

# 01NAEX - Lecture 01

## Introduction to Design of Experiments (DOE)

Jiri Franc

Czech Technical University  
Faculty of Nuclear Sciences and Physical Engineering  
Department of Mathematics

## What does DOE means and brings

**Experiment** is a test or a series of tests in which purposeful changes are made to the input variables or factors of a system so that we may observe and identify the reasons for changes in the output response.

**Statistical design of experiments (DOE)** is the process of planning experiments so that appropriate data can be analyzed by statistical methods that results in valid, objective, and meaningful conclusions from the data involves two aspects: design and statistical analysis.

**All experiments are designed experiments, some are poorly designed, some are well-designed.**

## What does DOE means and brings

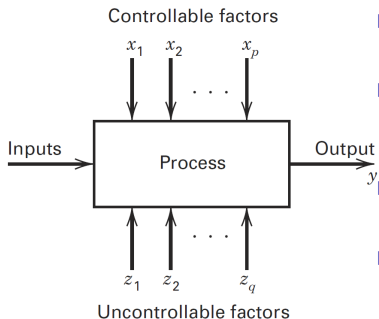
### DOE

- ▶ is a methodology for systematically applying statistics to experimentation.
- ▶ lets experimenters develop a mathematical model that predicts how input variables interact to create output variables or responses in a process or system.
- ▶ can be used for a wide range of experiments for various purposes including nearly all fields of engineering and even in business marketing.

## General model of a process or system

The process or system can be represented by the model, where we can visualize the process as a combination of operations, machines, methods, people, and other resources that transforms some input into an output that has one or more observable response variables.

**The objectives of the experiment may include the following:**



- ▶ Determining which variables are most influential on the response  $y$ .
- ▶ Determining where to set the influential  $x$ 's so that  $y$  is almost always near the desired nominal value.
- ▶ Determining where to set the influential  $x$ 's so that variability in  $y$  is small.
- ▶ Determining where to set the influential  $x$ 's so that the effects of the variables  $z$ 's are minimized.

## Short motivation example

Weighting problem: we want to weight 3 different items A, B, and C.

### Standard weighting procedure - 1

| No. measurement | Notation | Meaning | Model for responses           |
|-----------------|----------|---------|-------------------------------|
| 1               | (1)      | No item | $(1) = \mu + \varepsilon_1$   |
| 2               | a        | item A  | $a = \mu + A + \varepsilon_2$ |
| 3               | b        | item B  | $b = \mu + B + \varepsilon_3$ |
| 4               | c        | item C  | $c = \mu + C + \varepsilon_4$ |

Where  $\mu$  is offset of weighting device,  $A, B, C$  are weights of items A,B,C, and  $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$  are measurement errors with standard deviation  $\sigma_\varepsilon$ .

The estimate of A is given by  $\hat{A}^{(1)} = a - (1) = A + \varepsilon_2 - \varepsilon_1$  and the corresponding for B and C.

## Short motivation example

Weighting problem: we want to weight 3 different items A, B, and C.

### Alternative weighting procedure - 2

| No. measurement | Notation | Meaning      | Model for responses                |
|-----------------|----------|--------------|------------------------------------|
| 1               | (1)      | No item      | $(1) = \mu + \varepsilon_5$        |
| 2               | ac       | item A and C | $ac = \mu + A + C + \varepsilon_6$ |
| 3               | bc       | item B and C | $bc = \mu + B + C + \varepsilon_7$ |
| 4               | ab       | item A and B | $ab = \mu + A + B + \varepsilon_8$ |

The estimate of A is now given by  $\hat{A}^{(2)} = \frac{-(1)+ac-bc+ab}{2} = A + \frac{-\varepsilon_5+\varepsilon_6-\varepsilon_7+\varepsilon_8}{2}$  and the corresponding for B and C.

$$\text{Var} [\hat{A}^{(1)}] = 2\sigma_\varepsilon^2$$

$$\text{Var} [\hat{A}^{(2)}] = \sigma_\varepsilon^2$$

The second design is more precise than the first one, since in the first design not all measurements are used to estimate all parameters, which is the case in the second design.

