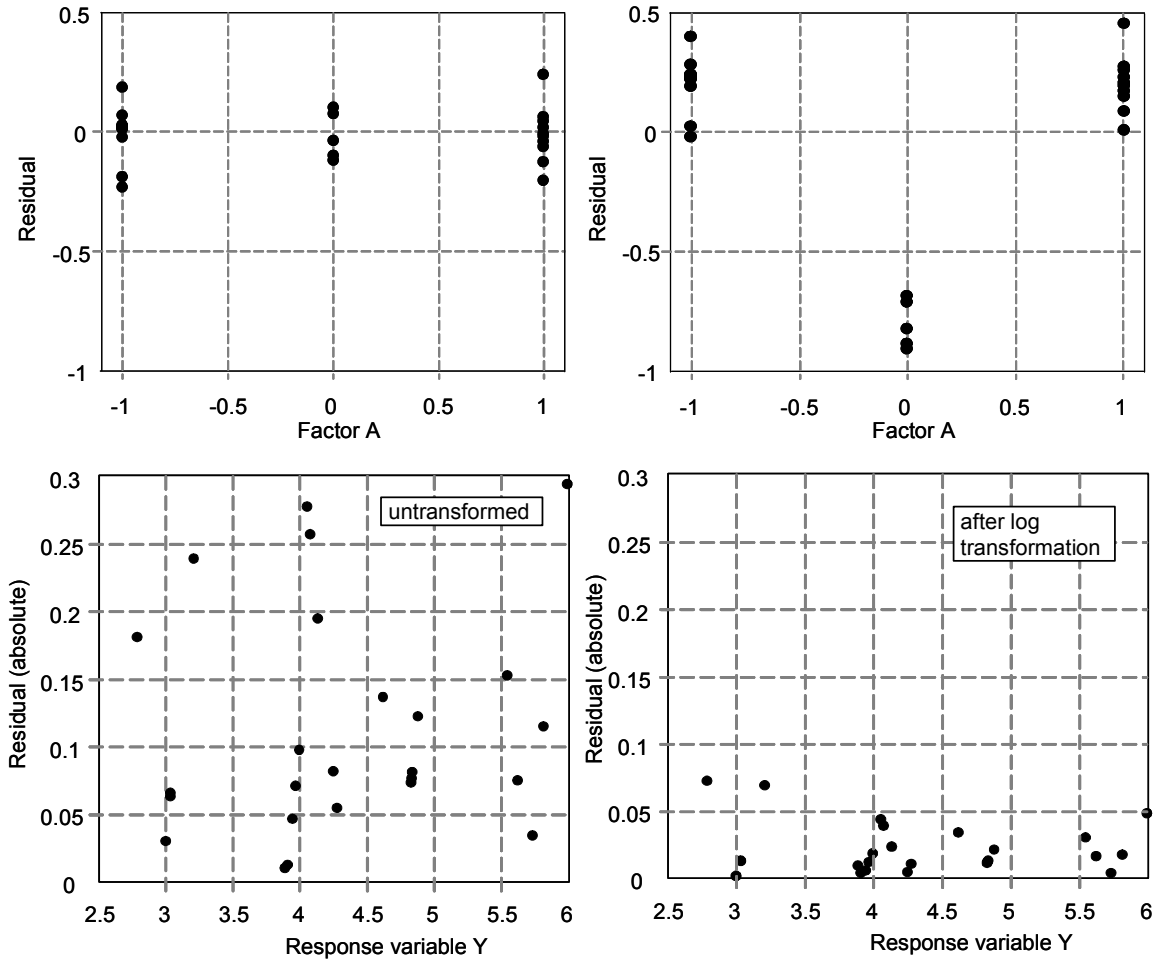



**Figure 6.5:** Top: Residuals as the function of a factor in a linear model. The data on the left confirm the linear profile. The data on the right indicate a non-linear profile. Bottom: Residuals (absolute value) as a function of an exponential response variable, without log transformation (left), with log transformation (right).

#### 6.3.2.3. Confidence interval and significance of regression coefficients

The coefficients of the regression model were determined on the basis of the experimental data, which were subject to random variation. It can thus be assumed that the coefficients themselves are random variables. It is possible to show that under certain conditions (see section 9.1.6 in the Appendix), the values determined using the least-squares method provide a good estimate of the true values.

With the standard error of the $i$th regression coefficient $S_i$, a two-sided confidence interval on the confidence level (1 - $\alpha$) can be stated using the technique of mean comparison. The mathematical foundations are not dealt with here, instead see section 9.1.6 in the Appendix. The results are normally represented together with the values of the coefficients, see Table 6.6.

The confidence intervals for the mean of the response variable $\mu_y(x)$ are dependent upon $x$ and are graphically represented as a “trumpet-shaped” enveloping of the regression curve. The significance of the calculated gradient can be evaluated in the same procedure as with the t-test, by calculating a test statistic and comparing it with values in the table.

| Variable | Coefficient | Std. error | t-value | p-value |
|---|---|---|---|---|
| $x_1$ | $b_1$ | $S_1$ | $t_1$ | $p_1$ |
| … | | | | |

**Table 6.6:** Coefficient table of a regression

Both the representation of the confidence interval of the mean and the definition of a confidence interval for the gradient illustrate that the calculated straight line is only an estimate of the true, but unknown, straight line. In a certain sense, the confidence intervals constitute the abundance of possible positions of the true straight line. During interpretation we must bear in mind that a significant gradient is not proof that the dependence is also linear, but simply that a linear proportion exists. This matter is illustrated further in diagram 6.6 below. Further details can be found in [1].



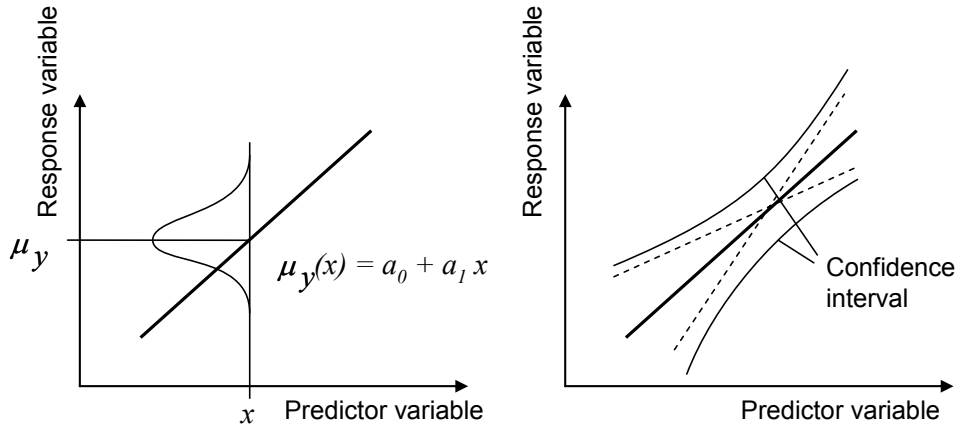



**Figure 6.6:** Confidence interval of the mean in a linear regression (schematic)

#### 6.3.2.4. Step-by-step regression

In step-by-step regression, the regression analysis is repeated iteratively in several steps, whereby with each step, effects and interactions that are not classed as significant are removed from the model, and the remaining degrees of freedom are employed for the purpose of improving statistical certainty. This process is repeated until only significant effects and interactions remain in the model, which then has the simplest possible form.

The following procedure is recommended, see [19], p.74:

- First, the significance of higher (e.g. quadratic) effects must be checked, if any are contained in the model. These may be removed first, if necessary.
- Higher order interactions are dealt with next.
- Finally, the significance of twofold interactions and effects is checked, taking into consideration the hierarchical structure of the model. Because of the latter, an effect must only be removed if all higher order terms and all its interactions have been removed first. This ensures the invariance of the model in the event that predictor variables are rescaled.

With each step of regression, the changes in the coefficient of determination $R^2$ must be tracked; $R^2$ must increase during the course of regression.

### 6.3.3. Validation through analysis of variance

In the preceding sections, we explained how the significance of effects can be evaluated using a mean value procedure based on the t-test for designs of $k$ factors on 2 levels. Where factors with several levels are concerned, statements in respect of significance could also be made on the basis of the t-test during a regression analysis, but only in the case of quantitative factors and quasi-linear dependencies. The analysis of variance now provides a tool for enabling conclusions about significance to be drawn in cases where the mean comparison technique is not equal to the task,

typically involving a qualitative factor with several levels. As this procedure is generally applicable and widely used in statistical software, we will deal with this subject briefly here. Further information about variance analyses can be found in the Appendix and in [27].

#### 6.3.3.1. Simple analysis of variance (ANOVA)

Simple in this context means that several levels of a single factor are under observation. This procedure can be employed for analyzing a single-factor test on several levels. As the first step in a multifactorial tests, we must check whether the variance of the results in the test lines differs significantly from the experimental variance, i.e. whether any factor is exerting an influence on the response variable.

Let's take a look at the design below with $n$ replications on $m$ levels:

| Level no. | Results $y_{ij}$ | Mean $\bar{y}_i$ | Variance $s_i^2$ |
|---|---|---|---|
| 1 | $y_{11}, y_{12}, \ldots, y_{1n}$ | $\bar{y}_1$ | $s_1^2$ |
| 2 | $y_{21}, y_{22}, \ldots, y_{2n}$ | $\bar{y}_2$ | $s_2^2$ |
| … | | | |
| m | $y_{m1}, y_{m2}, \ldots, y_{mn}$ | $\bar{y}_m$ | $s_m^2$ |

**Table 6.7:** Experimental design with $n$ replications on $m$ levels.

Simple analysis of variance can be performed in three stages:

1. Calculation of the mean variation of the individual values (this figure is a measure of experimental variance)

$$\overline{s_y^2} = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \ . \qquad (6.26)$$

2. Calculation of the variance of the means

$$s_{\bar{y}}^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{y}_i - \bar{\bar{y}}\right)^2 \quad \text{with} \quad \bar{\bar{y}} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i \quad \text{(overall mean).} \qquad (6.27)$$

3. F-test with the test statistic $F = \dfrac{n s_{\bar{y}}^2}{\overline{s_y^2}}$. (6.28)

   If $F$ is greater than the threshold value $F(f_1,f_2,99\%)$ of the F-distribution in the table in the Appendix with degrees of freedom $f_1=m-1$, $f_2=(n-1)m$, the test results are subject to a significant difference ($n$ = number of replications per level, $m$ = number of levels).

As with the mean comparison, statements about significance can be made dependent upon the interval in which the calculated test statistic is located:

| Lower limit | Test statistic | Upper limit | Rating |
|---|---|---|---|
| | $F \leq$ | $F(f_1, f_2, 99\%)$ | not significant |
| $F(f_1, f_2, 99\%)$ | $< F \leq$ | $F(f_1, f_2, 99.9\%)$ | significant |
| $F(f_1, f_2, 99.9\%)$ | $< F$ | | highly significant |

**Table 6.8:** Significance rating using the F-test

#### 6.3.3.2. Factorial analysis of variance

Factorial analysis of variance enables us to decide whether each factor of a multifactorial experimental design has a significant influence on the test result. The following procedure is recommended:

1. Calculation of the level means of the factor $\bar{y}_i$ , $i = 1...m$ . All test values with which the factor being evaluated has the level in question are subject to averaging. In this way, all other factors are included in the averaging process with values over all levels.
2. Calculation of experimental variance $\overline{s_y^2}$ . This is the mean variance of treatments with multiple realization.
3. Calculation of the variance $s_{\bar{y}}^2$ of the means of the factor calculated in item 1.
4. F-test with the test statistic $F = \frac{n s_{\bar{y}}^2}{\overline{s_y^2}}$ . (6.29)

   If the test statistic $F$ is greater than the threshold value $F(f_1,f_2,99\%)$ of the F-distribution in the table in the Appendix with degrees of freedom $f_1=m-1, f_2$, the factor has a significant influence on the test result, with $n$ = number of test values per level, $m$ = number of levels.

Statements about significance can be made dependent upon the interval in which the calculated test statistic is located, as shown in Table 6.8.



#### 6.3.3.3. Variance table, multiple analysis of variance

The analysis of variance can be comprehended as the partition of the sum of squares (SS) of the individual values from the overall mean in a proportion $Q_A$ and a remainder $Q_R$, whereby the first proportion is explained by the factor levels and the remaining proportion can be used to estimate random variation. An explanation of this partition can be found in section 9.1.5 of the Appendix, and is not discussed further at this point. The result of the partition is usually summarized in a table, which looks like this:

| | Degree of freedom | SS | Variance | F | p-value |
|---|---|---|---|---|---|
| Factor $A$ | $f_A = m-1 = f_1$ | $Q_A$ | $s_A^2 = \frac{Q_A}{f_A} = n \cdot s_{\bar{y}}^2$ | $F_A = \frac{s_A^2}{s_{\text{Rest}}^2}$ | $p_A$ |
| … | … | … | … | … | … |
| Remainder | $f_R = f_2$ | $Q_R$ | $s_{\text{Rest}}^2 = \frac{Q_R}{f_R} = \overline{s_y^2}$ | | |
| Total | $f = f_A + \ldots + f_R$ | $Q = Q_A + \ldots + Q_R$ | | | |

**Table 6.9:** MANOVA table ($n$ = number of test values per level, $m$ = number of levels)

Statements about significance can be made dependent upon the interval in which the calculated test statistic $F_A$ is located, as illustrated in Table 6.8. With some statistical tools, a variable $p_A$ is stated in addition to the test statistic $F_A$ for assessing significance, for which the following applies

$$F_A = F(1-p_A, f_1, f_2) . \quad (6.30)$$

This statistic is interpreted as the “probability of error” (type I error): the smaller this probability, the more significant is the factor’s influence. Assessment is the same as for the test statistic $F_A$: a significant result is present if $p_A \leq 1\%$, a highly significant result if $p_A \leq 0.1\%$.

As already mentioned in Table 6.9, this partition can also be carried out for several predictor variables and interactions; this is referred to as the multiple analysis of variance. Here, the mathematical procedure is considerably more time-consuming, but is easy to formalize, with the result that this analysis is included in the majority of statistical software packages.

### 6.4. Graphic representation

The calculated model equation and hence the results of the test evaluation are usually graphically presented for greater clarity. Response variables that are dependent upon two or more factors are mostly displayed in the form of level curves, characteristic maps and families of curves. The diagrams below illustrate this relationship.

The left-hand side of Figure 6.7 shows the 3D representation of a “mountain”. The right-hand side of Figure 6.7 shows the level curves of this “mountain” in a “view from above”, as you would see on a topographical map. In this example, the jump from one curve to a neighboring one corresponds to a defined difference in height of 20 units. Curves situated very close together signify a steep climb in one direction, perpendicular to the level curves. If we remain on one continuous level curve, we are moving - in terms of this map - around the mountain at a constant height.

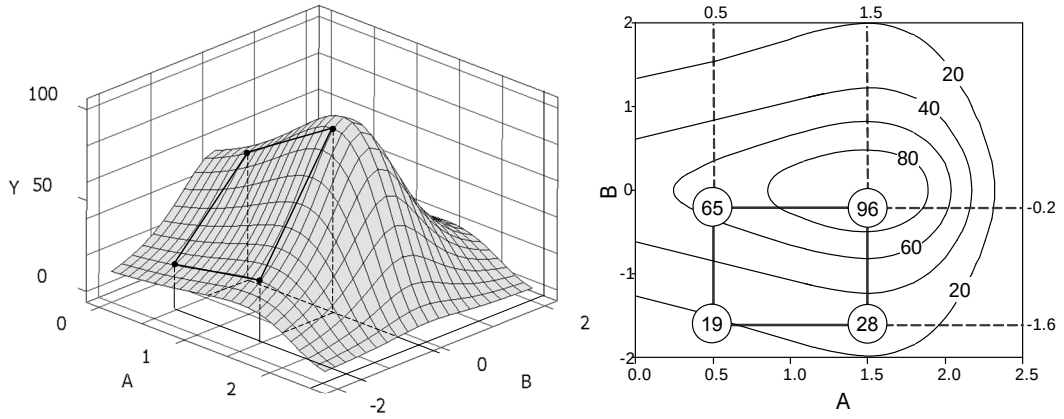



**Figure 6.7:** Function profile in a 3D (left) and 2D (right) presentation

Let’s leave this image of a “mountain” behind us for a while, and instead of its height observe a function $y$, which is dependent upon the variables $A$ and $B$: $y = f(A,B)$. This equation can act as a model for a response variable$y$, the value of which is determined by setting the factors $A$ and $B$. Then, each setting $(A, B)$ corresponds to a value $y$.

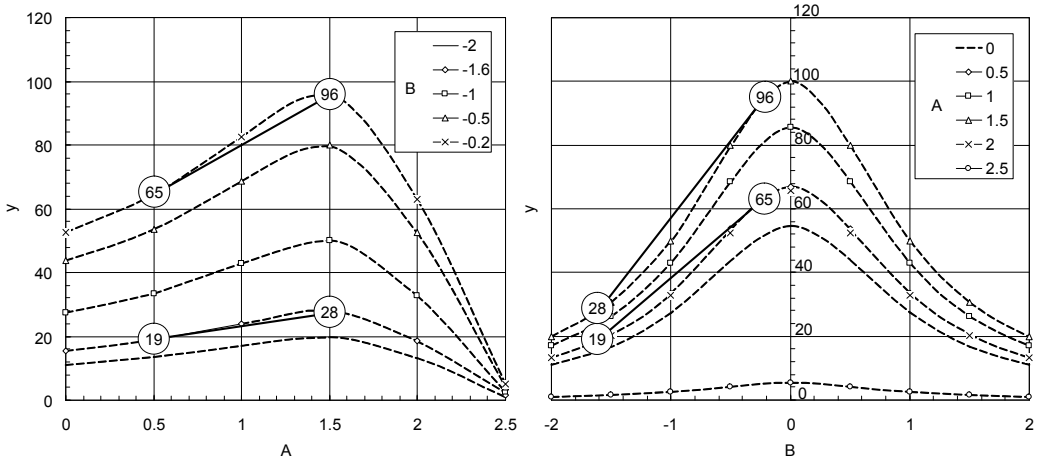



**Figure 6.8:** Parametric representation of the function profile shown in Figure 6.7

Figure 6.8 shows another possible representation for these results. The response variable $y$ is applied as a function of $A$ with $B$ as a fixed parameter, which is referred to as a “parametric representation”. The dotted curves in the left-hand graph of Figure 6.8 each represent the function $y$ with fixed $B$. They are, so to speak, the cutting lines of a vertical cut through the surface of the mountain with fixed $B$. In the same way, the dotted lines in the right-hand graph of Figure 6.8 each represent the function $y$ with fixed $A$.

These methods are basically often employed to illustrate test results. Typical representations are shown in Figure 6.9. We distinguish between the following three representations:

- The **Adjusted Response Graph** represents the mean effect of a factor. In other words, the effects of all other factors have been eliminated by averaging. If the represented factor displays significant interactions with other factors, this representation may be misleading, because the profile shown can depend to a great extent on how the other factors are currently set.
- The **Predicted Response Graph** is a parametric representation of the response variable as a function of a factor with predefined, fixed settings for all other factors.
- In the **Interaction Graph**, each factor is displayed at every level of all other factors. Here, significant interactions can be recognized in that the families of curves do not run in parallel. A non-parallel profile signifies that the effect of a factor depends significantly upon the setting of another factor, which is characterized as an interaction.