## Short history of DOE

- ▶ **The agricultural origins, 1920s - 1940s:**
  - ▶ Sir R. A. Fisher & his co-workers.
  - ▶ Profound influence on the use of statistics on agricultural.
  - ▶ Factorial designs, ANOVA.
- ▶ **The first industrial era, 1950s - 1970s:**
  - ▶ Box & Wilson, Response surface methodology (RSM).
  - ▶ Applications in chemical and process industries, finance, etc.
  - ▶ Total quality control (TQC), Statistical process control (SPC).
- ▶ **The second industrial era, 1970s - 1990s:**
  - ▶ Quality control, Japanese style of TQC.
  - ▶ Taguchi and robust parameter design, process robustness.
  - ▶ Total Quality Management (TQM), Continuous Quality Improvement (CQI).
- ▶ **The modern era, 1990s - till today:**
  - ▶ Wide use of computer technology and experiments in DOE.
  - ▶ Expanded use of DOE in Six-Sigma and in business.
  - ▶ Globalization is driving all sectors of the economy to be more competitive.

# The basic principles of design of experiments

## ▶ **Randomization:**

- ▶ Cornerstone underlying the use of statistical methods.
- ▶ Running the trials in an experiment in random order.
- ▶ Allocation of the experimental material is random.
- ▶ Notion of balancing out effects of hidden variables.
- ▶ Reduce bias and systematic errors.

## ▶ **Replication:**

- ▶ An independent repeat of each factor combination.
- ▶ Allows to obtain an estimate of the experimental error.
- ▶ Sample size (improving precision of effect estimation, estimation of error or background noise).
- ▶ Replication reflects sources of variability both between runs and within runs.

## ▶ **Blocking:**

- ▶ It is a design technique used to improve the precision with which comparisons among the factors of interest are made.
- ▶ Reduces or eliminates the variability transmitted from nuisance factors.
- ▶ Increases precision of experiment.



# Strategy of Experimentation

- ▶ **Best-guess approach:**
  - ▶ Wide-spread, trial and error experiments.
  - ▶ More successful than you might suspect.
  - ▶ **Disadvantage:**
    - ▶ cannot guarantee best solution has been found.
- ▶ **One-factor-at-a-time (OFAT) approach:**
  - ▶ Sometimes associated with the scientific or engineering method.
  - ▶ **Disadvantage:**
    - ▶ fails to consider any possible interaction between factors.
    - ▶ very inefficient, requires many test runs.
- ▶ **Statistically designed approach:**
  - ▶ Based on Fisher's factorial concept.
  - ▶ Modern and most efficient approach.
  - ▶ Can determine how factors interact.

# Factorial Designs

## **Treatment x Nuisance factors**

- ▶ Treatment factor - factor of primary interest to us.
- ▶ Nuisance factor - factor which are not our primary focus, but we have to deal with them.

## **Experimental x Classification factors**

- ▶ Experimental factor - we can specify them and set the levels.
- ▶ Classification factor - we can not change or randomly assign them.

## **Quantitative x Qualitative factors**

- ▶ Quantitative factor - we can assign any specified level of them.
- ▶ Qualitative factor - have categories which are different types.

# Planning, Conducting & Analyzing an Experiment

## Outline of Design and Analysis of Experiment:

1. Recognition and statement of problem.
2. Choice of factors, levels, and ranges.
3. Selection of the response variable(s).
4. Choice of experimental design.
5. Performing the experiment.
6. Statistical analysis of the data.
7. Drawing conclusions and recommendations.

Steps 2 and 3 are often done simultaneously or in reverse order.

Steps 1,2, and 3 are called Pre-experimental planning

# Guidelines for Designing Experiments

## 1. Recognition and statement of problem:

- ▶ Obvious, but in practice not so simple point.
- ▶ It is necessary to develop all ideas about the objectives of the experiment, get input from all concerned parties, the team approach.

## 2. Choice of factors, levels, and ranges:

- ▶ The potential design factors are those factors that we wish to vary in the experiment .
- ▶ Classify factors as: design factor, held-constant factor, allow-to-vary factors, nuisance factors.
- ▶ Nuisance factors may have large effects that must be accounted for (controllable, uncontrollable, noise).
- ▶ Need to use engineering judgment or prior test results.

## 3. Selection of the response variable(s):

- ▶ Be sure that the variable really provides useful information about the process under study.
- ▶ Multiple responses are not unusual.

# Guidelines for Designing Experiments

## 4. Choice of experimental design:

- ▶ Involves consideration of sample size (no. of replicates), selection of run order, whether blocking or randomization restriction, software to use etc

## 5. Performing the experiment:

- ▶ Vital to monitor the process carefully
- ▶ Up/for planning is crucial to success
- ▶ Easy to underestimate logistical and planning aspects of running and design experiment in development environment

## 6. Statistical analysis of the data:

- ▶ Should be used to analyze the data so that results and conclusions are objective rather than judgmental in nature
- ▶ Use simple graphics whenever possible

## 7. Drawing conclusions and recommendations:

- ▶ Follow-up test runs and confirmation testing to validate the conclusions from the experiment

## Useful advices in Planning, Conducting & Analyzing an Experiment

Use your **non-statistical knowledge** of the problem, physical laws, background knowledge, etc. It is crucial to success.