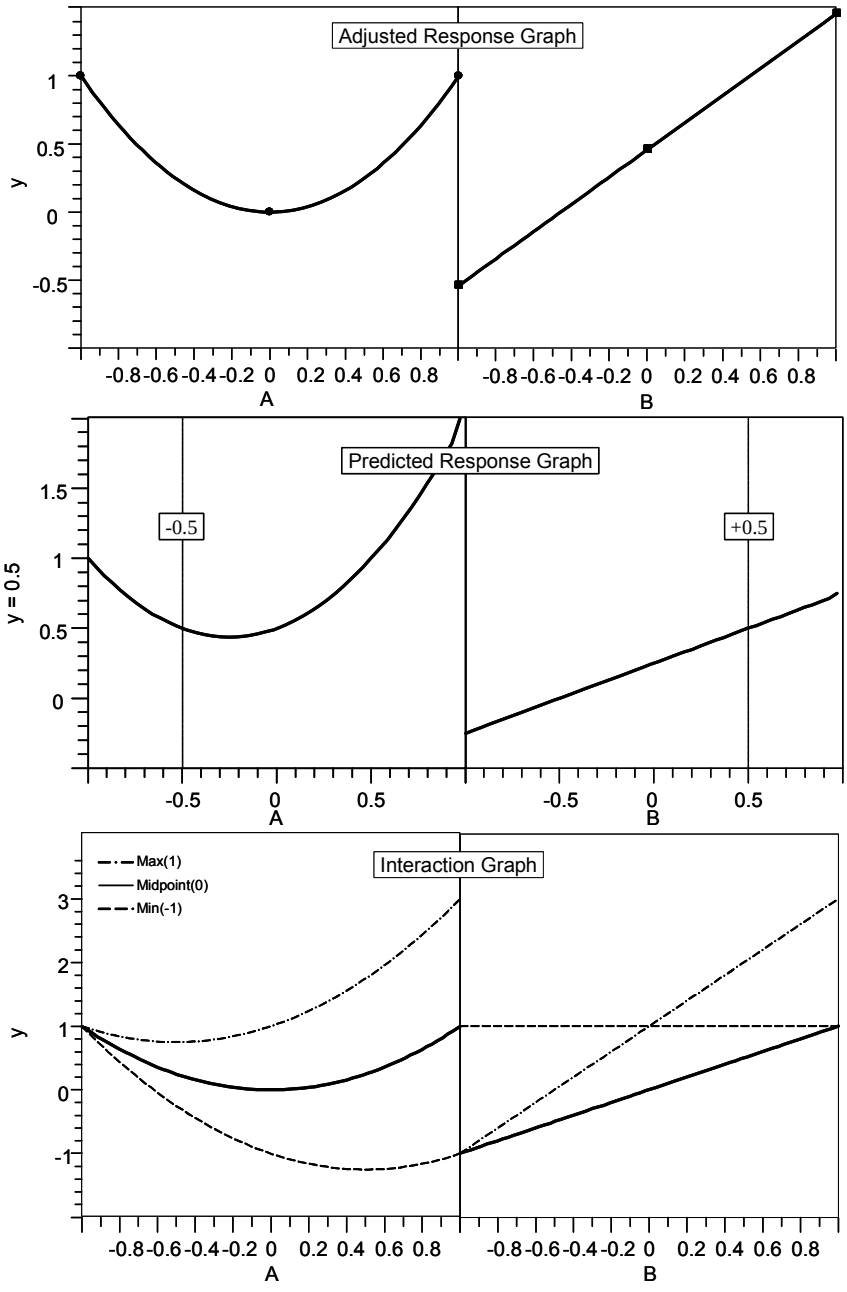



**Figure 6.9:** Typical representation of test results of the function $y = A^2 + B + AB$

For the communication with customers the predicted response graph is recommended. Here some effects can be discussed without going too deeply into specific DoE-details.

This type of representation can also be used when one (or more) of the factors under investigation is not a quantitative, adjustable variable but a qualitative variable with fixed levels (e.g. material 1 - material 2). In this case, of course, it would not make sense to interpolate the intermediate values.

Where three factors exist, the experimental area can be graphically represented in the form of a cube. In this cube, each corner corresponds to a combination of levels of the factors $A$, $B$ and $C$. When the factors number more than three, only two or three-dimensional projections of the multidimensional experimental area are possible.

### 6.5. Interpretation of results

Working together with experts, the results must be discussed, physically interpreted and plausibilized by means of a comparison with the expected result. Unexpected results always provide grounds for further research and a furthering of knowledge, with the aim, for example, of finding unknown factors or interactions. As a general rule, we should not trust any result that we do not understand.

First of all, we have to examine which factors have the greatest significant influence on the response variable. These are the factors with the greatest effects. If a change (e.g. optimization) of the response variable is desirable, precisely these factors must be used for this purpose. If, however, the response variable is to be kept as constant as possible or within narrow limits, particular effort must be made to keep precisely these factors constant or to limit their variation. However, it is important to note that the effects depend considerably upon the choice of factor ranges. Consequently, a small effect may be due to the selection of an overly small factor range.



When a linear approach is taken, larger effects are expressed by the steeper gradient of the effect straight lines. If the factors behave additively, the curves of a factor will run parallel at all other settings of the other factors. If, on the other hand, the effect of a factor depends upon the setting (level) of another, this signifies an *interaction* between these two factors, i.e. they are not behaving additively. The effect straight line of the factor is subject to different (and even contrary) gradients, depending on the setting of the other factors.

If significant interactions have been established, this means that all factors involved in the interaction must be examined together. The influence of the interaction may cause the effect of a factor to increase, diminish or even have the reverse effect. Therefore, we should know intuitively that neglecting the interaction can lead to completely false results or conclusions. Furthermore, when a strong interaction $AB$ is present, the mean effect of the two factors $A$ and $B$ can become zero, even though each factor can certainly exert a major influence. Higher order interactions may also occur in designs with more than 2 factors. Threefold interactions mean, for example, that a combination of 2 factors has an influence on a third factor, etc. Higher order interactions seldom have a major influence and are generally difficult to interpret. They may find expression, for instance, in the response curvature of factor $A$, changing with the settings of factor $B$. This behaviour will not be covered unless an interaction term $A^2B$ is included into the model.

*EXAMPLE: In this example, we will look at a design with 2 factors on two levels and results as follows:*

| *No.* | *A* | *B* | *AB* | $\bar{y}$ |
|---|---|---|---|---|
| *1* | - | - | + | $\bar{y}_1$ |
| *2* | + | - | - | $\bar{y}_2$ |
| *3* | - | + | - | $\bar{y}_3$ |
| *4* | + | + | + | $\bar{y}_4$ |

**Table 6.10:** Experimental design with 2 factors on two levels

*The effects of the factors can be calculated thus (also see section 6.2.3):*

$$\Delta_A = \frac{\sum \bar{y}(A_+)}{2^{k-1}} - \frac{\sum \bar{y}(A_-)}{2^{k-1}} = \frac{\bar{y}_2 + \bar{y}_4}{2} - \frac{\bar{y}_1 + \bar{y}_3}{2}$$

$$\Delta_B = \frac{\sum \bar{y}(B_+)}{2^{k-1}} - \frac{\sum \bar{y}(B_-)}{2^{k-1}} = \frac{\bar{y}_3 + \bar{y}_4}{2} - \frac{\bar{y}_1 + \bar{y}_2}{2} \qquad (6.31)$$

$$\Delta_{AB} = \frac{\sum \bar{y}(AB_+)}{2^{k-1}} - \frac{\sum \bar{y}(AB_-)}{2^{k-1}} = \frac{\bar{y}_1 + \bar{y}_4}{2} - \frac{\bar{y}_2 + \bar{y}_3}{2}.$$

*The graphic representation of the effects is shown in Figure 6.10 for the example of a two-factor experiment.*

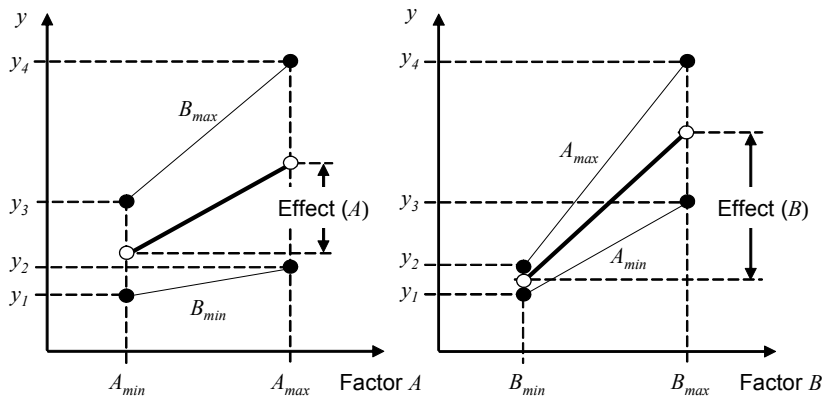



**Figure 6.10:** Graphic representation of the effects in this example

*In the above illustration, we can see (assuming the y-scale is the same in both graphs) that factor B has a greater effect on the response variable than factor A. Moreover, a clear interaction between the two factors is evident, as the gradient of the straight lines of a factor are obviously dependent upon the setting of the other factor in each case.*



## 6.6. Conclusions and further procedure

After the model has been calculated and validated and the results represented and interpreted, conclusions must be drawn from the tests. Whether or not the objective of the experiment was achieved can also be evaluated at this point.

If results are implausible or unsatisfactory, further tests may be conducted, e.g. experiments with different factors, factor levels, treatments or response variables. If necessary, an extended model may also be examined.

Last but not least, summary documentation must be compiled and knowledge transferred.

## 7. Applications

The aim of the evaluation described in Chapter 6 is to adapt the model approach to the test results. In this chapter, we take a brief look at the most important applications for the model, i.e. those that permit us to gain knowledge about the real system by means of simulation.

### 7.1. Prediction

Once a validated model is available, the simplest task is to calculate the value of the response variable when the factors have predefined values, in other words, to predict the behavior of the real system as a function of the predictor variables.

Here, please note that in the case of black-box models, such a prediction can only meaningfully be carried out as interpolation. Extrapolations can lead to entirely false statements, as no experimental information whatsoever is available about the behavior of the response variable outside the experimental area under investigation.



### 7.2. Optimization

Another important task a model has is to enable the response variable to be optimized. This task complements that of prediction: it involves finding those factor settings at which the response variables assume predefined values. These are frequently maxima or minima. If necessary, we may also define constraints that must not be violated. Here, the factors are usually referred to as design variables.

Various strategies can be used for optimization:

a) If treatments are generated according to a particular plan, deterministic methods can be used. If they are selected at random, stochastic methods are employed. Typical examples of deterministic methods are the simplex algorithm and the gradient-based approach. Stochastic methods are the genetic algorithm (GA), the evolutionary strategy (EV) and the artificial neural network (ANN).

b) Methods differ in that either only the function to be optimized is required for calculation, or this function and also its $1^{st}$ derivative, or even its $2^{nd}$ derivative as well. This is due to the fact that the calculation of derivatives can be time-consuming and even impossible. Typical examples of the first category are all stochastic methods and the simplex method. An initial derivative of the function in the form of a gradient, on the other hand, requires the classic gradient method. The quasi-Newton method even requires the second derivative of the function, although variants exist that manage without it.

c) Some methods are suitable for model-based optimization, others for test-based optimization, and some can be employed in both cases. Test-based methods require us to determine only the true response of the system with discrete settings of the predictor variable. This system response can be ascertained directly through tests. Model-based methods, on the other hand, require the calculation of a model equation first of all, i.e. the true response of the system ascertained in the test is approximated by means of an approach known as the response surface. Next, this approximation is optimized. These processes are frequently referred to as response surface methods. In the case of test-based optimization, all approaches, e.g. the simplex method, can be considered if they require function values only at discrete points of the experimental area. If, on the other hand, derivatives are needed as with gradient-based techniques, first of all we must produce a model equation, which can then be used to calculate the necessary gradients. Alternatively, however, gradients can also be approximated through differential quotients; then, only the function values would need to be determined.

d) So, we differentiate between response surface and adaptive response surface methods, depending on whether a global or local (possibly requiring iterative improvement) model equation is used for model-based optimization. In this context, global means that the model equation covers the entire experimental area. Local models, on the other hand, only cover certain areas of the experimental area.

e) Sometimes not just one but several target functions require optimization. This can be simply due to the fact that all target functions are to be weighted, superposed and combined in a new target function. A disadvantage of this technique is that arbitrary weighting factors are required, which clearly define the “compromise” between the individual response variables. Sometimes, however, it is necessary to find all combinations of predictor variables in the experimental area that are of “equal value”, i.e. no response variable can be improved without worsening another. From this quantity we can then use further criteria to establish a suitable optimum. Pareto optimality makes use of this approach.

We will now go on to explain some of the typical processes mentioned above in more detail. Reference is made to literature containing more in-depth information. After an optimum has been found using a suitable approach, it is generally also necessary to conduct a validation test, which confirms the results of optimization.

### 7.2.1. Gradient-based methods



The most commonly used procedure is model-based optimization, e.g. using a gradient method. Here, quadratic approaches must be used, at least, because the extremum of a linear approach always lies at the boundary of the experimental area and is not suitable for reasonably approximating a response surface with an extremum that is within the experimental area. With linear scenarios, much simpler approaches can be taken using so-called linear programming.

The classic **gradient method** is basically a method of non-linear optimization which, like most techniques, is based on a local search. A non-linear function can possess several local extrema, and the optimization process will identify one of these. However, this local extremum does not necessarily also constitute the global optimum. This, too, is one of the greatest disadvantages of gradient-based methods. A way out of this situation is offered by repeating the process several times with different starting values.

The gradient method is based on the fact that the gradient of a scalar function is a vector, which dictates the direction of the steepest ascent in the current position, and the length of which is a measure of the gradient at this point. This direction, which is always perpendicular to the level curves of the function, must be taken if we wish to get to the function maximum as rapidly as possible. (The considerations herein also apply to the search for a minimum - here, the same method can be employed.) Since the gradient is a local variable and varies from one location to another, an iterative procedure is applied, whereby the gradient is calculated in the current position and the direction of the ascent thus ascertained. Next, a step is taken in this direction, after which the gradient is determined once again in this new position, and the procedure is repeated. Several approaches exist for determining the increment; one effective method is to choose an increment in inverse proportion to the amount of the gradient, so that the smaller the increment, the steeper the ascent. The search is over when, within the confines of the predefined level of accuracy, no neighboring location is found that is higher than the momentary position. At this point, the maximum has been found.

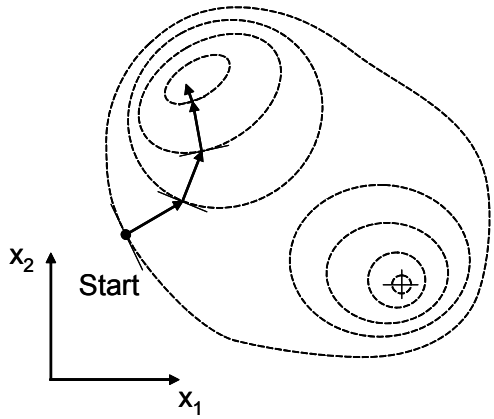



**Figure 7.1:** Gradient method

This method is one of the mostly widely used optimization algorithms, for it very quickly converges against a local extremum. A disadvantage of this method is that difficulties with convergence arise as soon as the response surface is affected by noise on the occurrence of confounders. Another limitation is that only smooth functions with a well defined gradient, that is, only continuous (quantitative) design variables can be treated in this way. This method is therefore typically used in the model-based optimization of response variables, which are determined through computer experiments, often using an FEM program as a solver.

A similar technique, which is also suitable for experimental optimization, is the **steepest ascent method**. The basic principle here is to approximately determine the gradient of the function from the difference quotients of the response variable in respect of the design variables, so that no global response surface is required. The procedure consists in applying, typically, a full factorial design on two levels, and approximately determining the direction of the steepest ascent (or descent) on this basis. If necessary, fractional factorial designs can also be used when more than 5 factors are present. In this direction, runs are now performed at certain increments until the moment of descent is reached. A new design is set up at this point and the procedure repeated until a local maximum (or minimum) has been found, within the constraints of the required accuracy. Experimental designs for non-linear relationships must be employed in the vicinity of the optimum, in order that its position can be determined with sufficient accuracy.



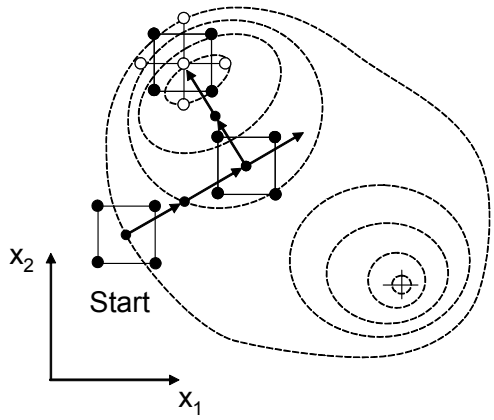



**Figure 7.2:** Steepest ascent method

The **Evolutionary Operations (EVOP)** method is a variant of the steepest ascent method that is adapted to suit the requirements of production. This method is employed during ongoing production, so the factor ranges are therefore smaller and the overall procedure more cautious. The idea at the heart of EVOP is that production and optimization can take place simultaneously within the framework of the continuous improvement of production. This method is explained in greater detail in [1].

### 7.2.2. Simplex method

The basic principle of the simplex method is to determine the direction for an improvement in the current position directly on the basis of the experiment, without calculating the gradient and with as few treatments as possible. It proposes the use of a structure for $k$ factors - the so-called simplex, from $k+1$ combinations. Where there are 2 factors, this is a triangle.

The procedure consists of the following steps, see [2], p.308:

- First, the test results are determined in the corners of simplex 1 in the current operating point.
- The point with the poorest result is reflected on the hyperplane (straight line in the case of 2 factors) determined by the remaining points, producing a new treatment. Together with the remaining points from the old simplex 1, this forms a new simplex 2, etc.
- The result is thus improved, step by step. The process is repeated until the desired optimum has been found.

The **advantages** of this simple method are that no time-consuming evaluation is required (and no determination of the gradient, in particular), and that each further operating point is better than the current one, with the result that an improvement can always be expected, regardless of how far we are from the actual optimum. These advantages mean that this technique is especially popular in the field of experimental optimization.

On the other hand, certain **disadvantages** also exist:

- The method only functions as described if the effects are considerably greater than the variation. If response surfaces suffer from noise, one remedial measure is to conduct several replications; then, however, the method quickly becomes time-consuming, as the experiments cannot be used more than once.
- The method converges poorly in the vicinity of the optimum, which is not what one would initially expect for an optimization process. However, an analysis shows that here, behavior is non-linear, so that the simplex cannot provide any major improvement to the response variable.
- The simplex algorithm cannot be used for optimization with several response variables.

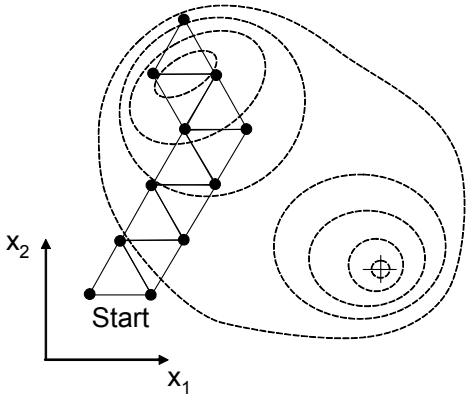



**Figure 7.3:** Simplex method



### 7.2.3. Evolutionary algorithms

This stochastic optimization technique was developed for those cases in which gradient-based optimization methods suffer to a greater or lesser extent from convergence difficulties. Consequently, it is able to handle both response surfaces suffering a large amount of noise, and qualitative and quantitative design variables. In addition, the method in question employs global search techniques, which are actually capable of determining the global optima among numerous local extrema. No calculation of the response surface is required; this is ascertained solely at discrete points through individual runs.

These approaches are based to a great extent on the natural principle of “survival of the fittest”, for nature must also counter the problem of maximum variance without time-consuming calculations. The price paid is the convergence speed, i.e. these methods generally require considerably more time to find an extremum.

Details about putting this method into practice can be found in [20].

## 7.3. Robustness and reliability analyses

BES-PE Glossary (issue 3-2009/11): Robustness is the ability of a system to preserve its function even under changing boundary conditions.

Typically, these boundary conditions are randomly varying geometric variables such as shape and position tolerances, material parameters such as the E modulus, ambient conditions such as temperature and, last but not least, varying loads. The effect of this parameter variance on design-relevant response variables is generally assessed by means of a robustness analysis. This is an integral part of the design process in the concept and drafting phase.

The central question posed by a robustness analysis is the extent to which variance in predictor variables can influence the response variable. A process is deemed robust if its response variables depend as little as possible on unavoidable fluctuations in predictor variables and any noise variables. Our understanding of robust products is the same, i.e. their characteristics should depend as little as possible on production and operating conditions. Here, different parameters can clearly influence robustness, so that certain settings can be advantageous for the variance in response variables, others are disadvantageous. Finally, robust optimization is a matter of finding favorable values for those

influence parameters that have a major influence on variance, without changing the mean of the process.