**KISS principle** (Keep it simple - Stupid!, Keep it simple and straightforward, Keep it short and simple). Don't use complex, sophisticated statistical techniques. If design is good, If design is good, analysis is relatively straightforward. If design is bad - even the most complex and elegant statistics cannot save the situation.

Think and experiment **sequentially**, pre-experimental planning is vital.

Recognize the difference between **practical and statistical significance**.

## General definitions from probability - probability distribution

Assume, probability triplet  $(\Omega, \mathcal{F}, p)$  and a random variable, say  $y$ .

**Probability distribution:** If  $y$  is discrete, we call the  $p(y)$  the probability mass function of  $y$ . If  $y$  is continuous, we call  $f(y)$  the probability density function for  $y$ .

**Mean, expected value:**

$$\mu = E(y) \begin{cases} \int_{-\infty}^{+\infty} yf(y)dy & y \text{ is continuous random variable} \\ \sum_{\text{all } y} yp(y) & y \text{ is discrete random variable} \end{cases}$$

**Variance, standard deviation:**

$$\sigma^2 = Var(y) \begin{cases} \int_{-\infty}^{+\infty} (y - \mu)^2 f(y) dy & y \text{ is continuous random variable} \\ \sum_{\text{all } y} (y - \mu)^2 p(y) & y \text{ is discrete random variable} \end{cases}$$

## General definitions from probability - Sampling

If the population contains  $N$  elements and a sample of  $n$  of them is to be selected, and if each of the possible samples has an equal probability of being chosen, then the procedure employed is called random sampling.

Suppose that  $y_1, y_2, \dots, y_n$  represents a sample. Then we define two statistics:

**Sample mean:**

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

The sample mean  $\bar{y}$  is a unbiased point estimator of the population mean  $\mu$

**Sample variance:**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

The sample variance  $S^2$  is a unbiased point estimator of the population variance  $\sigma^2$ . Sometimes  $S = \sqrt{S^2}$  is called **sample standard deviation**. The expression

$$SS = \sum_{i=1}^n (y_i - \bar{y})^2$$

is called the corrected **sum of squares** of the observations  $y_i$ .



## General definitions from probability - Degrees of Freedom

Number of degrees of freedom refers to number of independent elements in sum of squares  $SS$ :

$$E(SS) = E \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right] = (n - 1)\sigma^2.$$

The degrees of freedom associated with a sum-of-squares is the degrees-of-freedom of the corresponding component vectors.

Only  $(n - 1)$  elements from the set  $(y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})$  are independent.

## General definitions from probability - Normal distribution, $\chi^2$ distribution

Let  $y$  be a random variable with normal distribution then the probability density function of  $y$  is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}}$$

where  $\mu \in R$  is the mean of the distribution,  $\sigma^2 > 0$ ) is the variance.

If  $x_1, x_2, \dots, x_k$  are *iid* random variables with standard normal distribution then  $y = x_1^2 + x_2^2 + \dots + x_k^2$  is a random variable with  $\chi^2$  distribution with  $k$  degrees of freedom, denoted  $\chi_k^2$  and the probability density function of  $y$  is

$$f(y) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} y^{\frac{k}{2}-1} e^{-\frac{y}{2}}$$

The distribution is skewed with mean  $\mu = k$  and variance  $\sigma^2 = 2k$ .

## General definitions from probability - $t$ distribution

If  $z$  and  $\chi_k^2$  are independent standard normal and chi-squared r.v. then

$$t_k = \frac{z}{\sqrt{\frac{\chi_k^2}{k}}}$$

is r.v. with  $t$  distribution with  $k$  degrees of freedom and the probability density function of  $t_k$  is

$$f(y) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \frac{1}{(\frac{t^2}{k} + 1)^{\frac{k+1}{2}}}$$

The distribution is symmetric with mean  $\mu = 0$  and variance  $\sigma^2 = \frac{k}{k-2}$  for  $k > 2$ . If  $k = \infty$ , the  $t$  distribution becomes the standard normal distribution.

## General definitions from probability - $F$ distribution

If  $\chi_k^2$  and  $\chi_l^2$  are two independent chi-squared r.v. with  $k$  and  $l$  degrees of freedom then the ratio

$$F_{k,l} = \frac{\frac{\chi_k^2}{k}}{\frac{\chi_l^2}{l}}$$

follows the  $F_{k,l}$  distribution.

$F$  distribution is very important for us, suppose we have two independent normal populations with common variance  $\sigma$ .

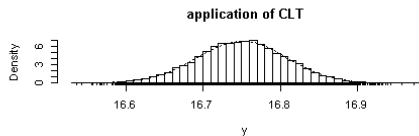
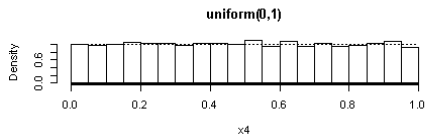
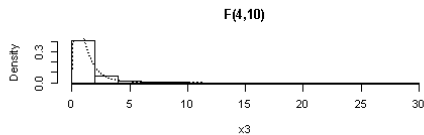
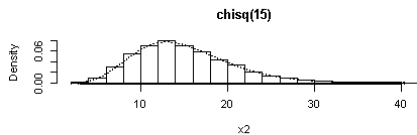
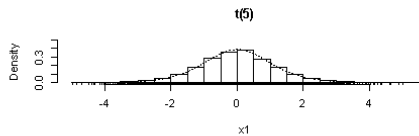
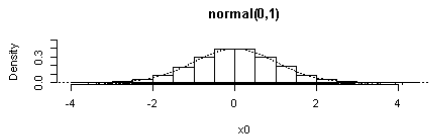
If  $y_{11}, y_{12}, \dots, y_{1n_1}$  is a random sample of  $n_1$  observations from first population with sample variance  $S_1^2$  and  $y_{21}, y_{22}, \dots, y_{2n_2}$  is a random sample of  $n_2$  observations from second population with sample variance  $S_2^2$  then

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-2}.$$

The result comes from following

$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

# General definitions from probability - CLT



## General definitions from probability - Hypothesis testing

A statistical hypothesis is a statement about the parameters of one or more populations.

Suppose that we partition the parameter space  $\Theta$  into two disjoint sets  $\Theta_0$  and  $\Theta_1$  and that we wish to test:

The null hypothesis  $H_0 : \theta \in \Theta_0$       *versus*  
The alternative hypothesis  $H_1 : \theta \in \Theta_1$

To test a hypothesis, we devise a procedure for taking a random sample, computing an appropriate test statistics, and then rejecting or failing to reject the null hypothesis  $H_0$

**Decisions in Hypothesis Testing**

| Decision     | $H_0$ Is True | $H_0$ Is False |
|--------------|---------------|----------------|
| Retain $H_0$ | no error      | type II error  |
| Reject $H_0$ | type I error  | no error       |

**Significance level:**  $\alpha = P(\text{type I error})$

**Power:**  $1 - \beta = 1 - P(\text{type II error}) = P(\text{reject } H_0 \mid H_0 \text{ is false})$

## General definitions from probability - p-value

The p-value is the smallest level of significance that would lead to rejection of the null hypothesis  $H_0$  with the given data.

- ▶ The p-value is the most frequent outcome from the hypothesis testing.
- ▶ If the test is rejected at the significance level  $\alpha$ , it will be also rejected at the level  $\alpha' > \alpha$ .
- ▶ The p-value is a measure of the evidence against  $H_0$  (the smaller the p-value, the stronger evidence against  $H_0$ ).
- ▶ The p-value is not the probability that the null hypothesis is true.
- ▶  $1 - (\text{p-value})$  is not the probability of the alternative hypothesis being true.
- ▶ The significance level of the test is not determined by the p-value.

# General definitions from probability - Test on means with variance known

## Tests on Means with Variance Known

| Hypothesis                                      | Test Statistic   | Fixed Significance Level<br>Criteria for Rejection | P-Value                  |
|---|--|--|--------------------------|
| $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$     | $Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$  | $ Z_0  > Z_{\alpha/2}$                             | $P = 2[1 - \Phi( Z_0 )]$ |
| $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$        |  | $Z_0 < -Z_{\alpha}$                                | $P = \Phi(Z_0)$          |
| $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$        |  | $Z_0 > Z_{\alpha}$                                 | $P = 1 - \Phi(Z_0)$      |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 \neq \mu_2$ |  | $ Z_0  > Z_{\alpha/2}$                             | $P = 2[1 - \Phi( Z_0 )]$ |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 < \mu_2$    |  | $Z_0 < -Z_{\alpha}$                                | $P = \Phi(Z_0)$          |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 > \mu_2$    | $Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | $Z_0 > Z_{\alpha}$                                 | $P = 1 - \Phi(Z_0)$      |