The fundamental relationships behind this approach are illustrated in Figure 7.4. The left-hand diagram portrays a non-linear dependence of the response variable on a factor $A$. Higher factor values visible reduce the variance of the response variable and are therefore advantageous for robustness. However, we can see that the response variable mean also moves considerably as $A$ rises, which may run counter to the requirements of the design. The right-hand section of Figure 7.4 shows the response variable as dependent upon two factors $A$ and $B$, indicating a clear interaction. By favorably setting factor $B$ on level $B(-)$, the response variable can now be rendered much more resilient to the influence of factor $A$ than with the setting $B(+)$, without changing the response mean. Interactions therefore play a very important role in robust optimization.

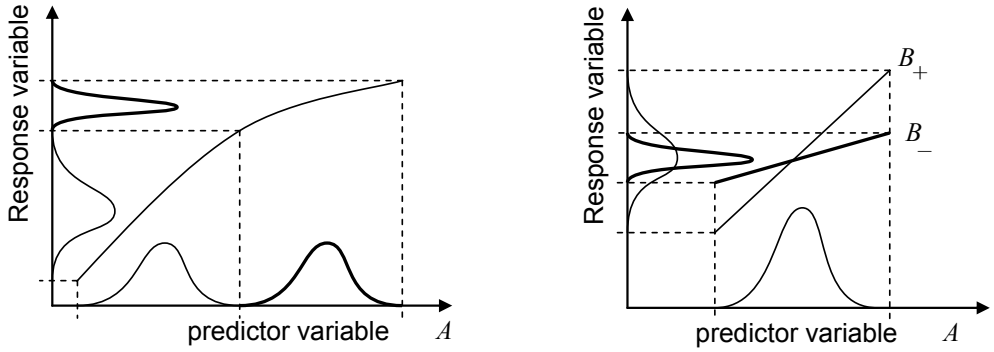



**Figure 7.4:** Robust processes (thick line) compared with insufficiently robust processes (thin line)

A robustness analysis typically takes place in three stages:

- First, we determine those parameters that exert a particularly high influence on design-relevant response variables and must therefore be kept within narrow tolerances in order to limit variance in response variables. These parameters can either be deduced from known cause-effect relationships between predictor and response variables, or determined through sensitivity tests (see chapter 8).
- At the heart of this analysis is the quantitative assessment of the design's robustness, achieved by calculating the distribution of the response variable as a function of the known distributions of relevant predictor variables. The proportion of permitted states among all possible states is commonly employed as the measure of robustness. This figure can be deduced from the distribution of the response variable within the tolerances stipulated by the requirements. A design is designated as robust, e.g. as shown in [20], if the interval of the response variable $\mu_y \pm 2\sigma_y$ lies inside the required tolerance, see [20], p. 236.
- Next, it may be necessary to optimize the robustness of the design, until the set requirements are satisfied.

The procedures for analyzing robustness and reliability are similar in that both methods evaluate the proportion of permitted states from all possible states. A difference is encountered during calculation, for in the robustness analysis, relatively frequent events occur when implicitly defined dependencies exist between response and predictor variables. These events can be dealt with using classic stochastic sampling methods (e.g. Monte Carlo or Latin hypercube, see chapter 8). In the reliability analysis, on the other hand, events are very rarely observed. These require the use of special techniques and are not discussed any further at this point. Further details can be found in [22], Chapters 7 and 8.

# 8. Advanced approaches

## 8.1. Fractional factorial designs

Fractional factorial designs are primarily used to determine those factors with the greatest effect on the response variable from a quantity of potentially relevant factors, in situations in which relatively little is known about the cause-effect relationships. We also refer to “screening” tests in this connection. Secondly, designs of this kind are also suitable for slightly reducing test expenditure in the case of tests with several factors.

### 8.1.1. Motivation

First of all, let’s take a look at problems with 3, 5 and 8 factors each on 2 levels, and analyze how many of the terms of a complete multilinear approach describe the primary effects as well as the two-fold, threefold and higher order interactions (IA). The terms also include a free constant, which is not explicitly detailed.



| Factors | No. of terms | | | |
|---|---|---|---|---|
| | Total | Primary effects | Twofold IA | Threefold and greater IA |
| 3 | 8 | 3 | 3 | 1 |
| 5 | 32 | 5 | 10 | 16 |
| 8 | 256 | 8 | 28 | 219 |

**Table 8.1:** Number of effects and interactions as a function of the number of factors

As can be seen in Table 8.1, the proportion of higher order interactions increases sharply as the number of factors rises. In a design with 8 factors on 2 levels with $m=2^8=256$ terms, the majority of these - 219 - are higher order interactions, of which it can be presumed that only a fraction will actually play a part as effects of a higher order. It is clear that just as many tests of a full factorial design would be required to determine these, with the result that as the number of factors increases, the test expenditure is increasingly used to determine higher order interactions.

The idea behind fractional factorial designs is to investigate further factors instead of some of these higher order interactions, so that the number of tests is not augmented.

### 8.1.2. Effects and confounding

We will now take a look at an example with 2 factors on 2 levels. The multilinear approach incorporates a total of 4 terms, including a twofold IA:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \,. \tag{8.1}$$

4 tests are therefore necessary in order to clearly determine the unknown coefficients. These can be conducted in a full factorial design. The idea of a fractional factorial design is now to examine a third factor instead of the twofold IA, using the following approach:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \,. \tag{8.2}$$

Above all, only 4 tests should be employed to determine the 4 unknown coefficients of the approach. This fractional factorial design is usually referred to as $2^{3-1}$, and signifies that 3 factors are to be investigated using the experiments of a double factorial design.

However, the complete multilinear approach with three factors looks like this:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{23} x_2 x_3 + a_{123} x_1 x_2 x_3 \,. \tag{8.3}$$

This contains 8 coefficients, the determination of which requires 8 tests. The example of the fractional factorial design $2^{3-1}$ illustrates the consequences of using fewer than 8 tests.

The design matrix of the full factorial design $2^3$ is displayed below. Since only 4 tests are available for our fractional factorial design $2^{3-1}$, only the first 4 tests should be used for this purpose.

| | A | B | C | AB | AC | BC | ABC |
|---|---|---|---|---|---|---|---|
| 1 | - | - | + | + | - | - | + |
| 2 | + | - | - | - | - | + | + |
| 3 | - | + | - | - | + | - | + |
| 4 | + | + | + | + | + | + | + |
| 5 | - | - | - | + | + | + | - |
| 6 | + | - | + | - | + | - | - |
| 7 | - | + | + | - | - | + | - |
| 8 | + | + | - | + | - | - | - |

**Table 8.2:** Design matrix of a full factorial design 23

The columns of the associated interactions $AB$, $AC$, $BC$ and $ABC$ can be calculated as a product of the corresponding columns for $A$, $B$ and $C$. If we compare these columns, we can see that $AB$ matches $C$, $AC$ matches $B$ and $BC$ matches $A$. In the full factorial design, the difference would only become clear if we incorporated the results from tests 5 to 8, which are not performed here. Consequently, columns $A$ and $BC$, $B$ and $AC$, $C$ and $AB$ of the evaluation matrix are indistinguishable. We would say that factor $A$ is *confounded* or *aliased* with interaction $BC$; the same applies to factor $B$ and interaction $AC$, as well as to factor $C$ and interaction $AB$. Furthermore, another column must be added to the evaluation matrix to clarify the confounding between the polynomial and the threefold interaction $ABC$, but this is not dealt with any further here for the sake of simplicity.

It is therefore not possible to calculate the effect of factor $A$ separately from the effect of interaction $BC$, for example, which constitutes a considerable disadvantage. An inadequate number of tests therefore leads to confounding, with the risk of misinterpretation. In particular, contradictory effects and/or IA can be eliminated, spurious effects feigned or actually present primary effects compensated. If higher level considerations (e.g. based on physics) allow us to decide which of the confounded effects determine the result, we will arrive at the desired result while economizing on the number of tests. If these deliberations are incorrect or if we ignore the alias, the result will be false!

The risk of misinterpretation greatly depends upon which effects are being confounded. If twofold IAs are confounded with the effects of the factors, this must be regarded as critical, because effects and twofold interactions play a major role in most applications. Designs of this kind have resolution III. Mutual confounding between twofold IAs in designs with resolution IV are less critical, and the aliasing of twofold IAs with threefold IAs in designs with resolution V are mostly uncritical. Table 8.3 provides a summary of this.

The use of Plackett-Burman screening designs is frequently recommended in the literature, as it enables important factors to be screened from a large number of possibilities. These designs are special, fractional factorial designs with resolution III and IV, in which the disadvantages mentioned above have partially been eradicated. In contrast to factorial designs, here the number of tests (number of lines) is an integral multiple of four, that is $N = 4,8,12,16,20,24,\cdots$. These screening designs are also structured according to different rules from factorial designs.

Nevertheless, Plackett-Burman designs are highly confounded arrangements, that incur the problems already discussed above. We would like to expressly point out that due to this aliasing structure, Plackett-Burman designs can only be used successfully if no interactions are present. Otherwise, they can lead to completely incorrect conclusions.

| Factors | No. of tests | | Confounding occurring with fractional factorial design | Resolution | Rating |
|---|---|---|---|---|---|
| | Full factorial design | Fractional factorial design | | | |
| 3 | 8 | $4 = 2^{3-1}$ | Factor with 2fIA | III | critical |
| 5 | 32 | $8 = 2^{5-2}$ | | | |
| 4 | 16 | $8 = 2^{4-1}$ | Factor with 3fIA<br>2fIA with 2fIA | IV | less critical |
| 6 | 64 | $16 = 2^{6-2}$ | | | |
| 5 | 32 | $16 = 2^{5-1}$ | Factor with 4fIA | V | largely uncritical |
| 8 | 256 | $64 = 2^{8-2}$ | 2fIA with 3fIA | | |
| 6 | 64 | $32 = 2^{6-1}$ | Factor with 5fIA<br>2fIA with 4fIA<br>3fIA with 3fIA | VI | uncritical |
| 9 | 512 | $128 = 2^{9-2}$ | | | |

**Table 8.3:** Resolution of fractional factorial designs



### 8.1.3. Recommended procedure

The following general recommendations apply to screening tests (see [1], p.145):

- It suffices to test each factor on 2 levels.
- Resolution III designs should not be used unless the alias pattern can be dispelled on the basis of physical considerations.
- Resolution IV designs offer the best compromise between expenditure and confounding. They have a clear alias pattern with no confounding of factors or twofold interactions, and are therefore especially suitable for screening tests.
- Designs with resolution V and higher can be employed to determine cause-effect relationships, as they are subject neither to confounding of factors and twofold interactions nor mutual confounding between twofold interactions.

If no more than 32 treatments can be examined (which is thoroughly realistic when we consider that these may also have to be conducted several times), the following is possible:

- Examinations of cause-effect relationships are possible with 6 parameters and 32 tests (or 5 parameters with 16 tests),
- Screening tests, on the other hand, can be conducted with up to 12 parameters and 32 tests (or 8 parameters with 16 tests).

## 8.2. Designs for non-linear relationships

In the preceding paragraphs we showed that factorial type $2^k$ designs always require a linear model. In this case, extreme values (maxima or minima) are situated at the boundary of the experimental area. The correctness of the model can only be checked by additional tests. If, for example, a local maximum lies inside the experimental area, proof of its existence can only be obtained by at least one test within this area.

Generally speaking, non-linearities can only be recognized when more than two levels are selected for each factor. In practical use, since the number of tests grows exponentially, the upper limit is mostly three levels, which means that quadratic approaches should be of the greatest interest to users. In this case, the complete model approach is:

$$y = a_0 + \sum_{i=1}^{k} a_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_{ij} x_i x_j + \sum_{i=1}^{k} a_{ii} x_i^2 \ . \qquad (8.4)$$

Curved surfaces can also be represented with this polynomial of the second degree.

One of the most common designs for non-linear relationships is presented in the sections below. These designs should only be used with few significant factors. If significance is unclear, screening tests must first be carried out using linear methods.

### 8.2.1. Type $3^k$ designs

The most obvious strategy for determining the quadratic model approach is to employ 3 levels per factor. Then, 3 factors automatically give rise to the $3^3$ design, whereby the three factor levels are usually referred to as -1, 0, +1 in the transformed coordinate system.

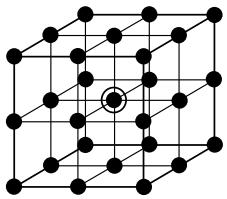


**Figure 8.1:** Full factorial $3^3$ design

The treatments of the $3^3$ design can be represented schematically by a cube in a three-dimensional experimental area in which, in addition to the corners of the cube of the $2^3$ design, the centers of the faces, the centers of the edges and the midpoint are to be examined.

The evaluation matrix of the $3^3$ design contains the fractions -2/3 and +1/3, as well as the integers -1, 0 and +1. Thus, the coefficients of the regression polynomial are not all calculated in the same way, see [2], p.209. In any case, a computer program must be used for evaluating a design of this kind.

An experimental design for three-level factors is mostly feasible if a test can be limited to few factors and an individual experiment is possible with comparatively low expenditure. In practice, it is conceivable that improved factor settings should be sought, for example, in the “vicinity” of the midpoint, which could represent serial production status.

Occasionally, one or more of the factors being examined are discrete predictor variables, e.g. material or machine. If so, the application of a regression polynomial and the calculation of theoretical intermediate values is completely meaningless. In this case, it is sufficient to limit oneself to evaluation based on variance analysis, and to ascertain the factor settings with the best test result according to the principle of “pick the winner”.

### 8.2.2. Central composite designs

A major disadvantage of full factorial designs on 3 levels is that the number of treatments very quickly grows with the number of factors. Moreover, there is high redundancy in testing, for considerably more experiments are conducted than there are unknown coefficients in the model equation, see Table 8.4.

| No. of factors | No. of coefficients in the quadratic model | No. of tests | | Redundancy factor | |
|---|---|---|---|---|---|
| | | $3^k$ | $2^k+2k+1$ | $3^k$ | $2^k+2k+1$ |
| 2 | 6 | 9 | 9 | 1.5 | 1.5 |
| 3 | 10 | 27 | 15 | 2.7 | 1.5 |
| 4 | 15 | 81 | 25 | 5.4 | 1.7 |

**Table 8.4:** Number of tests with central composite designs

A suitable alternative is offered by so-called type $2^k+2k+1$ central composite designs. The illustration below shows the structure of a central composite experimental design for three factors. In addition to the tests at the “points of the cube”, as prescribed by a complete $2^3$ design, experiments are conducted at the so-called star points and the center for $k$ factors of the star, the midpoint. This kind of design is known as central composite if the midpoint of the star coincides with the midpoint of the cube, as illustrated in Figure 8.2.

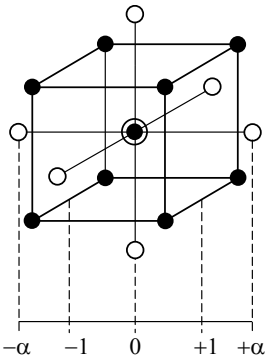



**Figure 8.2:** Central composite design with 3 factors

As we can see, each of the three factors is varied on five levels. Selecting the levels -2, -1, 0, +1, +2 for each level would be conceivable. However, the design would then lose its orthogonality, i.e. the coefficients of the regression polynomial could no longer be determined uncorrelated and with the same variance (see [2], p. 228). In order to retain this property, we must select the factor levels -$\alpha$, -1, 0, +1, +$\alpha$, whereby $\alpha$ is dependent upon the number of experiments conducted at the points of the cube, the star points and the midpoint:

$$\alpha_{orth}{}^2 = 0.5 \cdot \sqrt{N \cdot 2^k} - 2^k \,, \tag{8.5}$$

here, $N$ signifies the overall number of tests ($2^k$+2k+$n_0$) in this case, $k$ the number of factors and $n_0$ the number of replications of the midpoint.

Another important feature of an experimental design is its *rotatability*. A design is considered rotatable if the width of the calculated confidence interval depends only upon the distance from the center of the cube, not on the direction, see [1], p.191. The existence of this property can be shown if

$$\alpha_{dreh}{}^2 = \sqrt{2^k} \,. \tag{8.6}$$

It is possible to render both properties achievable by adapting the number of replications in the midpoint $n_0$.

Sometimes, a design can only be performed on three levels for technical reasons; in this case, $\alpha$=1 can be used. This must remain the exception to the rule, however, as a “face-centered” design of this kind is not orthogonal.

Where more than 3 factors exist, type $2^{k-p}$ fractional factorial designs of at least resolution V can also be employed as the basis for the cube. In this case, the conditions for orthogonality and rotatability must be adapted (the $2^k$ terms must be changed to $2^{k-p}$).

In practice, a central composite design is the preferred choice if an initial type $2^k$ design is to be reinforced by additional tests. In this case, only 2k+$n_0$ further treatments must be investigated.

### 8.2.3. D–optimal designs

D-optimal designs offer some major advantages:

- They provide a great deal of freedom as far as adapting the design to a predefined number of tests (e.g. for cost reasons) is concerned.
- They can be adapted to a predefined model and are therefore capable of registering non-linear behavior.
- Any number of levels that must be present in the design can be defined in advance. This permits the use of historic data.
- Areas of the experimental area that are not of interest can be excluded.
- Multi-level, quantitative and qualitative factors can be combined.

These designs have a certain disadvantage in that they are not orthogonal, although the difference is not very large in most cases. In addition, the procedure can only be carried out with software; however, most specialist DoE software tools now also offer D-optimal designs.

To determine the designs, those runs that enable the coefficients in the regression model to be ascertained as accurately as possible (with the smallest possible confidence intervals) are iteratively selected. A detailed description of this procedure would go beyond the scope of this volume, and is therefore dispensed with. Further details can be found in [4], for example.

D-optimal designs provide an interesting alternative to central composite designs. Their use is recommended where central composite designs cannot be employed. Typical applications are qualitative (categorical) factors with more than 2 levels, limitations in the experimental area, freely selectable block sizes, a different number of variants from the preset value, level values, etc.

### 8.2.4. Recommended procedure

Designs for non-linear relationships should only be employed where there are few significant factors, in order that the cause-effect relationships with the response variable can be established in detail and optimization achieved if necessary. For most applications, the number of significant factors should lie between 3 and 5. If these are not known, initial screening tests with linear approaches must be conducted.

The following recommendations should be noted, also see [1], p. 208:

- As a rule, a central composite design is the most suitable. With up to 4 factors, a full factorial design for the cube can be employed; for 5 factors, a resolution V fractional factorial design will suffice.
- As far as possible, the level values of the design must be set such that an orthogonal design is generated. Rotatability of the design must also be achieved by selecting a suitable number of replications in the midpoint.
- If, in exceptional cases, only 3 levels are feasible for technical reasons, a face-centered design with $\alpha$=1 can be used.
- If the use of central composite designs is contraindicated by categorical factors with more than 2 levels, impossible combinations in the experimental area, level values that deviate from the presetting, or similar, D-optimal designs must be employed.

## 8.3. Alternative model approaches

If physical modeling was not possible, section 4.4 mostly assumed a linear or quadratic polynomial a priori as a general model approach. This is due to the fact that every continuous function in the environment of a point can be approximated by a power series with a higher order error. This process is also referred to as Taylor series expansion.

However, this approximation is only valid in the environment of the expansion point, that is *locally*. Problems can arise if different behavior occurs in sections of the experimental area, such that the *global* relationships cannot be registered by simple polynomials of the first or second degree. Such difficulties are commonly overcome by incorporating higher order terms in the model. The example below illustrates that this strategy is not always successful.