# General definitions from probability - Test on means with variance unknown

## Tests on Means of Normal Distributions, Variance Unknown

| Hypothesis                                      | Test Statistic   | Fixed Significance Level<br>Criteria for Rejection | P-Value   |
|---|--|--|---|
| $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$     |  | $ t_0  > t_{\alpha/2, n-1}$                        | sum of the probability above $t_0$ and below $-t_0$ |
| $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$        | $t_0 = \frac{\bar{y} - \mu_0}{S/\sqrt{n}}$   | $t_0 < -t_{\alpha, n-1}$                           | probability below $t_0$                             |
| $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$        |  | $t_0 > t_{\alpha, n-1}$                            | probability above $t_0$                             |
| <hr/>   |  |  |   |
|   | if $\sigma_1^2 = \sigma_2^2$   |  |   |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 \neq \mu_2$ | $t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$<br>$v = n_1 + n_2 - 2$                            | $ t_0  > t_{\alpha/2, v}$                          | sum of the probability above $t_0$ and below $-t_0$ |
| <hr/>   |  |  |   |
|   | if $\sigma_1^2 \neq \sigma_2^2$  |  |   |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 < \mu_2$    | $t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$   | $t_0 < -t_{\alpha, v}$                             | probability below $t_0$                             |
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 > \mu_2$    | $v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$ | $t_0 > t_{\alpha, v}$                              | probability above $t_0$                             |

# General definitions from probability - Test on Variance of Normal Distributions

## Tests on Variances of Normal Distributions

| Hypothesis  | Test Statistic                           | Fixed Significance Level<br>Criteria for Rejection                              |
|---|--|---|
| $H_0: \sigma^2 = \sigma_0^2$<br>$H_1: \sigma^2 \neq \sigma_0^2$     |  | $\chi_0^2 > \chi_{\alpha/2, n-1}^2$ or<br>$\chi_0^2 < \chi_{1-\alpha/2, n-1}^2$ |
| $H_0: \sigma^2 = \sigma_0^2$<br>$H_1: \sigma^2 < \sigma_0^2$        | $\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$ | $\chi_0^2 < \chi_{1-\alpha, n-1}^2$   |
| $H_0: \sigma^2 = \sigma_0^2$<br>$H_1: \sigma^2 > \sigma_0^2$        |  | $\chi_0^2 > \chi_{\alpha, n-1}^2$   |
| $H_0: \sigma_1^2 = \sigma_2^2$<br>$H_1: \sigma_1^2 \neq \sigma_2^2$ | $F_0 = \frac{S_1^2}{S_2^2}$              | $F_0 > F_{\alpha/2, n_1-1, n_2-1}$ or<br>$F_0 < F_{1-\alpha/2, n_1-1, n_2-1}$   |
| $H_0: \sigma_1^2 = \sigma_2^2$<br>$H_1: \sigma_1^2 < \sigma_2^2$    | $F_0 = \frac{S_2^2}{S_1^2}$              | $F_0 > F_{\alpha, n_2-1, n_1-1}$  |
| $H_0: \sigma_1^2 = \sigma_2^2$<br>$H_1: \sigma_1^2 > \sigma_2^2$    | $F_0 = \frac{S_1^2}{S_2^2}$              | $F_0 > F_{\alpha, n_1-1, n_2-1}$  |

## General definitions from probability - Confidence interval

Confidence interval is a kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate.

To obtain an interval estimate of an unknown parameter  $\theta$ , we need to find two statistics  $L$  (lower limit) and  $U$  (upper limit) such that the probability statement

$$P(L \leq \theta \leq U) = 1 - \alpha$$

is true. The interval  $(L \leq \theta \leq U)$  is called a  $100(1 - \alpha)$  percent confidence interval for parameter  $\theta$ .

## Simple Comparative Experiment

We consider experiments to compare two conditions (also called treatments, levels, factors, etc).