*EXAMPLE 1: We wish to approximate the following function once in the interval $I_1$ = (-0.6; +0.6) and once in the interval $I_2$ = (-3.0; +3.0) by polynomials of different degrees. The function is shown in the table in 7 evenly distributed points of the interval in question:*

$$f(x) = \frac{1}{1+x^2}$$

| $x$ | $y=f(x)$ |
|---|---|
| -0,6 | 0,735 |
| -0,4 | 0,862 |
| -0,2 | 0,962 |
| 0,0 | 1,000 |
| 0,2 | 0,962 |
| 0,4 | 0,862 |
| 0,6 | 0,735 |

| $x$ | $y=f(x)$ |
|---|---|
| -3,0 | 0,1 |
| -2,0 | 0,2 |
| -1,0 | 0,5 |
| 0,0 | 1,0 |
| 1,0 | 0,5 |
| 2,0 | 0,2 |
| 3,0 | 0,1 |

**Table 8.5:** Data points of the example

*The calculated compensating curves are shown in Figure 8.3. As we can see, a quadratic polynomial is completely adequate for a local approximation in the interval $I_1$. However, this is not the case for a global view in the interval $I_2$; furthermore, increasing the polynomial approach from degree 1 (linear) to degree 6 only partially improves the quality of interpolation. Higher polynomials, in particular, tend towards "oscillation" at the interval limits, and are hardly suitable for reasonably reflecting the global behavior of the function. This phenomenon is also known as "overfitting".*

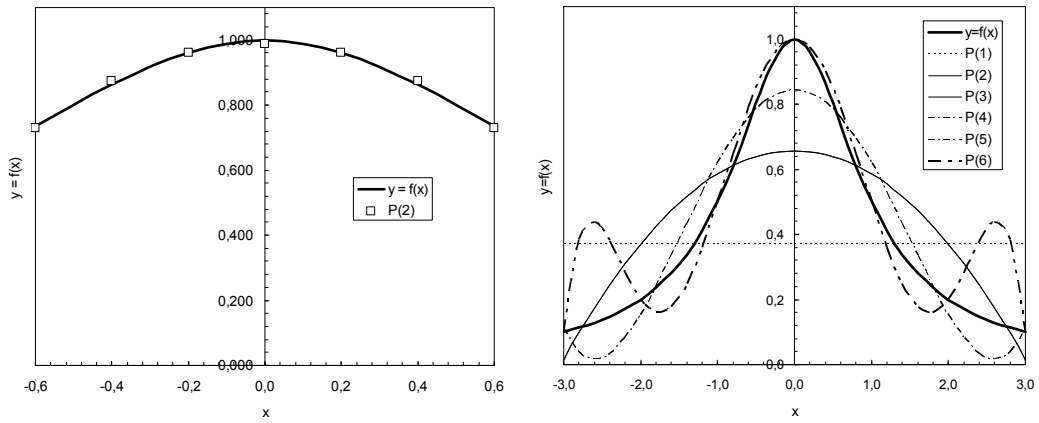



**Figure 8.3:** Calculated compensating curves

The problem obviously lies in the polynomial's fundamental inability to approximate a complex global function profile. This situation has led to the development of alternative model approaches, some of which are briefly summarized below. Further information can be found in [26,27].

- **Trigonometric functions**: Sine and cosine functions can be used as basis functions. This is similar to an approach using Fourier series development of the function. Here, as with traditional procedures, the approach is global in the sense that each of the basis functions makes a contribution in the overall experimental area.
- **Splines**: A spline of the order $k$ is a smooth curve, which is defined piece by piece from different polynomials of the same order. It is intended to ensure that locally different behavior can reasonably be reflected. The problem here is guaranteeing the smooth course of the spline at the "seams" - that is, the grid points where the different polynomials adjoin. Cubic splines ($k=3$) are frequently used, which employ additional conditions to ensure that the curve can be continuously differentiated twice at the grid points, with the result that the function profile does not suffer from any jumps or "kinks". We do not provide an explicit illustration of the spline basis function at this point, see [26]. We wish only to mention that the functions in question are the result of $(k+1)$ polynomial pieces of the order $k$, which are composed such that they can be continuously differentiated $(k-1)$ times and differ from zero only in a section of the experimental area formed by $k+2$ neighboring knots. The status whereby a function differs from zero only in a limited section is referred to as a "function with compact support".
- **Radial functions**: Here, the functions used as the basis can have a global character but rapidly decrease to a local point in the experimental area as the radial distance increases (hence their name), and are normalized to 1. These functions are also referred to as "kernels". A typical example is functions shaped by the probability density of Gaussian distribution (Gaussian bell curve), which possess the important characteristic of having a very smooth profile, so that they naturally resolve the problem of the continuous course of the model equation. This supports the advantages of a local model approach, without having to put up with the disadvantages of an insufficiently smooth model curve.
- **Sigmoid functions:** Functions of this kind typically emerge within the framework of the artificial neural network (ANN) concept. An ANN is a network composed of artificial structures, which are modeled on natural neurons (nerve cells). Each neuron can be mathematically described by its activation function, which defines its task dependent on the inputs. A sigmoid function is frequently used as an activation function.

The example below demonstrates the suitability of the above-mentioned alternative model approaches:

*EXAMPLE 2: For the function from Example 1 above, a regression is to take place in zone $I_2$ on the basis of the 7 grid points contained therein. Instead of polynomials, trigonometric functions, spline bases, radials and sigmoid functions are to be employed as basis functions for the model.*

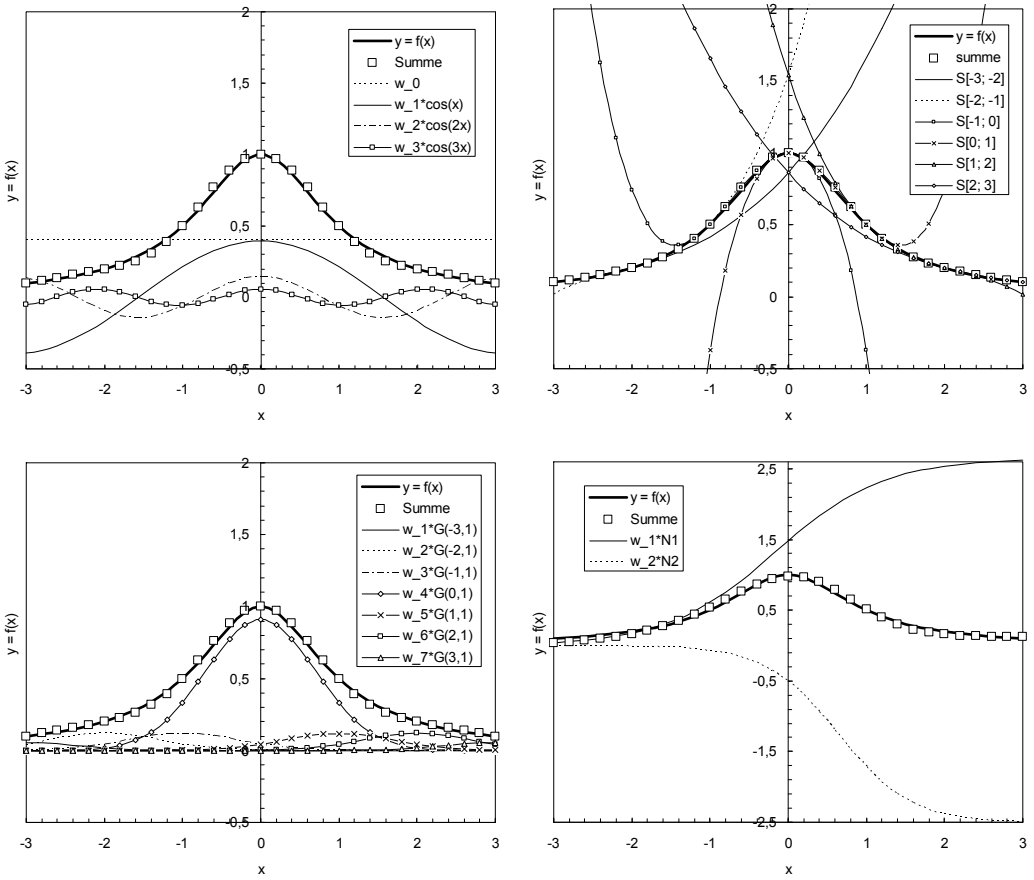



**Figure 8.4:** Compensating curves with alternative model approaches

*The results of the 4 approaches are illustrated in Figure 8.4. The thickest curve represents the actual function, the curve with the large squares the respective approximation. These diagrams show that all approaches are perfectly capable of reflecting the function's global behavior.*

*Further, the thinner curves represent the weighted functions used (or part polynomials in the case of the spline). 3 cosine terms and one constant were used as trigonometric functions (Figure 8.4, top left). The cubic spline in the top right diagram of Figure 8.4 was defined as 6 polynomials of the 3rd degree S[a,b]; each of these is valid in the interval [a,b]. For the radial basis functions, 7 Gaussian functions were used (Figure 8.4, bottom left). Finally, in the diagram at the bottom right of Figure 8.4, an ANN with sigmoid functions was employed as the activation function for the approximation. The weight coefficients (with the exception of the spline) were calculated by minimizing the sum of the residual squares using the MS Excel optimizer (solver). The spline was calculated using the Java applet (http://www.arndt-bruenner.de/mathe/scripts/kubspline.htm).*

The above univariate examples flag up alternative methods for selecting which model approach to deploy. The most important difference from classic polynomial approaches consists in employing functions that differ locally, instead of adapting a global, complex polynomial to the test results through parameter identification. Within this context, such approaches are often also referred to as "non-parametric models". In the case of multi-factorial tasks with extensive experimental data, the process is complicated, so that the use of suitable software is indispensable. A suitable solution based on the Gaussian basis function approach is contained in the Appendix.

### 8.4. Stochastic experimental designs

In the preceding chapters, we have presented a great variety of methods with which the relationship between predictor and response variables can be determined. All these techniques can be referred to as sampling methods, because they attempt to discretize a continuous experimental area through the generation of isolated grid points (treatments). Next, these grid points assist with the calculation of a response surface (the model equation), which is intended to approximately describe the true but unknown cause-effect relationships between predictor and response variables. These are thoroughly deterministic approaches, which generate the grid points according to a certain, predefined plan. In a full factorial design, for example, these are the corners of the experimental area. Consequently, all the approaches presented thus far can be termed *deterministic sampling methods*.



However, a whole group of applications, such as the analysis of robustness and reliability, require not just the functional relationship between predictor and response variables, but also the distribution of the response variable as a function of the predictor distributions. The distribution of the response variable answers the key question of the robustness and reliability analysis - namely, what proportion of all possible states is permitted within the context of the design requirements.

If no explicit models are available, the distribution of a response variable is typically determined by means of a Monte Carlo simulation, in which a random sample of values is generated for each predictor variable, in accordance with a distribution defined at the outset. In this way, each combination of values of the different predictor variables forms a possible "design", which already occurs with the "correct" frequency in the random sample of all designs. The distribution of the response variable can therefore be calculated by statistically evaluating the random sample of all designs.

A key task in Monte Carlo simulation is the generation of permitted combinations of predictor variables in the design space, whereby the random sample of each variable must originate from the distribution determined in advance. This task is very similar to the creation of an experimental design, on the other hand, it also differs from classic DoE approaches in that the combinations of predictors have to be formed not according to a fixed plan, but as the random sample of a distribution. Consequently, these techniques are referred to as *stochastic sampling methods* and the generated designs as *stochastic experimental designs*. We now go on to present the two most important approaches. Further details on this subject can be found in [20,22].

#### 8.4.1. Plain Monte Carlo (PMC)

The PMC approach is the most obvious strategy in which, firstly, a random number generator generates $N$ pseudo-random numbers $u_i$, which are evenly distributed in the interval [0,1]. Next, these numbers are interpreted as probabilities of the associated attribute values $x_i$ of a certain distribution function $F(x)$. The attribute values can thus be calculated by the application of the inverse cumulative distribution function.

$$x_i = F_x^{-1}(u_i)\,. \tag{8.7}$$

With this approach, pseudo-random numbers can be calculated for every distribution for which an inverse of its cumulative distribution function is defined [22].

The chief advantage of this method is evident in its robustness and simplicity. It betrays a certain disadvantage in that, due to the random nature of the samples, random deviations from the ideal distribution can also occur which, because of the law of large numbers, are only eradicated when very large numbers of samples are used. As a result, the Monte Carlo simulation converges only relatively slowly. An enhancement of the PMC method provides assistance here, and is described in the next section.

### 8.4.2. Latin hypercube (LHC)

When the PMC method is used, random deviations from the ideal form of distribution can occur, as mentioned above, as predefined by the density function $F(x)$. For example, with the random sample of a normal distribution, the values $F(x)$ in a normal distribution plot would not be positioned on an ideal straight line, but random minor deviations would occur.

These deviations can be described as follows: If the scope of a random sample is $N$, and $N$ equiprobable (but generally not necessarily equidistant) classes are formed, ideally each class may only have one member during classification. However, random deviations from the ideal state can result in not all of these classes being occupied, or certain classes having several members. This is where the LHC method comes in. If the volume of the random sample is $N$, the probability interval is initially subdivided into $N$ equiprobable subintervals

$$D_i = (x_i, x_{i+1});\ F(x_{i+1}) - F(x_i) = 1/N;\ i = 1...N \tag{8.8}$$

Next, in *each* of these intervals, *one* random number is generated, as with the PMC method, whereby either the actual distribution is taken into consideration or, for the sake of simplicity (with greater quantities of random samples), the midpoint of the interval is used. The LHC method therefore endeavors, figuratively speaking, to give chance a helping hand.

The LHC technique boasts clear advantages over the PMC method in terms of convergence speed, so that smaller quantities of random samples can be used with the same accuracy, see [20]. Since the scope of random samples also determines the required number of simulations (calls of external solvers) in the Monte Carlo approach, the advantages are obvious.



## 8.5. Special experimental designs

Components are subjected to trials for verifying and validating quality attributes such as functionality, durability, robustness and safety. For this purpose, special experimental techniques are employed, a description of which would exceed the boundaries of this volume. Such methods are therefore not dealt with any further at this point, for details, see e.g. [17].

However, we would like to mention at this point that nothing would generally contraindicate the use of DoE approaches for such trials, so that the maximum of information about the cause-effect relationships could be obtained without wasting resources. In some cases, to gain new knowledge of the product it would suffice to systematically record the predictor variables that take effect during a typical endurance run and to take these into account in the evaluation. Alternatively, a DoE scheme would enable influences on the service life of components to be examined, with the endurance test constituting a complete series of runs.

This gives rise to a hierarchical procedure, in which the structure of the tests is predefined by a DoE scheme at meta-level, and any number of tests can be performed at the operative level, enabling us to determine the response variables under examination. These do not have to be simple tests, but may themselves incorporate complex tests with long runs, backed up by simulations.

## 8.6. Taguchi experimental designs

The approach presented here was developed by the Japanese quality engineer and statistician G. Taguchi (born in 1924) after the 2nd World War, with the aim of rendering production more robust against disturbing influences and avoiding losses in quality, a key feature in the Japanese quality philosophy. In combination with the Six Sigma quality philosophy, this procedure quickly became very popular, and in some circles is virtually synonymous with experimental design.

Without wishing to call into question Taguchi's undisputed contribution to the development and popularization of the design of experiments and, in particular, his understanding of the relationship between quality and robustness, some of the approaches he proposed are regarded more critically today. Despite this, within the framework of his basic idea, knowledge of classic experimental design can undoubtedly be put to good use, as described in the preceding chapters. This gives rise to a method that reflects the spirit but not the disputed disadvantages of the original approach, so that even today, this strategy is still contemporary. A brief introduction is provided below; further details can be found in [3,24], and numerous applications in [25].



### 8.6.1. Motivation: robust products and processes

The origin of many experimental tests is a quality problem whereby a high proportion of rejects is produced because of high variance in quality attributes. We then ask not only "why" this is happening, but "how" this variance can be reduced. For example, the product of a manufacturing process must be as insensitive (or robust) as possible to fluctuating ambient conditions, such as temperature and air humidity.

A robust process is one with a result that deviates as little as possible from a predefined target value, i.e. its variance is as small as possible. The objective of process optimization is hence to reduce variance.

The fundamental question as to how predictor variables can influence robustness was already discussed in section 7.3. At this point, we simply wish to explain how favorable values can be found for those parameters that have a major influence on variance, without changing the mean of the process.

### 8.6.2. Taguchi experimental designs

In his quest for robust processes, Taguchi distinguishes between two types of factors, see [1], p.149:

- **Control factors** are factors that can be adjusted both in laboratory tests and in production.
- **Noise factors**, on the other hand, can be adjusted in laboratory tests, but not later on in production, e.g. for cost reasons. When a process is robust, it will be as insensitive as possible to these factors, e.g. position in the manufacturing appliance, ambient temperature, etc. These factors must not be confused with factors (also referred to as noise factors), which are basically unknown and are therefore not adjustable, neither in experiments, nor in production.

The procedure basically consists in finding those treatments that minimize the response variance (due to the variance in noise factors). Two types of designs are employed for this purpose:

- For control factors, a so-called **inner array** is used, also referred to by Taguchi as an "orthogonal array". In classic statistical experimental design, this constitutes a fractional factorial experimental design with resolution III, see [1], p.159. As discussed in section 8.1, designs with this resolution bring disadvantages and can lead to problems due to the confounding of primary effects and twofold interactions. The use of a resolution IV design is therefore recommended.
- For noise factors, an **outer array** is used, which concerns either one factor in several levels or a resolution III fractional factorial design for several factors. The low resolution is not critical at this point, because the outer array is deployed as a kind of test replication in further procedure.

### 8.6.3. Evaluation and result

Evaluation takes place as follows:

- To record variance, the signal-noise ratio

$$S/N = 10 \cdot \log_{10}\left(\frac{\bar{y}^2}{s^2}\right), \tag{8.9}$$

  or another suitable variable (e.g. standard deviation $\sigma$, or variance $s^2$) can be recorded as a separate response variable in addition to the mean $y$.

- The effects of the mean and S/N are calculated as if the outer array were a replication of the experiment, even though this is not the case because the noise factors are being varied in a targeted manner. A conservative conclusion about significance can be drawn, for the significance of the effects is slightly underestimated. For the response variable S/N there is only one run; this response variable is therefore treated the same as in tests with one run $n$=1.

- Depending on the result, the factors are apportioned as follows:
  - Factors that influence variance (and possibly the mean as well) are selected such that variance is minimized.
  - Factors that only influence the mean are used to adjust the response variable in line with design stipulations.
  - Finally, factors that influence neither the mean nor variance are used to minimize expenditure.

*EXAMPLE: Table 8.6 presents an example with 3 control factors C1 to C3 and 3 noise factors N1 to N3. A full factorial design on 2 levels is used as the inner array, the $2^{3-1}$ design as the outer array. The significance of the effects of the mean and S/N were rated as "---", "*" "**" (not significant, significant, highly significant). With this evaluation, the factors C1 and C2 would appear the obvious choice for increasing robustness, as they have a significant effect on S/N. In an optimization, they must therefore be adjusted on the level that displays a smaller variance (high S/N). This would change the mean of the response variable, however. Correction can be achieved using the factor C3, which influences only the mean of the response variable, but not S/N. C3 is therefore adjusted in such a way that the response variable demonstrates the necessary mean.*

| | | | N1 | - | + | - | + | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | N2 | - | - | + | + | | | |
| | | | N3 | - | + | + | - | | | |
| No. | C1 | C2 | C3 | 1 | 2 | 3 | 4 | y | s | S/N |
| 1 | + | + | + | $y_{11}$ | $y_{12}$ | $y_{13}$ | $y_{14}$ | $y_1$ | $s_1$ | $S/N_1$ |
| … | | | | | | | | | | |
| 8 | - | - | - | $y_{81}$ | $y_{82}$ | $y_{83}$ | $y_{84}$ | $y_8$ | $s_8$ | $S/N_8$ |
| $\Delta_y$ | * | --- | * | | | | | | | |
| $\Delta_s$ | | | | | | | | | | |
| $\Delta_{S/N}$ | ** | * | --- | | | | | | | |

**Table 8.6:** Evaluation of a Taguchi design

Alternatively, robustness to noise factors can also be investigated using classic experimental designs. Here, the noise factors must be taken into consideration as the factor, and the variance as the response variable. Robustness can be influenced by interactions between control and noise factors.

## 8.7. Shainin® method

The American consultant Dorian Shainin (1914 - 2000) put forward a method that differs starkly from the other approaches presented here. In a wider sense, however, the Shainin® method can also be understood as a variant of DoE, which offers alternative approaches to designing, conducting and evaluating experiments. The terms tagged with ® and ™ within Chapter 8.7 are trademarks of Red X® Holdings LLC.

### 8.7.1. Motivation: the search for the Red X®

The Shainin® method commences with the assumption that the so-called Pareto principle applies to most problems. The Pareto principle states that a phenomenon that can theoretically have many causes only has very few causes in reality.

Shainin® recommends five screening processes for detecting these relevant causes. These are characterized by a great simplicity, but can only be applied to specific questions. Instead of drawing on relevant prior knowledge and experience, he recommends "letting the parts speak". This recommendation is therefore based on the assumption that the problem in question concerns an ongoing production process. With this method, the response variable is commonly referred to as Green Y®. The factor that has the greatest influence (50% of the effect) on the result Y is called Red X®.



As a general guide, the approach can be briefly described as follows:

1. From many parameters, identify max. 20 as significant through systematic observation
2. From these 20, determine the 4 most important through simple tests
3. On this basis, implement a complete experimental design with subsequent optimization

### 8.7.2. Systematic observation

The first task is that of systematic observation, of “letting the parts speak”, and finding out as much as possible about the matter under investigation. This approach is particularly apt in the latter phases of the PCP or during production, if prior knowledge, samples and data are available.

#### 8.7.2.1. Multivari charts

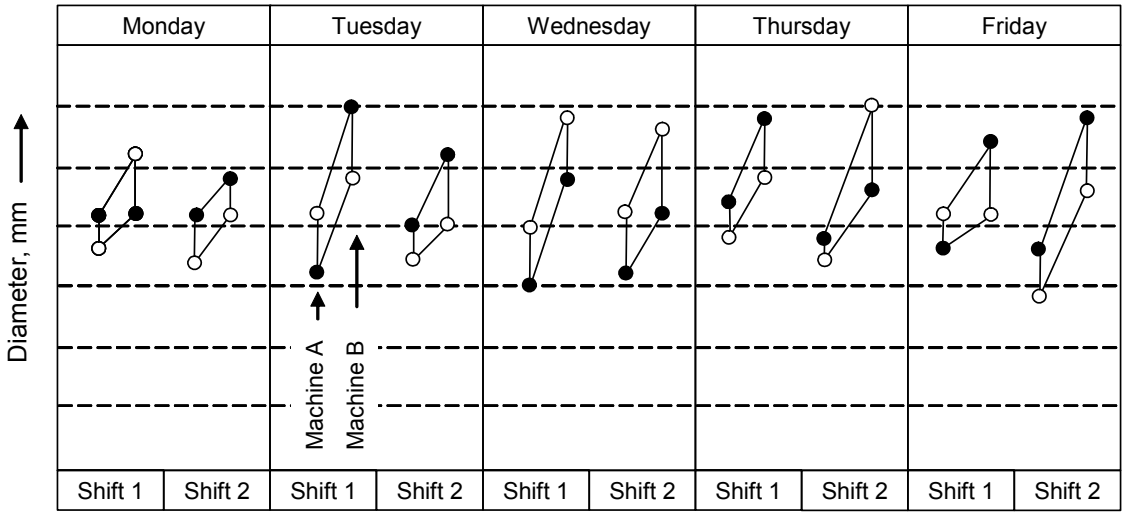



**Figure 8.5:** Multivari chart

The Multivari chart method constitutes an important systematic observation technique. It was developed by L. Seder in 1950. Process fluctuations are graphically represented. Using original values or vertical lines, the fluctuations are plotted independently from one another on the basis of location, batch, time, etc. In certain circumstances, this clarifies where the primary causes of fluctuations lie. Shainin® makes the following distinctions:

a) Position-dependent variations, e.g. position within a unit, machine to machine, clamping point to clamping point, operator to operator, line to line, plant to plant,

b) Cyclic variations, e.g. lot to lot, batch to batch,

c) Variations over time, e.g. hour to hour, shift to shift, day to day or week to week.

A Multivari chart is simply an intelligent graphic representation of data arranged according to groups of variance groups. The reading of the attribute under test (the Green Y®) is entered on the ordinate (y-axis). In Figure 8.5, the means of 3 readings are illustrated as dots. The black dots denote clamping point 1, the white dots clamping point 3. It is clear that machine B systematically produces higher results than machine A.

#### 8.7.2.2. Concentration charts

The observation of markedness is sometimes evident as a “concentration” of a peculiarity. This is the principle behind concentration charts.

Insofar as the cause of a problem can be found within a unit, it makes sense to record the precise location and frequency of the error(s) in an error location diagram. To this aim, a simple schematic representation of the product is used, in which the point at which the error or deviation occurs is marked. The example in Figure 8.6 originates from the Elementary Tools volume from Quality Assurance.



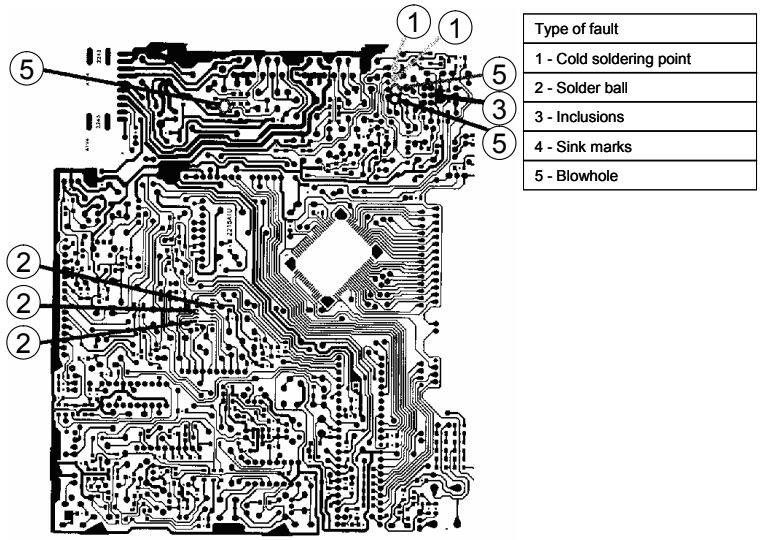



**Figure 8.6:** Concentration chart

### 8.7.3. Simple tests

#### 8.7.3.1. Component search™

With component search™, individual parts are systematically swapped between a “good” and a “bad” unit, in order to locate the part that led to quality deficiencies in the bad unit. The component search method is used often in the repair of faulty equipment.

Prerequisites:

- Used primarily in assembly
- The response variable must be measurable with a measurement uncertainty equivalent to no more than 20% of the tolerance or the process variance.
- It must be possible to dismantle and reassemble good and bad parts, without altering anything about the “good” or “bad” result.

Procedure (12 steps):

1. Select 2 parts from a day's production - one very good and one very bad (BOB = best of the best, WOW = worst of the worst). The further apart the measurement results of these parts, the easier it is to find the Red X®.
2. Dismantle and reassemble both parts and measure the response variable again twice.
3. Perform a significance test on the basis of the measurement results. The results of the "good" part must be considerably better than the results of the "bad" one. Criterion:

   $D/d \geq 5$, with $D = \bar{x}_{BOB} - \bar{x}_{WOW}$ and $d = \frac{R_{BOB} + R_{WOW}}{2}$, see Chapter 8.7.4.1. (8.10)

4. If $D/d < 5$, the problem lies not with the components but with their assembly. Suggestion: Examine the assembly step by step. Otherwise, one or more of the components or sub-components must be the Red X®. Then commence the component search.
5. Set the components in order according to the size of the suspected influence.
6. Swap component A between BOB and WOW (A = most probable candidate).
7. Analyze the results.
8. Always reassemble, to see whether the original state is restored.
9. Repeat steps 6, 7 and 8 with components B, C, etc.
10. Identify one or more components as the Red X®. A component is the Red X® if a complete reversal takes place, i.e. BOB becomes WOW and vice versa, when a simultaneous swap of this component between BOB and WOW takes place. An interaction is present if a WOW becomes a BOB or a BOB becomes a WOW, but not the other way round.
11. Confirmation test: If you take a BOB product and a WOW product, and swap the components considered unimportant as a group, the BOB result and the WOW result should not change in any way. If, on the other hand, you swap all components considered important as a group, a noticeable, complete reversal should occur.
12. Calculate the primary effects and interactions

Component swap test results are represented schematically according to [12], as shown in Figure 8.7.

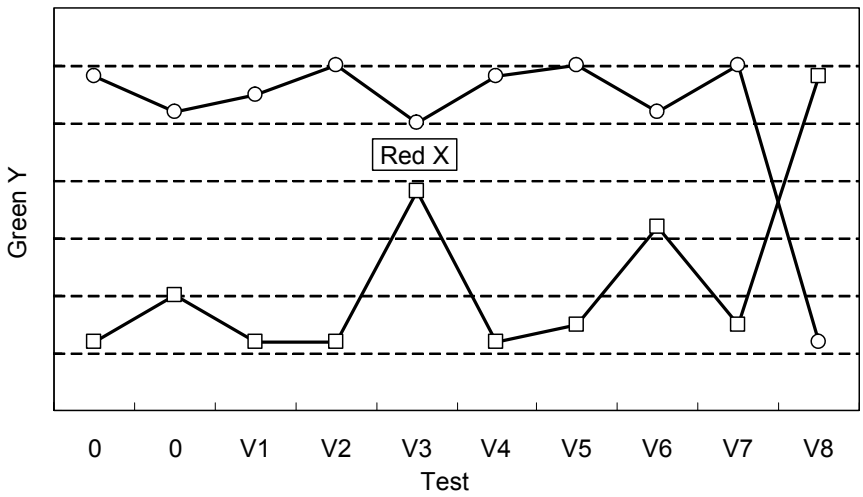



**Figure 8.7:** Component search. The first two tests simply consist of dismantling and subsequent reassembly. In test V3, the Red X® has been discovered. The final step is a confirmation test.

#### 8.7.3.2. Paired comparisons™

The paired comparisons™ is a very similar method to component search™. It is employed when units cannot be dismantled into individual components. Good and bad units from corresponding populations are mutually compared with respect to measurable quality attributes. Quality attributes that repeatedly

differ from one another to a noticeable extent may be responsible for the difference in quality between the units. This suspicion is then put to the test in further comparisons.

#### 8.7.3.3. Variable search™

The variable search™ is a technique that screens the most important factors from a mean number (5 - 15) of them as effectively as possible. In a certain sense, the variable search is a single-factor test. The variable search can only be applied when we basically know, when establishing two levels for each factor, which level is "better" and which level is "worse". This vital prerequisite is, of course, quite far-reaching. In other words, the variable search is not the search for an optimum, but rather the question as to which factors make a decisive contribution to the optimum.

Procedure:

1. Conduct comparative tests of "all factors on the bad level" $V_S$ with "all factors on the good level" $V_G$. If no major difference is perceptible, no further investigation is necessary.
2. The influence of the individual factors is now examined separately. If $A$ is presumed to be the strongest factor, a test is conducted in which $A$ is assigned to the good level and all other factors to the bad level. If the result roughly corresponds to the result of the test $V_G$ in 1., we can assume that *A* has a dominant influence on the response variable.

   This is then confirmed by means of a cross check, in which $A$ is assigned to the bad level and all other factors to the good level. If the dominance of $A$ prevails, the result of $V_S$ would be repeated. If $A$ is dominant, the test does not need to be continued. If $A$ is not dominant, $B$, etc. are examined in the same manner.
3. If none of the factors is dominant, the factors with a perceptible influence are grouped and assigned once to the good and once to the bad level, while the remaining factors are assigned to the opposite level.

   If the results of 1. are repeated at this point, we have demonstrated that it suffices to assign the factors that revealed an influence in the single-factor comparisons to their good level, while the remaining factors can be assigned to their bad level.

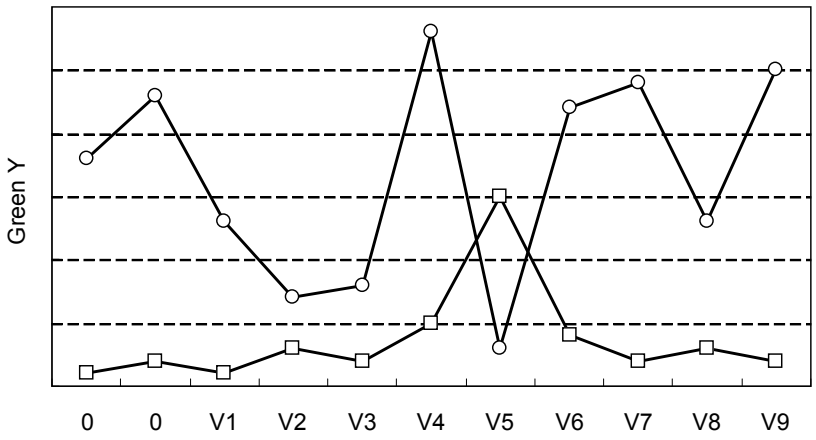



**Figure 8.8:** Variable search. The first two tests simply consist of dismantling and subsequent reassembly. In test V5, the Red X® has been discovered.

Now that we have reduced the number of factors to a maximum of 2 - 4, Shainin® recommends the use of the familiar full factorial designs with 2 to 4 factors on 2 levels, preferably with no more than 2 replications per treatment. These designs are evaluated in the usual manner, e.g. through the analysis of variance. In this way, it is possible to determine which factors or which interactions are particularly important. Since Shainin® recommends relatively small numbers of random samples, it is obvious that only major effects can be demonstrated. One of the recommended procedures is the single-factor comparison (B vs. C™).

Finally, it is also worth mentioning that Shainin rejects the use of fractional factorial designs, because confounded designs carry the risk of misinterpretation. An example variable search is presented in Figure 8.8.

#### 8.7.3.4. Product/process search

The aim of this approach is to separate important and unimportant process parameters, on the understanding that parts can be tracked during the process and the process parameters can be measured while these parts are undergoing production.

Procedure:

1. List the possible causal process parameters in order of suspected probability of influence
2. Document which process parameters are measured precisely where and by whom
3. Make sure that the measurement uncertainty is < 20% of the tolerance
4. Ensure that the actual parameter values are measured, not just the settings
5. If, during monitoring, an individual process parameter displays no variance whatsoever, it can be eliminated from subsequent testing.
6. Conduct a 100% test of the manufactured products, until at least 8 good and 8 bad parts have been found. The span between the best and the worst part must equal at least 80% of the process scatter range observed so far.
7. During the production of a part, each of the process parameters from the list must be measured and recorded
8. Paired comparison of the process parameter values that resulted in “good “ parts with those that led to “bad” parts
9. Perform a B vs. C™ test.
10. Variable search™
11. Further optimization and, if necessary adjustment of tolerances

### 8.7.4. Further tools

As an alternative to classic significance tests such as the t-test, for example, Shainin® mentions a few simple test procedures that should supply the same information with less expenditure. A brief explanation of these and some other representation tools now follows.

#### 8.7.4.1. D:d rule

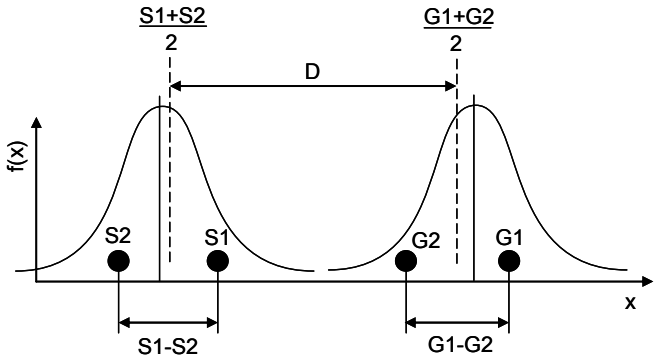



**Figure 8.9:** D/d rule

In component search, the D:d rule can be used to decide whether or not the difference in test results from the good and bad units is significant. A significant difference exists if D/d ≥ 5. This rule can, of course, equally be applied to the process of variable comparison, for comparing the test results on the good and the bad level.

The calculation below elucidates the importance of the variables D and d and the representation of the rule D:d ≥ 5:1:

$$d = \frac{G_1 - G_2}{2} + \frac{S_1 - S_2}{2},\ D = \frac{G_1 + G_2}{2} - \frac{S_1 + S_2}{2}. \tag{8.11}$$

#### 8.7.4.2. B vs. C™

B vs. C™ stands for "better versus current", so it simply signifies the comparison between these two states. The object of this comparison may be a product, a process or a method, for example.

B vs. C™ is a simple (non-parametric) statistical significance test, and purely a version of the Tukey test. If we have 3 "B" values and 3 "C" values that do not overlap, their difference on the 95% level is significant. Where overlapping is present, the Tukey test must be conducted with a larger number of random samples.

#### 8.7.4.3. Isoplot®



For examining and assessing measurement uncertainty in proportion to product variation, Shainin® suggests as an alternative to the classical approach (see [15,16]) the so-called Isoplot®.

For this method, a random sample of 30 parts is taken (parts are numbered). The attribute of interest in each part is measured (1st series of measurements). A repeat measurement is conducted in random sequence (2nd series of measurements). The two results ascertained for one part form the $x$ and $y$ coordinates of the corresponding point in the plot (representation of all 30 pairs of variates).

It must be borne in mind that abscissa and ordinates have the same scale, and the graph will therefore be square. The points are contained in a kind of "Zeppelin". The length $L$ of this area (in mm) is a measure of process variation (see remarks), its width $\Delta M$ (also in mm) an indicator of the measurement uncertainty.

According to Shainin®, measurement uncertainty is sufficiently small if $L/\Delta M \geq 8.5$.

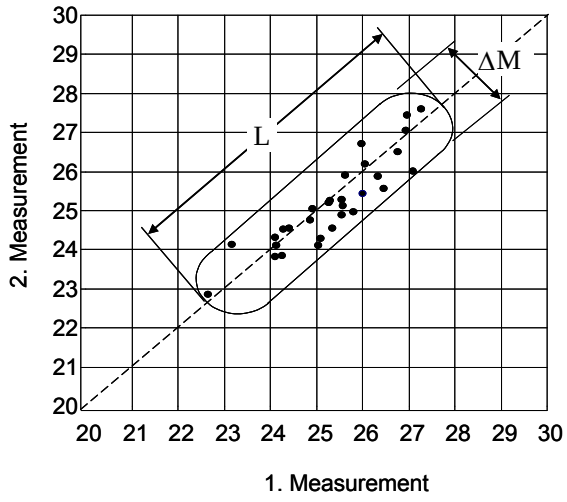



**Figure 8.10:** Isoplot®.

# 9. Appendix

## 9.1. Fundamental concepts in statistics

In this section, we will briefly discuss some of the fundamental ideas that are required in order to understand the subject of experimental design.

### 9.1.1. Data series and their characteristic values

Even when an experiment is meticulously conducted, the results will differ - they will vary - if the test is performed more than once, although the conditions remain constant. A data series

$$x_i\ ,\ i \in 1...n\ , \tag{9.1}$$

with $n$ values of the response variable under observation can therefore be regarded as the result of a stochastic process. Within the context of descriptive statistics, data series of this kind can basically be described by the following characteristic values:

- Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , \tag{9.2}$$

- Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{9.3}$$

  or their square root, the standard deviation $s$, which characterizes the distribution of the values around the mean.

Since the individual values of the data series are subject to variance, this series will itself differ if determined several times. Thus, its characteristic values will also be different. It is important to recognize that the mean and the standard deviation are also subject to variance. However, the variance of the mean is considerably smaller than that of the individual values, and it becomes smaller the larger the number of random samples. We can therefore demonstrate (see [22], p.122) that

$$s_{\bar{x}}^2 = \frac{s^2}{n}\ . \tag{9.4}$$

Data series can be represented in histograms, which are produced when the values of the data series are classified. A histogram therefore presents frequencies of the occurring classes of attributes. Total frequencies can be deduced by adding the classes together.

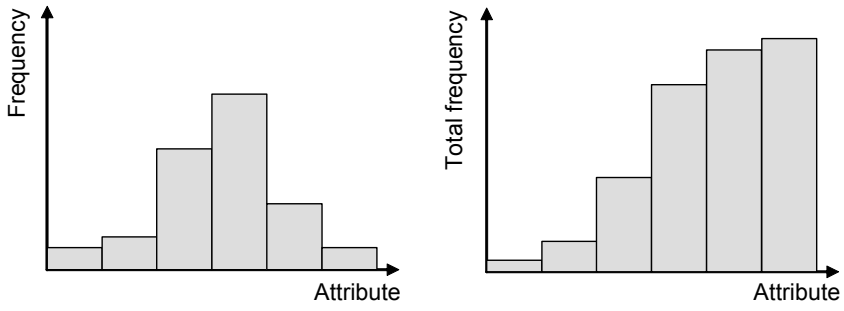



**Figure 9.1:** Histogram and total frequency curve (schematic diagram).

### 9.1.2. Distributions and their characteristic values

#### 9.1.2.1. Distribution and distribution density, moments

Things are seen in a somewhat different light if the data are viewed as a random sample from an infinitely large population. Thus

$$x_i\ ,\ i \in 1...n\ , \tag{9.5}$$

is construed as manifestations of a random variable $X$. A theoretical model for describing the population of all possible values of the variable is provided by the distribution density function

$$f(x)\ . \tag{9.6}$$

This assigns a number $f(x)$ to the value $x$, similarly to the manner in which a relative frequency is assigned to the value $x$ in a histogram. We can show that if the number of random samples under observation is very large and the class widths of the histogram are declining in size, the distribution density function can be derived directly from the histogram, with the result that the latter can be regarded as an approximation of the density function.