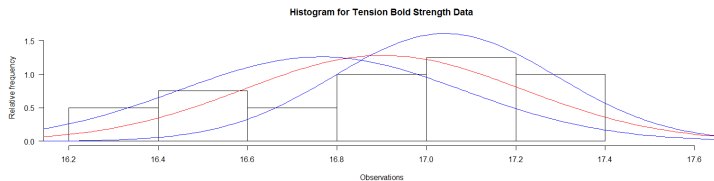
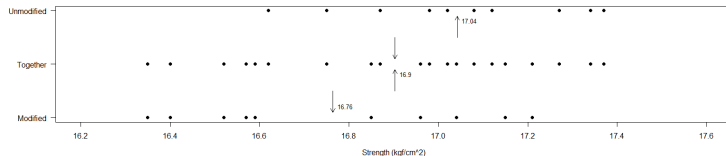
In the simple comparative experiment we use previous mentioned statistical concepts, such as random samples, sampling distributions, hypothesis testing etc.

## Simple Comparative Experiment - data

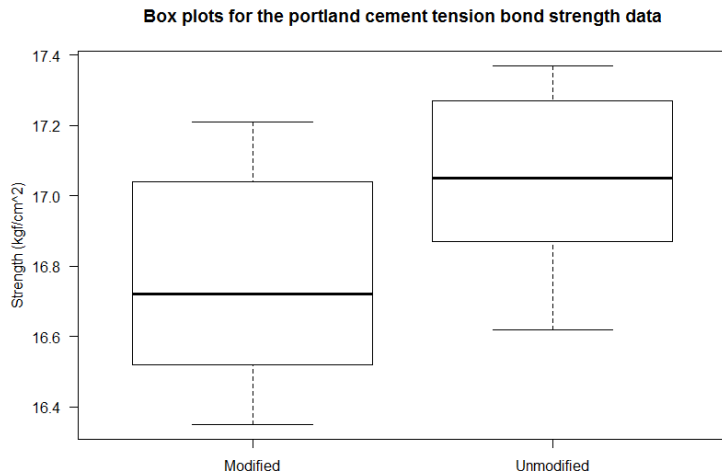
**Tension Bond Strength Data**  
**Portland Cement Data**

| Observation | Modified Mortar | Unmodified Mortar |
|-------------|-----------------|-------------------|
| $j$         | $y_{1j}$        | $y_{2j}$          |
| 1           | 16.57           | 18.15             |
| 2           | 16.59           | 17.96             |
| 3           | 17.15           | 17.90             |
| 4           | 16.96           | 18.22             |
| 5           | 17.04           | 17.75             |
| 6           | 16.52           | 17.86             |
| 7           | 16.35           | 18.00             |
| 8           | 17.21           | 18.25             |
| 9           | 16.40           | 17.63             |
| 10          | 16.85           | 17.50             |

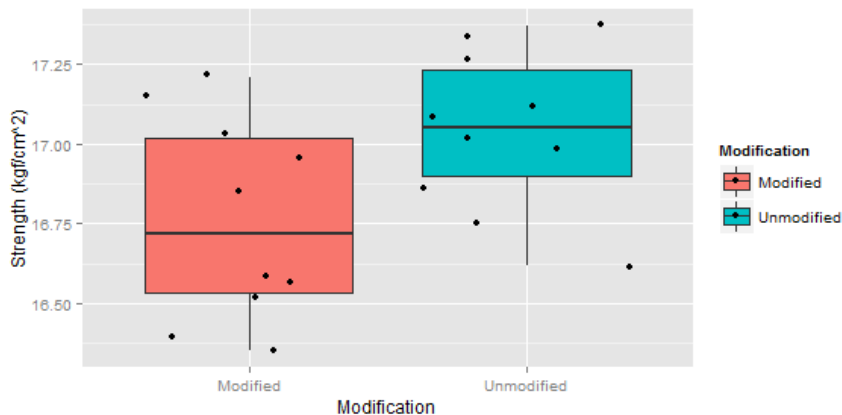
# Simple Comparative Experiment - data plot



## Simple Comparative Experiment - boxplot



## Simple Comparative Experiment - boxplot (from gplot package)





## Simple Comparative Experiment - F-test

If we do not know the variances of the data sets, we have to test, if they are identical or not.

```
var.test(y1, y2, ratio = 1, alternative="two.sided",  
conf.level = 0.95)
```

F test to compare two variances

data: y1 and y2

F = 1.6293, num df = 9, denom df = 9, p-value = 0.4785

alternative hypothesis:

true ratio of variances is not equal to 1

90 percent confidence interval:

0.5125235 5.1792350

95 percent confidence interval:

0.4046845 6.5593806

99 percent confidence interval:

0.2490804 10.6571184

sample estimates:

ratio of variances 1.629257

## Simple Comparative Experiment - The Two-Sample $t$ -test

If we could assume that the variances of tension bond strengths are identical, or the sample  $sd$  are reasonably similar, we can use simple two-sample  $t$ -test to compare means.