From the total frequency curve, a function can also be derived - distribution function $F(x)$ - by crossing the limit to infinity. For each value $x$, this states the probability that a random variable $X$ will assume a value lower than $x$. The distribution function of a distribution is generally obtained by integrating its density function:

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(v)dv\ . \tag{9.7}$$

Here, the independent variable of the function $f(v)$ is designated $v$, in order that it can be differentiated from the upper integration limit $x$. Thus, the function $F(x)$ is a function of the upper integration limit, while $v$ is a variable by which integration takes place.

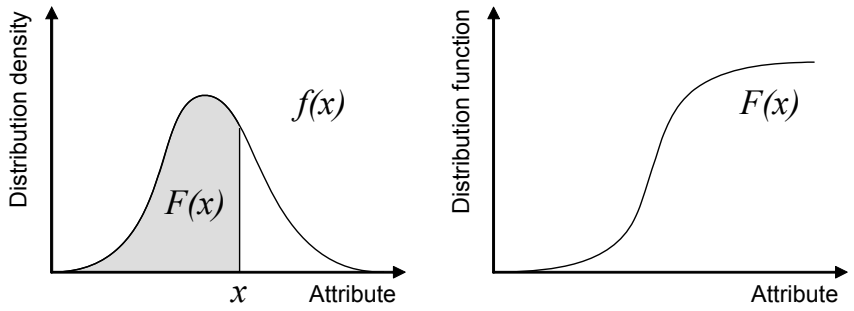



**Figure 9.2:** Distribution density and distribution function (schematic diagram).

The distribution function assigns a probability to the value $x$ of the variables $X$. Certain preconditions must therefore be satisfied for this to take place, so that a function can be employed as a distribution function:

- Within the limits, the following must apply

  $F(-\infty) = 0$ , $F(+\infty) = 1$, from this, we can deduce $\int_{-\infty}^{+\infty} f(v)dv = 1$. (9.8)

- For all $-\infty < x < +\infty$ , $F(x) \ge 0$ and $F(x)$ must be a monotonously increasing function of $x$.

The density and distribution function of a distribution possess characteristic values (also known as moments), which determine the appearance of the distribution. The most important are

- the mean (sometimes known as the population mean)

$$\mu = \int_{-\infty}^{+\infty} x \cdot f(x)dx \tag{9.9}$$

- and the variance

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)dx\,, \tag{9.10}$$

which describe the position and width of the distribution.

#### 9.1.2.2. Examples of distributions

The simplest means of illustrating the above matter is the example of normal distribution, which possesses the well-known “bell-shaped” distribution density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,. \tag{9.11}$$

By integrating the density function, we obtain the s-shaped distribution function of normal distribution

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2}\, dv\,. \tag{9.12}$$

As the integral cannot be further analyzed, the above equation cannot be simplified any further.

The equations show that with normal distribution, the mean and standard deviation characteristics occur as parameters in the distribution function; they are often also referred to as $N(\mu,\sigma)$. This makes working with them especially simple. The fact that this is not necessarily the case for every distribution, however, is demonstrated by the example of exponential distribution with the density function

$$f(x) = \lambda e^{-\lambda x}\,,\ x \geq 0\,, \tag{9.13}$$

in which a parameter $\lambda$ occurs. The relationship between this parameter and the above characteristic values is described as follows:

$$\mu = \frac{1}{\lambda}\,;\ \sigma^2 = \frac{1}{\lambda^2}\,. \tag{9.14}$$

Here, the relationship between distribution parameters and characteristic values is relatively simple, so that conversion can take place without problem. It becomes more complicated in the case of the Weibull distribution, which plays an important role in the description of service life; there, conversion is no longer that straightforward. We do not go into further detail at this point, however.

#### 9.1.2.3. Characteristic values of independent and dependent variables

If a variable $Y$ indicates a linear dependence of $n$ independent random variables $X_i$

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i\,, \tag{9.15}$$

then $Y$ will also be a random variable. The distribution of $Y$ cannot generally be calculated without problem, particularly if the variables $X_i$ have different types of distribution. Nevertheless, it is possible to acquire information about the characteristic values of the $Y$ distribution; this will depend upon the characteristic values of the distribution of $X_i$ [22] p.150:

$$\begin{aligned} \mu_Y &= a_0 + \sum_{i=1}^{n} a_i \mu_{Xi} \\ \sigma_Y^2 &= \sum_{i=1}^{n} a_i^2 \sigma_{Xi}^2 \end{aligned} \tag{9.16}$$

In particular, the mean of $n$ normally distributed, independent variables

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \ i = 1..n \ , \ X_i \in N(\mu,\sigma) \ , \tag{9.17}$$

is also normally distributed with

$$\mu_{\overline{x}} = n\frac{\mu}{n} = \mu \text{ and } \sigma^2_{\overline{x}} = n\frac{\sigma^2}{n^2} = \frac{\sigma^2}{n} \ . \tag{9.18}$$

#### 9.1.2.4. Estimating the mean, confidence interval, t-distribution

We are generally up against the question as to how to determine the characteristic values of the unknown population. It can be demonstrated that

- the mean of a random sample $\overline{y}$ can be used as an estimate for the mean of the population $\mu$, and
- the variance of a random sample $s^2$ can be used as an estimate for the variance of the population $\sigma^2$.

In addition, both estimates are also unbiased, i.e. with many replications, they approximate the values of the population.

The quality of this point estimate is initially questionable, but can be evaluated as described below. If a random variable $X$ is normally distributed, the variable

$$u = \frac{X - \mu}{\sigma} \ , \ X \ \in N(\mu,\sigma) \tag{9.19}$$

is also normally distributed with the mean 0 and the standard deviation 1. This distribution is known as “standard normal distribution”. An estimate of the probability P with which $X$ lies within a given interval $\mu - u \cdot \sigma \le X \le \mu + u \cdot \sigma$ can be made with the aid of the parameter $u$:

$$P(\mu - u \cdot \sigma \le X \le \mu + u \cdot \sigma) = 1 - \alpha \ . \tag{9.20}$$

For example, the variable $X$ will lie in the interval $\mu \pm 1 \cdot \sigma$ with a probability of 68.3%. It is therefore clear that $u$ equals the attribute value of standard normal distribution with probabilities (1-$\alpha$/2) and $\alpha$ /2.

We will now examine the mean of $n$ independent, normally distributed variables

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \ X_i \in N(\mu,\sigma) \ , \ i = 1..n \ , \ \overline{X} \in N(\mu,\frac{\sigma}{\sqrt{n}}) \ . \tag{9.21}$$

The variable

$$u = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \tag{9.22}$$

has, by definition, a standard normal distribution. However, the standard deviation of the population is generally unknown, with the result that it must be estimated on the basis of a random sample: $\sigma \approx s$. At the same time, we can shown that in this case, the quantity

$$t = \sqrt{n}\,\frac{\overline{X} - \mu}{s} \tag{9.23}$$

is no longer normally distributed; its distribution is described by the so-called t-distribution (or student’s t-distribution). For small values of the parameter $n$, the t-distribution manifests a greater width and edge definition than normal distribution; for the large values of $n$, it approaches normal distribution.

An estimate of the probability $P$ with which $\overline{X}$ lies within a given interval can be made with the aid of the parameter $t$. By the same token, an indirect conclusion can also be drawn regarding an interval

$$\overline{X} - t \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t \cdot \frac{s}{\sqrt{n}} \tag{9.24}$$

in which the unknown population mean is located. Here, the probability

$$P\left[ \overline{X} - t \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t \cdot \frac{s}{\sqrt{n}} \right] = 1 - \alpha \tag{9.25}$$

is usually referred to as the *confidence level*, and the interval as the confidence interval. The more certain one wishes to be that the interval contains the true mean, the higher the confidence level and the larger the parameter $t$ that must be selected. Tabular values of $t$ are attached.

In the same way, a confidence interval can also be stated for the variance, although this is not discussed any further here. Details can be found e.g. in [22].

### 9.1.3. t-test, comparison of two means



The t-distribution can also be used to decide, on the basis of a so-called t-test, whether two random samples differ from one another significantly or not.

To clarify this, let's pursue the following thought experiment: Let us take a normally distributed population N($\mu$ , $\sigma$) two random samples each with scope n, calculate the means $\bar{y}_1$ and $\bar{y}_2$ and the standard deviations $s_1$ and $s_2$ (or variances $s_1^2$ and $s_2^2$), and thus the value

$$t = \sqrt{n} \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{s_1^2 + s_2^2}} . \tag{9.26}$$

$t$ can take on values between 0 and $+\infty$. If we repeat this process very often, we can expect most values to be close to zero, with large values occurring only rarely.

This 'thought experiment' was performed using computer simulation. For $n$=10 and 3000 random sample pairs ($t$-values), the histogram presented in Figure 9.3 was generated.

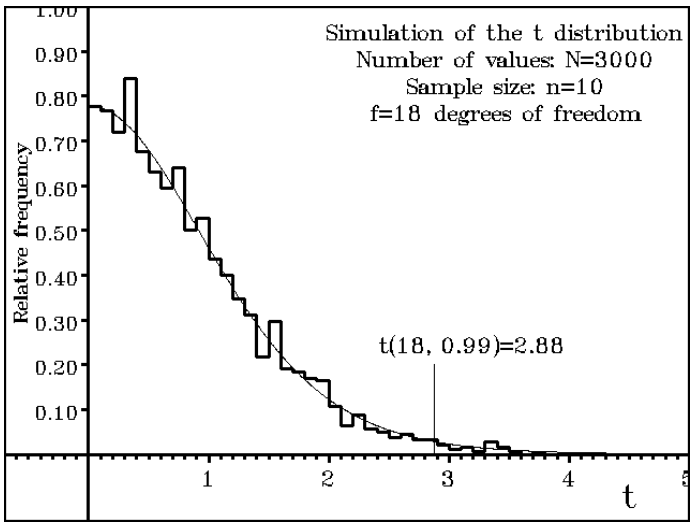



**Figure 9.3:** Histogram and density function of the t-distribution.

If the number of random samples tends to infinity and the class width simultaneously tends to zero, the histogram increasingly approaches the solid line that represents the density function of the t-distribution.

In this example, the upper limit of the 99% random scatter range (threshold value) is $t(18;99\%)$=2.88, i.e. larger values than 2.88 can only randomly occur in 1% of all cases. The threshold values of the t-distribution are set out in a table for various significance levels as a function of the degree of freedom $f=2(n-1)$.

The t-test procedure is based on the relationships presented above. This test is a statistical technique that can be used to decide whether the arithmetic means of two series of measurements each with a scope of $n$ belong to one and the same population, so that the means of the random samples differ from one another or not only by chance, and the difference between the means of the random samples is not random. The null hypothesis assumes that the means of the respective population are identical.

Preconditions for comparison are:

- The individual values of both random samples are representative
- Both random samples originate from one normal distribution
- The standard deviation of this normal distribution is the same for both random samples. Compliance with this precondition can be checked with the aid of an F-test.

With the means $\bar{y}_1$ and $\bar{y}_2$ and variances $s_1^2$ and $s_2^2$, the next step is to calculate the test statistic

$$t = \sqrt{n}\,\frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{s_1^2 + s_2^2}} \tag{9.27}$$



If the result is $t > t_{2(n-1);0,99\%}$, i.e. $t$ lies outside the 99% random scatter range, the null hypothesis can be discarded.

Remarks:

- The expression of the test statistic $t$ can only be applied in this simple form if the number of random samples is assumed to be the same ($n_1=n_2=n$). However, an equivalent calculation rule also exists for different numbers of random samples.
- In the form illustrated here, the t-test tests the null hypothesis $\mu_1=\mu_2$ against the alternative $\mu_1\neq\mu_2$. This question therefore has two sides. For this reason, the amount of the mean difference is incorporated in the expression for $t$. Consequently, $t$ can only assume values $\geq 0$, so that the distribution shown above is produced.
- In response to this two-sided question, Table 2 in section 9.3 specifies the threshold values 95%, 99%, and 99.9% of the t-distribution. They correspond to the one-sided threshold values 97.5%, 99.5% and 99.95%.

### 9.1.4. Minimum number of random samples

In the preparatory phase of experimental tests, the experimenter frequently asks himself which minimum mean difference $\Delta = \bar{y}_1 - \bar{y}_2$ is of interest to him in terms of his objective (system optimization, simplification of production, cost reduction), and what minimum number of random samples $n$ he should decide on in order for this minimum mean distance - if at all present - to be significant within the framework of the test evaluation.

The expression of the test statistic $t$

$$t = \sqrt{n}\,\frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{s_1^2 + s_2^2}} \tag{9.28}$$

shows that the larger $n$ must be in order for a “t-test response”, the smaller the mean difference $\bar{y}_1 - \bar{y}_2$ and the larger the variances $s_1^2$ and $s_2^2$ of the two series of measurements that are the subject of comparison. At the same time, we must bear in mind that the tabular value $t_{table}$ becomes smaller as the degree of freedom $f = 2(n-1)$ grows.

Ostensibly, a small difference between means with simultaneously "large" variance among the distributions means that the two groups of values are not or are scarcely visually distinguishable in a graphic representation of the two series of measurements.

On this basis, a rough estimate of the minimum number of random samples $n$ would be conceivable, and could be achieved by stating the mean difference as a multiple of a "mean variance" $(s_1^2 + s_2^2)/2$, and comparing the calculated test statistic $t$ with $t_{\text{table}}$ for various $n$ (note the degree of freedom and significance level!). As well as "trying things out" in this way, however, the minimum number of random samples can be deduced in a statistically accurate manner. This procedure is outlined only roughly here (mathematical foundations in [2]).

During the comparison of means of two series of measurements and the resulting test decision, two types of error are possible.

- Firstly, both series of measurements originate from the same population, i.e. there is no significant difference. If, on the basis of a t-test, we decide that a difference exists between the two means, we are committing a type I error ($\alpha$). This corresponds to the significance level of the t-test (for example $\alpha$).
- Secondly, if a difference between the means actually exists, i.e. the series of measurements originate from two different populations, the test does not demonstrate this with absolute certainty. The test result may purport, purely at random, that this difference does not exist. This constitutes a type II error ($\beta$).



For the experimenter, both types of error are disagreeable, because the suspected effect of an factor may cause him to propose expensive further tests, or even changes to a production process (type I error), for example, or because he does not recognize an effect that is actually present and thus loses the opportunity to suggest possible process improvements (type II error).

Under considerations of plausibility, as outlined above, the minimum number of random samples $n$ that is required to recognize a genuine mean difference on the basis of a comparison of 2 random samples of the same size $n_1=n_2=n$ depends upon the distance between the two means

$$D = \frac{\Delta}{\sigma} \tag{9.29}$$

stated in units of standard deviation $\sigma$, and the significance levels $\alpha$ and $\beta$:

$$n = 2\,\frac{\left(u_\alpha + u_\beta\right)^2}{D^2}\,. \tag{9.30}$$

During the concrete comparison of two series of measurements, the means $\mu_1$ and $\mu_2$ and the standard deviation $\sigma$ of the population (consequently also $D$) are not known. They are estimated by the empirical values $\bar{y}_1$, $\bar{y}_2$ and $s$. For this reason, the t-distribution must form the basis for the calculation of $n$ in accordance with the given formula. Thus, $u_\alpha$ and $u_\beta$ are the x-coordinates $u$ at which the t-distribution assumes the values $\alpha$ (two-sided) or $\beta$ (one-sided).

Smaller significance levels, i.e. smaller type I ($\alpha$) and type II ($\beta$) errors mean that the two distributions that are being compared - and consequently the associated distributions of the means - may only slightly overlap. To this aim, the selected number of random samples $n$ must be sufficiently large at the given mean distance $D$.

If our intention is to conduct an experiment using a full factorial experimental design with factors on 2 levels, the number of replications per treatment must be selected in such a way that the overall number of tests is

$$N_{tot} \geq 4\,\frac{\left(u_\alpha + u_\beta\right)^2}{D^2}\,, \tag{9.31}$$

since the effect of every factor in this case is defined by a difference of means of two samples with size $n = N_{tot}/2$.

### 9.1.5. Analysis of variance

#### 9.1.5.1. F-test, F-distribution

The F-test is a statistical technique that can be used to determine whether the variances of two random samples differ significantly. As with the t-test, how this test functions can best be explained using the results of a computer simulation.

From a normally distributed population $N(\mu,\sigma)$, we will take two random samples of size $n_1$ and $n_2$, calculate the random sample variances $s_1^2$ and $s_2^2$ finally the variable

$$F = \frac{s_1^2}{s_2^2} \tag{9.32}$$

$F$ can assume values between 0 and $+\infty$, but will only randomly deviate from 1. It is entirely plausible that if this process is repeated frequently, small values close to zero and very large values will only seldom occur.

The result of a computer simulation in which the $F$-values were determined for $N$=3000 random sample pairs with a sampling scope of $n_1$=$n_2$=9 is presented in Figure 9.4 as a histogram.

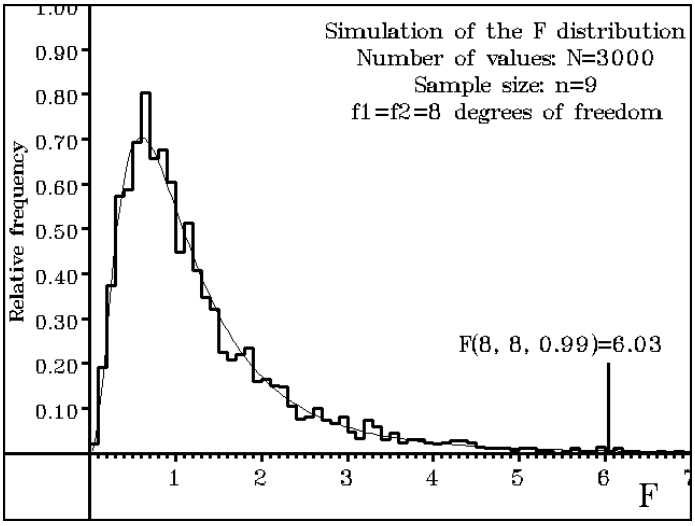



**Figure 9.4:** Histogram and density function of the F-distribution.

If we permit the number of random samples to tend to infinity and simultaneously allow the class width to tend to zero, the histogram increasingly approaches the solid line (density function of the F-distribution). The shape of the histogram depends upon the random sample sizes $n_1$ and $n_2$ of the random sample pairs under test; accordingly, the curve shape of the density function of the F-distribution depends upon the degrees of freedom $f_1 = n_1 - 1$ and $f_2 = n_2 - 1$.

In the example calculated here, the upper limit of the 99% random scatter range (threshold value) is $F_{8;8;0,99}$=6,03, i.e. only in 1% of all cases (significance level) is it, purely at random, $s_1^2 \geq 6{,}03 s_2^2$.

The threshold values of the F-distribution for the various significance levels as a function of the degrees of freedom $f_1$ and $f_2$ are set out in an attached table.

The relationship illustrated above clearly explains the procedure of the F-test. The decision must be taken as to whether or not two measurement series with a scope of $n_1$ and $n_2$ originate from two normally distributed populations with the same variance (the means do not need to be known). The null hypothesis therefore assumes that the variances of the populations in question are identical: $\sigma_1^2 = \sigma_2^2$.

Next, from the variances $s_1^2$ and $s_2^2$ of the two series of measurements, the test statistic

$$F = \frac{s_1^2}{s_2^2} \tag{9.33}$$

is calculated and compared with the threshold value of the F-distribution. If the result is $F > F_{n_1-1;n_2-1;0,99}$, i.e. $F$ lies outside the 99% random scatter range, the null hypothesis can be discarded.

*REMARKS:*

- *The alternative hypothesis is $\sigma_1^2 > \sigma_2^2$, it is therefore a one-sided question.*
- *When the larger of the two variances $s_1^2$ and $s_2^2$ is written above the fraction line, $F$ can only assume values greater than 1; the question is therefore two-sided. With a significance level $\alpha$=1%, the threshold value will then be determined to 99.5% based on Table 2 in section 9.3.*

#### 9.1.5.2. Analysis of variance

The t-test can be used to verify whether the mean values of two series of measurements differ significantly. The series of measurements to be compared can each be regarded formally as test results for the two levels 1 (e.g. material $A$) and 2 (material $B$) of a single factor (material). If the single-factor test is expanded to more than two levels (generally $m$ levels), a comparison of the mean values using the t-test is no longer possible. In this case, evaluation can take place by means of an analysis of variance.

Let us now take a look at the following test results from $m$ levels of a factor with $n$ replications per level:

| Level | Results | Mean | Variance |
|---|---|---|---|
| 1 | $y_{11}, y_{12}, \ldots, y_{1n}$ | $\bar{y}_1$ | $s_1^2$ |
| 2 | $y_{21}, y_{22}, \ldots, y_{2n}$ | $\bar{y}_2$ | $s_2^2$ |
| | | | |
| $m$ | $y_{m1}, y_{m2}, \ldots, y_{mn}$ | $\bar{y}_m$ | $s_m^2$ |

**Table 9.1:** Test results from $m$ levels of a factor with $n$ replications per level

If the factor has no influence on the measurement result, all individual results can be viewed as originating from the same population. They - and therefore the means, too - are then subject only to random deviations ("test noise") of the shared mean.

If, on the contrary, the factor has a significant influence on the measurement result, the means of the distributions belonging to the factor levels will differ significantly.

During the analysis of variance, we assume $m$ independent, normally distributed populations with the same variance, and put forward the null hypothesis that all readings originate from populations with the same mean $\mu_1 = \mu_2 = ... = \mu_m = \mu$. Since the precondition for this is identical variances, the null hypothesis means that all readings originate from one and the same population.

The basic objective of the analysis of variance is to estimate the variance of the means in two different ways. Firstly, the variance of the means can be directly calculated from the experimental data:

$$^1s_{\bar{y}}^2 = s_{\bar{y}}^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{y}_i - \bar{\bar{y}}\right)^2 \text{, with } \bar{\bar{y}} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i \quad \text{(overall mean)} \tag{9.34}$$

On the other hand, the variance of the means can be estimated from the variance of the individual values. The latter method produces

$$\overline{s_y^2} = \frac{1}{m}\sum_{i=1}^{m}s_i^2 \tag{9.35}$$

as a mean over the variances of all $m$ levels, and acts as a measure of experimental variance. If no true differences exist between the individual levels, the variance of the means is therefore

$$^2s_{\bar{y}}^2 = \frac{\overline{s_y^2}}{n}\,. \tag{9.36}$$

Finally, the test statistic

$$F = \frac{{}^{1}s_{\bar{y}}^2}{{}^{2}s_{\bar{y}}^2} = \frac{n s_{\bar{y}}^2}{\overline{s_y^2}} \tag{9.37}$$

is used to conduct an F-test (comparison of the two estimates), and the null hypothesis formulated above is rejected if $F > F(m\text{-}1;\ m(n\text{-}1);\ 99\%)$. (Further threshold values for $F$ are attached).

The larger the test statistic, the more improbable it is that the effect is random (not significant), and the stronger the implication of a true (significant) difference between the levels. Rejecting the null hypothesis means: a significant difference exists between the means of the measurement results for the factor levels: the factor has a significant influence on the measurement result.

#### 9.1.5.3. Derivation of the test statistic

The term 'analysis of variance' denotes the partitioning of the variance of all readings, using the formulae expressed above, into either random variance ("test noise") or systematic deviation of the means. This partitioning, or analysis, is described below. If $m$ is the number of lines and $n$ is the number of readings (tests) per line, the total variance of all $m*n$ readings is expressed by

$$s^2 = \frac{1}{n \cdot m - 1} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \bar{\bar{y}}\right)^2 . \tag{9.38}$$

The variable $Q = (n*m\text{-}1)s^2$ is the sum of squares (SS).

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \bar{\bar{y}}\right)^2 \tag{9.39}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \bar{y}_j + \bar{y}_j - \bar{\bar{y}}\right)^2 \text{ (extension with zero)} \tag{9.40}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \left(y_{ij} - \bar{y}_j\right)^2 + 2\left(y_{ij} - \bar{y}_j\right)\left(\bar{y}_j - \bar{\bar{y}}\right) + \left(\bar{y}_j - \bar{\bar{y}}\right)^2 \right) \tag{9.41}$$

It is possible to show that the mean term is zero. Thus:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \bar{y}_j\right)^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\bar{y}_j - \bar{\bar{y}}\right)^2 \tag{9.42}$$

$$Q = \sum_{j=1}^{m} (n-1) s_j^2 + \sum_{i=1}^{n} (m-1)\, s_{\bar{y}}^2 \tag{9.43}$$

$$(n \cdot m - 1) s^2 = m(n-1)\overline{s_y^2} + n\,(m-1) s_{\bar{y}}^2 \tag{9.44}$$

$$Q = Q_R + Q_A \tag{9.45}$$

"Total variance = test noise + mean variance"

Degree of freedom of $Q_R$: $f_R = m(n-1)$ (9.46)

Degree of freedom of $Q_A$: $f_A = m - 1$ (9.47)

Degree of freedom of $Q$: $f = n \cdot m - 1$ (9.48)

Statement of degrees of freedom:

$$f = f_A + f_R \tag{9.49}$$

$$n \cdot m - 1 = m - 1 + m(n-1) \tag{9.50}$$

$$n \cdot m - 1 = n \cdot m - 1 \tag{9.51}$$

Test statistic:

$$F = \frac{\frac{Q_A}{f_A}}{\frac{Q_R}{f_R}} = \frac{\frac{n(m-1)}{m-1} s_{\bar{y}}^2}{\frac{m(n-1)}{m(n-1)} \overline{s_y^2}} = \frac{n s_{\bar{y}}^2}{\overline{s_y^2}} \tag{9.52}$$

### 9.1.6. Regression analysis

In the case of quantitative factors on $n$ levels, the task at hand is to adapt a predefined model approach to the test results in the optimum manner through regression. Let us assume, for example, that we have $N$ tests and hence $N$ values of the response variable $y$ at $N$ values of a factor $x$:

$$(y_i, x_i)\,;\; i = 1...N\,. \tag{9.53}$$

For $y = f(x)$, we suspect a linear relationship, from which the distributed response variable deviates only minimally and randomly by the amount $\varepsilon$:

$$y_i = b_0 + b_1 x_i + \varepsilon_i\,. \tag{9.54}$$

During regression, it is now a matter of calculating the coefficients $b_0$ and $b_1$ in such a way as to enable the best possible prediction of the response variable. With

$$\hat{y}_i = b_0 + b_1 x_i \tag{9.55}$$

we express the value of the response variable predicted by the model at the point $x_i$. The “quality” of the prediction is assured using the familiar least-squares method, by minimizing the sum of squares between the test value and the prediction at all $N$ points $x_i$:

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - (b_0 + b_1 x_i))^2 \rightarrow \min\,. \tag{9.56}$$

The necessary condition for minimization of the above function is that the first partial derivations in terms of $b_0$ and $b_1$ are zero. On this basis, the coefficients can be calculated as follows (see [22], p.158):

$$b_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{\bar{x}})(y_i - \bar{\bar{y}})}{\sum_{i=1}^{N} (x_i - \bar{\bar{x}})^2}, \quad b_0 = \bar{\bar{y}} - b_1 \bar{\bar{x}}\,, \tag{9.57}$$

with the designations

$$\bar{\bar{x}} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad \bar{\bar{y}} = \frac{1}{N} \sum_{i=1}^{N} y_i\,. \tag{9.58}$$

Under the precondition that

- the response variable is normally distributed for each setting of the predictor variables,
- a linear relationship exists between the predictor variables and the mean of the response variables, $\mu_y(x) = a_0 + a_1 x$,
- the deviations between this linear relationship are only random in nature, and
- the standard deviation characterizing this random variance is constant throughout the experimental area,

we can demonstrate that the estimates for the coefficients of the model equation, calculated using a linear regression, are unbiased, i.e. their means over numerous tests concur with the actual values, confidence intervals can be expressed and therefore conclusions about the significance of the coefficients can be made, see [1], p.171ff.

To rate the quality of the adaptation, the sum of squares (SS) of the $y$ values can be partitioned as follows:

$$\sum_{i=1}^{N}(y_i-\bar{\bar{y}})^2=\sum_{i=1}^{N}(\hat{y}_i-\bar{\bar{y}})^2+\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$$
$$Q=Q_{\text{Regr}}+Q_{\text{Resid}} \tag{9.59}$$

The first term can be explained by the regression line, the second term is assigned to residuals. As mentioned above, the least-squares method entails minimizing the second term. The RMS error (root mean square error)

$$s_R=\sqrt{\frac{Q_{\text{Resid}}}{f_{\text{Resid}}}}\,, \tag{9.60}$$

also referred to as the standard error or standard deviation of regression, is therefore suitable for characterizing the result of this minimization. The degree of freedom $f_{\text{Resid}} = N - p$ depends upon the number $p$ of free coefficients (2 with a linear scenario).

The quality of the adaptation increases the greater the proportion that is explained by the regression. It therefore stands to reason that the following quotients can be used as a measure of quality:

$$R^2=\frac{Q_{\text{Regr}}}{Q}\,. \tag{9.61}$$

This variable is the coefficient of determination, and assumes values between 0 and 1. Values close to 1 indicate a high proportion of total variance that is explained by the regression. Alternatively, the so-called correlation coefficient is another measure that can be used:

$$R=\text{sgn}(b_1)\cdot\sqrt{R^2}\,. \tag{9.62}$$

The correlation coefficient assumes values between -1 and +1.

The question as to whether the proportion of total variance that is explained by the regression is significantly larger than the proportion assigned to residuals can be analyzed by means of an F-test with the test statistic

$$F_{\text{RegrT}}=\frac{Q_{\text{Regr}}/f_{\text{Regr}}}{Q_{\text{Resid}}/f_{\text{Resid}}}=\frac{Q_{\text{Regr}}/f_{\text{Regr}}}{s_R^2} \tag{9.63}$$

$f_{\text{Regr}} = p - 1$ and $f_{\text{Resid}} = N - p$ denote the degrees of freedom, $p$ is the number of free model parameters. The rating takes place as with every F-test: if $F_{\text{RegrT}} > F(f_{\text{Regr}}; f_{\text{Resid}}; 99\%)$, the null hypothesis “The proportion of variation that is explained by the regression is as high as the proportion of variation of the residuals” is rejected, with the result that the regression significantly explains part of the variation. Threshold values for $F$ are attached.

The term $Q_{\text{Resid}}$, as mentioned above, is assigned to the residuals, i.e. that part of $y$ values in the SS that is not explained by the regression. The question we are facing is basically that of whether this proportion is due to shortcomings on the part of the model, or constitutes pure experimental variance. If the first hypothesis is true, further reduction of the residual variance can be achieved by adding more (higher) terms in the model. If the second case is correct, improvement of the model would only be possible if so far unknown noise variables are taken into consideration. In order to decide which of the above is correct, several replications of one point of the experimental area are required, at least, in order that the pure error can be estimated. We are assuming that the $N$ tests were conducted in such a way that the $i^{\text{th}}$ of $m$ treatments were repeated $n_i$ times. We can now partition $Q_{\text{Resid}}$ as follows:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2=\sum_{i=1}^{m}n_i(\bar{y}_i-\hat{y}_i)^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$$
$$Q_{\text{Resid}}=Q_{\text{LoF}}+Q_{\text{PE}} \tag{9.64}$$

The first proportion $Q_{\text{LoF}}$ corresponds to the weighted SS of the deviation of the group mean of all replications in the $i^{\text{th}}$ treatment from the predicted value. It is referred to as “lack of fit”. The second proportion $Q_{\text{PE}}$ corresponds to the weighted SS of the deviation of the individual values from the respective group mean of all replications in the $i^{\text{th}}$ treatment. This term is not dependent upon the value predicted on the basis of the existing model, and is therefore known as a “pure error”. The relationships are graphically illustrated in Figure 9.5:

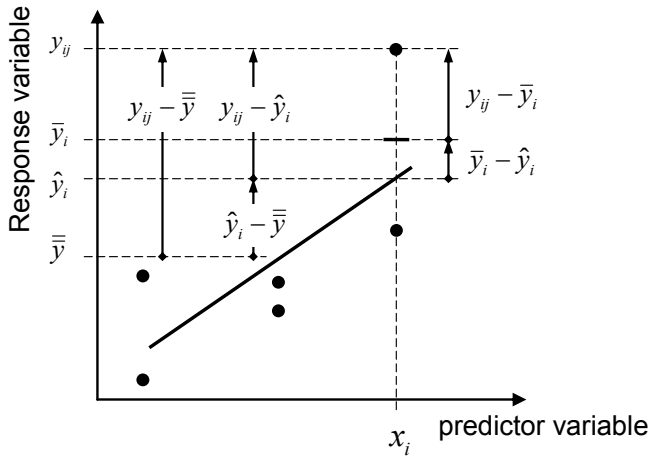



**Figure 9.5:** Components in a regression partition of the SS (schematic)

The quality of the model in terms of whether the approach used has a sufficiently high order (but not in terms of whether all relevant factors have been taken into consideration!) can be analyzed by means of an F-test with the test statistic

$$F_{LoFT} = \frac{Q_{\mathrm{LoF}} / f_{LoF}}{Q_{\mathrm{PE}} / f_{PE}} \tag{9.65}$$

This test is also referred to as the lack-of-fit test. $f_{\mathrm{LoF}} = m - p$ and $f_{\mathrm{PE}} = N - m$ denote the degrees of freedom, $p$ is the number of free model parameters. If $F_{\mathrm{LoFT}} > F(f_{\mathrm{LoF}}; f_{\mathrm{PE}}; 99\%)$, the null hypothesis "Variance due to lack of fit and pure error are the same" is rejected, with the result that the lack of fit is significantly greater than pure experimental variance (threshold values for $F$ are attached).

| | Degree of freedom | SS | Variance | F-value | p-value |
|---|---|---|---|---|---|
| Regression | $f_{\mathrm{Regr}}$ | $Q_{\mathrm{Regr}}$ | $s^2_{\mathrm{Regr}} = \frac{Q_{\mathrm{Regr}}}{f_{\mathrm{Regr}}}$ | $F_{\mathrm{RegrT}} = \frac{s^2_{\mathrm{Regr}}}{s^2_{\mathrm{R}}}$ | $p_{\mathrm{Regr}}$ |
| Residuals | $f_{\mathrm{Resid}} = f_{\mathrm{LoF}} + f_{\mathrm{PE}}$ | $Q_{\mathrm{Resid}} = Q_{\mathrm{LoF}} + Q_{\mathrm{PE}}$ | $s^2_{\mathrm{R}} = \frac{Q_{\mathrm{Resid}}}{f_{\mathrm{Resid}}}$ | | |
| Lack of fit | $f_{\mathrm{LoF}}$ | $Q_{\mathrm{LoF}}$ | $s^2_{\mathrm{LoF}} = \frac{Q_{\mathrm{LoF}}}{f_{\mathrm{LoF}}}$ | $F_{\mathrm{LoFT}} = \frac{s^2_{\mathrm{LoF}}}{s^2_{\mathrm{PE}}}$ | $p_{\mathrm{LoFT}}$ |
| Pure error | $f_{\mathrm{PE}}$ | $Q_{\mathrm{PE}}$ | $s^2_{\mathrm{PE}} = \frac{Q_{\mathrm{PE}}}{f_{\mathrm{PE}}}$ | | |
| Sum total | $f = N - 1 = f_{\mathrm{Regr}} + f_{\mathrm{Resid}}$ | $Q = Q_{\mathrm{Regr}} + Q_{\mathrm{Resid}}$ | $s^2 = s^2_{\mathrm{Regr}} + s^2_R$ | | |

**Table 9.2:** ANOVA regression table

With some statistical tools, a corresponding variable $p_{\mathrm{Regr}}$ or $p_{\mathrm{LoFT}}$ is used in addition to the test statistic $F_{\mathrm{LoFT}}$ or $F_{\mathrm{RegrT}}$ in order to assess significance, whereby

$$F_{\mathrm{RegrT}} = F(1 - p_{\mathrm{RegrT}}; f_1; f_2) \text{ or } F_{\mathrm{LoFT}} = F(1 - p_{\mathrm{LoFT}}; f_1; f_2)\,. \tag{9.66}$$

This value can be interpreted as the "probability of error" (type I error): the smaller this probability, the more significant the regression or the lack of fit. The assessment progresses in the same way as the test statistic: a significant result is present if, for example, $p_{\mathrm{Regr}} \leq 1\%$, a highly significant result if $p_{\mathrm{Regr}} \leq 0.1\%$.

The variables described are typically set out in a table. As with the table for the analysis of variance, this table is referred to as a (global) ANOVA table of the regression model, see Table 9.2.

## 9.2. Software tools

If modeling is to be carried out with several factors on the basis of experimental data, the formulation of designs and most evaluations is time-consuming and, for many users, simply insurmountable without the help of software. For users without much background knowledge, in particular, a tabular calculation program such as MS Excel will not necessarily be of further assistance. Some users will also be in need of software support for linear regression, or for an analysis of variance with one factor. For this reason, the use of suitable software is therefore universally recommended for experiment design and evaluation. In general, the following tools are worthy of recommendation:

- **Minitab®**
  This is a standard solution for statistical evaluations in the context of Six Sigma initiatives, which also features DoE functionalities and can be employed to create and evaluate DoE experimental designs. This solution is especially suitable for occasional DoE tasks, and users that already have extensive experience of this tool. A company license for Minitab is available.

  Further information can be found on the internet at http://www.additive-net.de/software/minitab

- **Cornerstone®**
  This is a software tool that specializes in DoE, and enables even more extensive DoE experiments to be designed and evaluated. Experienced users, above all, can benefit from some of this software's editing and display functions, which have been optimized with DoE in mind and give a tremendous help when handling real experimental data. A company license is not available.

  Further information and a test license for one month can be found on the internet at http://www.versuchsplanung.de/Versuchsplanung/Cornerstone

- **ASCMO®**
  This is a DoE tool that has been developed at Bosch since 2004. A special feature of this software is that the model used is based not on polynomials but on functions with local supports, which can be very useful for response-surface modelling and detailed optimization of complex behaviour. The software is available as a MATLAB and a standalone version. Operation is interactive via a user interface (including helpful editing and display functions like Cornerstone), or via scripts in batch mode. This software is currently free of charge for internal Bosch users.

  Further information is available from the PJ-ASCTeam-Mailbox@de.bosch.com

- **Optislang®**
  This is a software tool for sensitivity studies, optimization and robustness tests in the field of CAE, which also features DoE functionalities, with an emphasis on computer experiments. This tool has interfaces to many numerical simulation tools, which can be automatically activated as solvers for deterministic or stochastic optimizations. A company license for Optislang is available.