The test statistic:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -2.186876,$$

where  $S_p$  is an estimate of common variance  $\sigma^2 = \sigma_1^2 = \sigma_2^2$

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.2842534.$$

Because  $t_0 = -2.186876 < -t_{0.025,18} = -2.100922$ , we would reject the  $H_0$  hypothesis at the significance level  $\alpha = 0.05$ .

## Simple Comparative Experiment - The Two-Sample $t$ -test

Confidence interval on the true difference in means  $\mu_1 - \mu_2$  for the Portland cement problem.

The statistic:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is distributed as  $t_{n_1+n_2-2}$  and thus,

$$P(\bar{y}_1 - \bar{y}_2 - t_{\frac{\alpha}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + t_{\frac{\alpha}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$$

is  $100(1 - \alpha)$  percent confidence interval for  $\mu_1 - \mu_2$ .

For  $\alpha = 0.05$  we obtain  $-0.5450734 \leq \mu_1 - \mu_2 \leq -0.01092661$ .

## Simple Comparative Experiment - The Two-Sample *t*-test

If we could assume that the variances of tension bond strengths are identical, or the sample *sd* are reasonably similar, we can use simple two-sample *t*-test to compare means.

```
> t.test(y1,y2,alternative="two.sided",mu=0,  
paired=F,var.equal=T,conf.level=0.95)
```

Two Sample *t*-test

data: y1 and y2

$t = -2.1869$ ,  $df = 18$ ,  $p\text{-value} = 0.0422$

alternative hypothesis:

true difference in means is not equal to 0

95 percent confidence interval:

-0.54507339 -0.01092661

99 percent confidence interval:

-0.64391308 0.08791308

sample estimates:

mean of x: 16.764, mean of y: 17.042

## Simple Comparative Experiment - t-test

If the variances are not identical, the classical test statistic is not distributed exactly as  $t$  and some approximation of degrees of freedom must be used.

```
> t.test(y1,y2,alternative="two.sided",mu=0,  
paired=FALSE,var.equal=FALSE,conf.level=0.95)
```

Welch Two Sample t-test

data: y1 and y2

$t = -2.1869$ ,  $df = 17.025$ ,  $p\text{-value} = 0.043$

alternative hypothesis:

true difference in means is not equal to 0

95 percent confidence interval:

-0.546174139 -0.009825861

99 percent confidence interval:

-0.64636235 0.09036235

sample estimates:

mean of x: 16.764, mean of y: 17.042

## Simple Comparative Experiment - checking assumptions

In using classical t-test we make following assumptions:

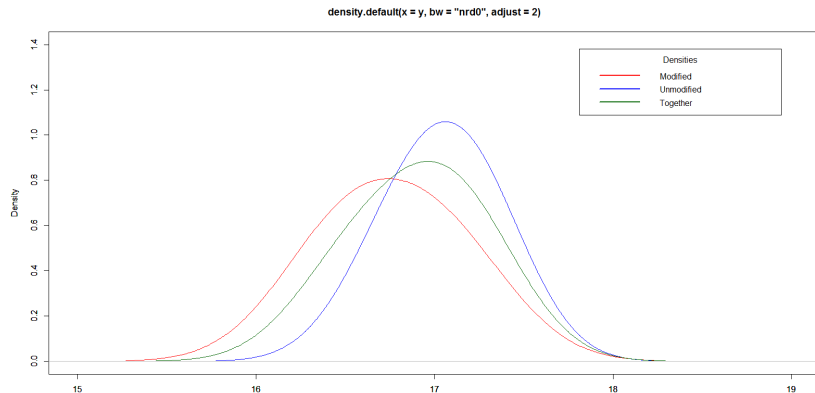
- ▶ observations are independent random variables.
- ▶ both samples are drawn from independent populations that are **normal distributed**.
- ▶ variances of both populations are equal.

### Strategy for Checking Model Assumptions:

- ▶ check the form of the model.
- ▶ check for outliers.
- ▶ check for independence.
- ▶ check for constant variance.
- ▶ check for normality.