  Further information can be found on the internet at http://www.dynardo.de/software/optislang/

### 9.3. Tables

Table 1:

Threshold values of the t-distribution (two-sided)

| | Data reliability | | |
|---|---|---|---|
| $f$ | 95% | 99% | 99,9% |
| 1 | 12.7 | 63.7 | 636.6 |
| 2 | 4.3 | 9.93 | 31.6 |
| 3 | 3.18 | 5.84 | 12.9 |
| 4 | 2.78 | 4.60 | 8.61 |
| 5 | 2.57 | 4.03 | 6.87 |
| 6 | 2.45 | 3.71 | 5.96 |
| 7 | 2.37 | 3.50 | 5.41 |
| 8 | 2.31 | 3.36 | 5.04 |
| 9 | 2.26 | 3.25 | 4.78 |
| 10 | 2.23 | 3.17 | 4.59 |
| 11 | 2.20 | 3.11 | 4.44 |
| 12 | 2.18 | 3.06 | 4.32 |
| 13 | 2.16 | 3.01 | 4.22 |
| 14 | 2.15 | 2.98 | 4.14 |
| 15 | 2.13 | 2.95 | 4.07 |
| 16 | 2.12 | 2.92 | 4.02 |
| 17 | 2.11 | 2.90 | 3.97 |
| 18 | 2.10 | 2.88 | 3.92 |
| 19 | 2.09 | 2.86 | 3.88 |
| 20 | 2.09 | 2.85 | 3.85 |
| 25 | 2.06 | 2.79 | 3.73 |
| 30 | 2.04 | 2.75 | 3.65 |
| 35 | 2.03 | 2.72 | 3.59 |
| 40 | 2.02 | 2.70 | 3.55 |
| 45 | 2.01 | 2.69 | 3.52 |
| 50 | 2.01 | 2.68 | 3.50 |
| 100 | 1.98 | 2.63 | 3.39 |
| 200 | 1.97 | 2.60 | 3.34 |
| 300 | 1.97 | 2.59 | 3.32 |
| 400 | 1.97 | 2.59 | 3.32 |
| 500 | 1.97 | 2.59 | 3.31 |
| $\infty$ | 1.96 | 2.58 | 3.30 |

Table 2:

Threshold values of the F-distribution (one-sided to the value 95%)

| $f_2$ | $f_1=1$ | $f_1=2$ | $f_1=3$ | $f_1=4$ | $f_1=5$ | $f_1=6$ | $f_1=7$ | $f_1=8$ | $f_1=9$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 |
| 2 | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 |
| 3 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 |
| 32 | 4.15 | 3.30 | 2.90 | 2.67 | 2.51 | 2.40 | 2.31 | 2.24 | 2.19 |
| 34 | 4.13 | 3.28 | 2.88 | 2.65 | 2.49 | 2.38 | 2.29 | 2.23 | 2.17 |
| 36 | 4.11 | 3.26 | 2.87 | 2.63 | 2.48 | 2.36 | 2.28 | 2.21 | 2.15 |
| 38 | 4.10 | 3.24 | 2.85 | 2.62 | 2.46 | 2.35 | 2.26 | 2.19 | 2.14 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 |
| 50 | 4.03 | 3.18 | 2.79 | 2.56 | 2.40 | 2.29 | 2.20 | 2.13 | 2.07 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 |
| 70 | 3.98 | 3.13 | 2.74 | 2.50 | 2.35 | 2.23 | 2.14 | 2.07 | 2.02 |
| 80 | 3.96 | 3.11 | 2.72 | 2.49 | 2.33 | 2.21 | 2.13 | 2.06 | 2.00 |
| 90 | 3.95 | 3.10 | 2.71 | 2.47 | 2.32 | 2.20 | 2.11 | 2.04 | 1.99 |
| 100 | 3.94 | 3.09 | 2.70 | 2.46 | 2.31 | 2.19 | 2.10 | 2.03 | 1.97 |
| 150 | 3.90 | 3.06 | 2.66 | 2.43 | 2.27 | 2.16 | 2.07 | 2.00 | 1.94 |
| 200 | 3.89 | 3.04 | 2.65 | 2.42 | 2.26 | 2.14 | 2.06 | 1.98 | 1.93 |
| 1000 | 3.85 | 3.00 | 2.61 | 2.38 | 2.22 | 2.11 | 2.02 | 1.95 | 1.89 |

Table 2 (continued)

Threshold values of the F-distribution (one-sided to the value 95%)

| $f_2$ | $f_1=10$ | $f_1=15$ | $f_1=20$ | $f_1=30$ | $f_1=40$ | $f_1=50$ | $f_1=100$ | $f_1\to\infty$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 242 | 246 | 248 | 250 | 251 | 252 | 253 | 254 |
| 2 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 |
| 3 | 8.79 | 8.70 | 8.66 | 8.62 | 8.59 | 8.58 | 8.55 | 8.53 |
| 4 | 5.96 | 5.86 | 5.80 | 5.75 | 5.72 | 5.70 | 5.66 | 5.63 |
| 5 | 4.74 | 4.62 | 4.56 | 4.50 | 4.46 | 4.44 | 4.41 | 4.37 |
| 6 | 4.06 | 3.94 | 3.87 | 3.81 | 3.77 | 3.75 | 3.71 | 3.67 |
| 7 | 3.64 | 3.51 | 3.44 | 3.38 | 3.34 | 3.32 | 3.27 | 3.23 |
| 8 | 3.35 | 3.22 | 3.15 | 3.08 | 3.04 | 3.02 | 2.97 | 2.93 |
| 9 | 3.14 | 3.01 | 2.94 | 2.86 | 2.83 | 2.80 | 2.76 | 2.71 |
| 10 | 2.98 | 2.85 | 2.77 | 2.70 | 2.66 | 2.64 | 2.59 | 2.54 |
| 11 | 2.85 | 2.72 | 2.65 | 2.57 | 2.53 | 2.51 | 2.46 | 2.40 |
| 12 | 2.75 | 2.62 | 2.54 | 2.47 | 2.43 | 2.40 | 2.35 | 2.30 |
| 13 | 2.67 | 2.53 | 2.46 | 2.38 | 2.34 | 2.31 | 2.26 | 2.21 |
| 14 | 2.60 | 2.46 | 2.39 | 2.31 | 2.27 | 2.24 | 2.19 | 2.13 |
| 15 | 2.54 | 2.40 | 2.33 | 2.25 | 2.20 | 2.18 | 2.12 | 2.07 |
| 16 | 2.49 | 2.35 | 2.28 | 2.19 | 2.15 | 2.12 | 2.07 | 2.01 |
| 17 | 2.45 | 2.31 | 2.23 | 2.15 | 2.10 | 2.08 | 2.02 | 1.96 |
| 18 | 2.41 | 2.27 | 2.19 | 2.11 | 2.06 | 2.04 | 1.98 | 1.92 |
| 19 | 2.38 | 2.23 | 2.16 | 2.07 | 2.03 | 2.00 | 1.94 | 1.88 |
| 20 | 2.35 | 2.20 | 2.12 | 2.04 | 1.99 | 1.97 | 1.91 | 1.84 |
| 22 | 2.30 | 2.15 | 2.07 | 1.98 | 1.94 | 1.91 | 1.85 | 1.78 |
| 24 | 2.25 | 2.11 | 2.03 | 1.94 | 1.89 | 1.86 | 1.80 | 1.73 |
| 26 | 2.22 | 2.07 | 1.99 | 1.90 | 1.85 | 1.82 | 1.76 | 1.69 |
| 28 | 2.19 | 2.04 | 1.96 | 1.87 | 1.82 | 1.79 | 1.73 | 1.65 |
| 30 | 2.16 | 2.01 | 1.93 | 1.84 | 1.79 | 1.76 | 1.70 | 1.62 |
| 32 | 2.14 | 1.99 | 1.91 | 1.82 | 1.77 | 1.74 | 1.67 | 1.59 |
| 34 | 2.12 | 1.97 | 1.89 | 1.80 | 1.75 | 1.71 | 1.65 | 1.57 |
| 36 | 2.11 | 1.95 | 1.87 | 1.78 | 1.73 | 1.69 | 1.62 | 1.55 |
| 38 | 2.09 | 1.94 | 1.85 | 1.76 | 1.71 | 1.68 | 1.61 | 1.53 |
| 40 | 2.08 | 1.92 | 1.84 | 1.74 | 1.69 | 1.66 | 1.59 | 1.51 |
| 50 | 2.03 | 1.87 | 1.78 | 1.69 | 1.63 | 1.60 | 1.52 | 1.44 |
| 60 | 1.99 | 1.84 | 1.75 | 1.65 | 1.59 | 1.56 | 1.48 | 1.39 |
| 70 | 1.97 | 1.81 | 1.72 | 1.62 | 1.57 | 1.53 | 1.45 | 1.35 |
| 80 | 1.95 | 1.79 | 1.70 | 1.60 | 1.54 | 1.51 | 1.43 | 1.32 |
| 90 | 1.94 | 1.78 | 1.69 | 1.59 | 1.53 | 1.49 | 1.41 | 1.30 |
| 100 | 1.93 | 1.77 | 1.68 | 1.57 | 1.52 | 1.48 | 1.39 | 1.28 |
| 150 | 1.89 | 1.73 | 1.64 | 1.53 | 1.48 | 1.44 | 1.34 | 1.22 |
| 200 | 1.88 | 1.72 | 1.62 | 1.52 | 1.46 | 1.41 | 1.32 | 1.19 |
| 1000 | 1.84 | 1.68 | 1.58 | 1.47 | 1.41 | 1.36 | 1.26 | 1.08 |

Table 2 (continued)

Threshold values of the F-distribution (one-sided to the value 99%)

| $f_2$ | $f_1=1$ | $f_1=2$ | $f_1=3$ | $f_1=4$ | $f_1=5$ | $f_1=6$ | $f_1=7$ | $f_1=8$ | $f_1=9$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4052 | 4999 | 5403 | 5625 | 5764 | 5859 | 5928 | 5982 | 6022 |
| 2 | 98.5 | 99.0 | 99.2 | 99.3 | 99.3 | 99.3 | 99.4 | 99.4 | 99.4 |
| 3 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 | 27.5 | 27.3 |
| 4 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 | 14.8 | 14.7 |
| 5 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 | 10.3 | 10.2 |
| 6 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 |
| 7 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 |
| 8 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 |
| 9 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 |
| 10 | 10.0 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.70 | 4.46 | 4.28 | 4.14 | 4.03 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 |
| 18 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 |
| 22 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 |
| 26 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 |
| 28 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 |
| 32 | 7.50 | 5.34 | 4.46 | 3.97 | 3.65 | 3.43 | 3.26 | 3.13 | 3.02 |
| 34 | 7.44 | 5.29 | 4.42 | 3.93 | 3.61 | 3.39 | 3.22 | 3.09 | 2.98 |
| 36 | 7.40 | 5.25 | 4.38 | 3.89 | 3.57 | 3.35 | 3.18 | 3.05 | 2.95 |
| 38 | 7.35 | 5.21 | 4.34 | 3.86 | 3.54 | 3.32 | 3.15 | 3.02 | 2.92 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 |
| 50 | 7.17 | 5.06 | 4.20 | 3.72 | 3.41 | 3.19 | 3.02 | 2.89 | 2.79 |
| 60 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 |
| 70 | 7.01 | 4.92 | 4.08 | 3.60 | 3.29 | 3.07 | 2.91 | 2.78 | 2.67 |
| 80 | 6.96 | 4.88 | 4.04 | 3.56 | 3.26 | 3.04 | 2.87 | 2.74 | 2.64 |
| 90 | 6.93 | 4.85 | 4.01 | 3.54 | 3.23 | 3.01 | 2.84 | 2.72 | 2.61 |
| 100 | 6.90 | 4.82 | 3.98 | 3.51 | 3.21 | 2.99 | 2.82 | 2.69 | 2.59 |
| 150 | 6.81 | 4.75 | 3.92 | 3.45 | 3.14 | 2.92 | 2.76 | 2.63 | 2.53 |
| 200 | 6.76 | 4.71 | 3.88 | 3.41 | 3.11 | 2.89 | 2.73 | 2.60 | 2.50 |
| 1000 | 6.66 | 4.63 | 3.80 | 3.34 | 3.04 | 2.82 | 2.66 | 2.53 | 2.43 |

Table 2 (continued)

Threshold values of the F-distribution (one-sided to the value 99%)

| $f_2$ | $f_1=10$ | $f_1=15$ | $f_1=20$ | $f_1=30$ | $f_1=40$ | $f_1=50$ | $f_1=100$ | $f_1\to\infty$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 6056 | 6157 | 6209 | 6261 | 6287 | 6300 | 6330 | 6366 |
| 2 | 99.4 | 99.4 | 99.4 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 |
| 3 | 27.2 | 26.9 | 26.7 | 26.5 | 26.4 | 26.4 | 26.2 | 26.1 |
| 4 | 14.5 | 14.2 | 14.0 | 13.8 | 13.7 | 13.7 | 13.6 | 13.5 |
| 5 | 10.1 | 9.72 | 9.55 | 9.38 | 9.29 | 9.24 | 9.13 | 9.02 |
| 6 | 7.87 | 7.56 | 7.40 | 7.23 | 7.14 | 7.09 | 6.99 | 6.88 |
| 7 | 6.62 | 6.31 | 6.16 | 5.99 | 5.91 | 5.86 | 5.75 | 5.65 |
| 8 | 5.81 | 5.52 | 5.36 | 5.20 | 5.12 | 5.07 | 4.96 | 4.86 |
| 9 | 5.26 | 4.96 | 4.81 | 4.65 | 4.57 | 4.52 | 4.42 | 4.31 |
| 10 | 4.85 | 4.56 | 4.41 | 4.25 | 4.17 | 4.12 | 4.01 | 3.91 |
| 11 | 4.54 | 4.25 | 4.10 | 3.94 | 3.86 | 3.81 | 3.71 | 3.60 |
| 12 | 4.30 | 4.01 | 3.86 | 3.70 | 3.62 | 3.57 | 3.47 | 3.36 |
| 13 | 4.10 | 3.82 | 3.66 | 3.51 | 3.43 | 3.38 | 3.27 | 3.17 |
| 14 | 3.94 | 3.66 | 3.51 | 3.35 | 3.27 | 3.22 | 3.11 | 3.00 |
| 15 | 3.80 | 3.52 | 3.37 | 3.21 | 3.13 | 3.08 | 2.98 | 2.87 |
| 16 | 3.69 | 3.41 | 3.26 | 3.10 | 3.02 | 2.97 | 2.86 | 2.75 |
| 17 | 3.59 | 3.31 | 3.16 | 3.00 | 2.92 | 2.87 | 2.76 | 2.65 |
| 18 | 3.51 | 3.23 | 3.08 | 2.92 | 2.84 | 2.78 | 2.68 | 2.57 |
| 19 | 3.43 | 3.15 | 3.00 | 2.84 | 2.76 | 2.71 | 2.60 | 2.49 |
| 20 | 3.37 | 3.09 | 2.94 | 2.78 | 2.69 | 2.64 | 2.54 | 2.42 |
| 22 | 3.26 | 2.98 | 2.83 | 2.67 | 2.58 | 2.53 | 2.42 | 2.31 |
| 24 | 3.17 | 2.89 | 2.74 | 2.58 | 2.49 | 2.44 | 2.33 | 2.21 |
| 26 | 3.09 | 2.82 | 2.66 | 2.50 | 2.42 | 2.36 | 2.25 | 2.13 |
| 28 | 3.03 | 2.75 | 2.60 | 2.44 | 2.35 | 2.30 | 2.19 | 2.06 |
| 30 | 2.98 | 2.70 | 2.55 | 2.39 | 2.30 | 2.25 | 2.13 | 2.01 |
| 32 | 2.93 | 2.66 | 2.50 | 2.34 | 2.25 | 2.20 | 2.08 | 1.96 |
| 34 | 2.89 | 2.62 | 2.46 | 2.30 | 2.21 | 2.16 | 2.04 | 1.91 |
| 36 | 2.86 | 2.58 | 2.43 | 2.26 | 2.17 | 2.12 | 2.00 | 1.87 |
| 38 | 2.83 | 2.55 | 2.40 | 2.23 | 2.14 | 2.09 | 1.97 | 1.84 |
| 40 | 2.80 | 2.52 | 2.37 | 2.20 | 2.11 | 2.06 | 1.94 | 1.80 |
| 50 | 2.70 | 2.42 | 2.27 | 2.10 | 2.01 | 1.95 | 1.82 | 1.68 |
| 60 | 2.63 | 2.35 | 2.20 | 2.03 | 1.94 | 1.88 | 1.75 | 1.60 |
| 70 | 2.59 | 2.31 | 2.15 | 1.98 | 1.89 | 1.83 | 1.70 | 1.54 |
| 80 | 2.55 | 2.27 | 2.12 | 1.94 | 1.85 | 1.79 | 1.66 | 1.49 |
| 90 | 2.52 | 2.24 | 2.09 | 1.92 | 1.82 | 1.76 | 1.62 | 1.46 |
| 100 | 2.50 | 2.22 | 2.07 | 1.89 | 1.80 | 1.73 | 1.60 | 1.43 |
| 150 | 2.44 | 2.16 | 2.00 | 1.83 | 1.73 | 1.66 | 1.52 | 1.33 |
| 200 | 2.41 | 2.13 | 1.97 | 1.79 | 1.69 | 1.63 | 1.48 | 1.28 |
| 1000 | 2.34 | 2.06 | 1.90 | 1.72 | 1.61 | 1.54 | 1.38 | 1.11 |

## 10. Literature


[1] W. Kleppmann: Taschenbuch Versuchsplanung. Produkte und Prozesse Optimieren, Hanser, Munich, 1998

[2] E. Scheffler: Einführung in die Praxis der statistischen Versuchsplanung, Deutscher Verlag für Grundstoffindustrie, Leipzig, 1986

[3] J. Krottmaier: Versuchsplanung. Der Weg zur Qualität des Jahres 2000, Verlag Industrielle Organisation Zürich, Verlag TÜV Rheinland, Cologne, 1990

[4] H. Petersen: Grundlagen der statistischen Versuchsplanung, Vol.2, ecomed Verlag Landsberg/Lech, 1991

[5] W.G. Cochran, G.M. Cox: Experimental Designs, Wiley, New York, 1957

[6] VDA Reihe Qualitätsmanagement in der Automobilindustrie, Vol. 4 Sicherung der Qualität während der Produktrealisierung – Methoden und Verfahren, Chapter 5: Versuchsmethodik (DoE), 4th Edition 2003

[7] VDA Reihe Qualitätsmanagement in der Automobilindustrie, Vol. 3 “Zuverlässigkeitssicherung bei Automobilherstellern und Lieferanten”, Chapter 2 “Zuverlässigkeitsmethoden und Hilfsmittel”

[8] VDA Reihe Qualitätsmanagement in der Automobilindustrie, Vol. 4 “Sicherung der Qualität während der Produktrealisierung – Methoden und Verfahren”, Chapter 2 “Entwicklungsabläufe”

[9] ISO/TS 16949 “Qualitätsmanagementsysteme - Besondere Anforderungen bei Anwendung von ISO 9001:2000 für die Serien- und Ersatzteilproduktion in der Automobilindustrie”

[10] Chrysler, Ford, GM: QS-9000

[11] R. A. Fisher: The Design of Experiments, Oliver&Boyd, London 1935

[12] K. R. Bhote: Qualität – Der Weg zur Weltspitze, IQM, Grossbottwar, 1990

[13] H. Bossel: Systeme, Dynamik, Simulation. Modellbildung, Analyse und Simulation komplexer Systeme, Books on Demand, Norderstedt, 2004

[14] G. Ropohl: Eine Systemtheorie der Technik, Hanser, Munich, 1979

[15] Qualitätsmanagement in der Bosch-Gruppe, Vol. 8: Messunsicherheit

[16] Qualitätsmanagement in der Bosch-Gruppe, Vol. 10: Fähigkeit von Mess- und Prüfprozessen

[17] E. Haibach: Betriebsfestigkeit, VDI, Düsseldorf, 1989

[18] J. Riedel: Gewichtsoptimierung eines Kreuzfahrtschiffes unter Spannungsrestriktionen, Colloquium of the Bauhaus University of Weimar, 2000 (see www.optislang.de/Projekte/Anlagenbau.htm)

[19] T. Wember: Skript Technische Statistik und statistische Versuchsplanung, self-published

[20] OptiSlang Version 3.0, Theory Manual und User’s Guide, dynardo Dynamic Software and Engineering GmbH, 2009

[21] J. Gamweger, O.Jöbstl, M. Strohmann, W. Suchowerskyj: Design for Six Sigma, Hanser, Munich, 2009

[22] A. Haldar, S. Mahadevan: Probability, Reliability and Statistical Methods in Engineering Design, J. Wiley, 2000

[23] G. Taguchi: Introduction to Quality Engineering; Designing Quality into Products and Processes, Unipub Kraus International Publications and American Supplier Institute, 1986

[24] G. Taguchi: System of experimental design, Vol. 1 and 2, Unipub Kraus International Publications and American Supplier Institute, 1987

[25] T. Bendell (Editor): Taguchi Methods; Proceedings of the 1988 European Conference, Elsevier, 1989

[26] L. Fahrmeir, T. Kneib, S. Lang: Regression. Modelle, Methoden und Anwendungen. 2nd Edition, Springer, 2009

[27] H. Pruscha: Statistisches Methodenbuch, Springer, 2006

## 11. 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**

BVE 17241-1 Stand 10.2010