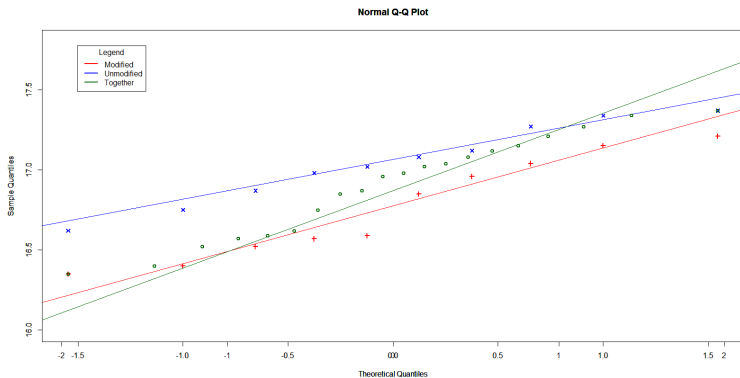
# Simple Comparative Experiment - checking assumptions

## Kernel density estimations of data sets



## Simple Comparative Experiment - checking assumptions

The equal variance and normality assumptions are easy to check using **normal probability plot** (QQ-plot).





## Simple Comparative Experiment - checking assumptions

**Test for the hypothesis of normality.**

**One sample Kolmogorov - Smirnov test** tests the null hypothesis that a probability distribution of given sample is the same as a reference probability distribution.

**Shapiro - Wilk normality test** and **Anderson - Darling normality test** test the null hypothesis that a given sample came from a normally distributed population.

**Hypothesis testing of normality**

| Data       | Shapiro-Wilk test |         | Kolmogorov-Smirnov test |         |
|------------|-------------------|---------|-------------------------|---------|
|            | Statistics $W$    | p-value | Statistics $D$          | p-value |
| Modified   | 0.9186            | 0.3457  | 0.2088                  | 0.7027  |
| Unmodified | 0.9626            | 0.8153  | 0.1211                  | 0.9943  |
| Together   | 0.9540            | 0.4326  | 0.1226                  | 0.8891  |

## Simple Comparative Experiment -Non-parametric alternative

For two-sample comparisons, two non-parametric tests can be used, depending on the way data are collected. If both sample are independent, we use Mann-Whitney-Wilcoxon rank sum test, while for paired sample the corresponding test is called Wilcoxon signed rank test.

Both are called using R function:

```
wilcox.test
```

and the option

```
paired={TRUE|FALSE}.
```

## Choice of sample size

In hypothesis testing we eliminate the probability of type I error by setting the significance level  $\alpha$ , but how to decrease the probability of type II error  $\beta$ ?

In t-test, the Power depends on the significance level  $\alpha$ , on the true difference in means  $\delta$  and the sample size  $n$ .

| Two-sample t test power calculation |                         |
|-------------------------------------|-------------------------|
| Sample size calculation             | Power calculation       |
| n = 28.17                           | n = 10                  |
| delta = 0.28                        | delta = 0.28            |
| sd = 0.28                           | sd = 0.28               |
| sig.level = 0.05                    | sig.level = 0.05        |
| power = 0.95                        | power = 0.54            |
| alternative = two.sided             | alternative = two.sided |

```
#calculation of required sample size
power.t.test(power = .95, delta = diff.t, sd = sigma.t,
sig.level = 0.05, type = "two.sample", alternative = "two.sided")
#calculation of the power of the test
power.t.test(n = 10, delta = diff.t, sd = sigma.t,
sig.level = 0.05, type = "two.sample", alternative = "two.sided")
```

## Choice of sample size

Calculation of Power for equal but unknown populations SD's:

```
C      = qt(0.975, n1 + n2 - 2)
sd     = sigma*sqrt(1/n1 + 1/n2)
power  = 1-pt(C, n1+n2-2, ncp=delta/se)+pt(-C, n1+n2-2, ncp=delta/sd)
```

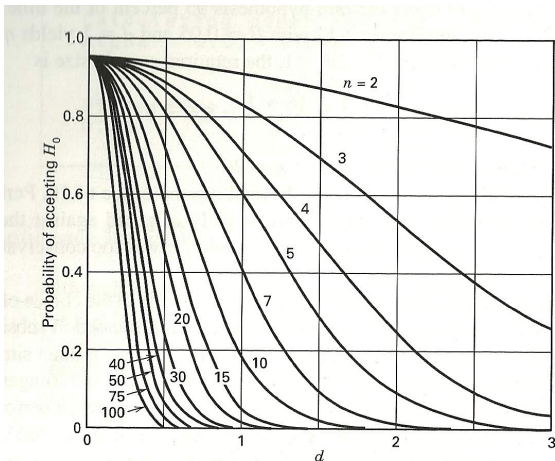
We have four connected quantities:

- ▶ sample size;
- ▶ difference between means;
- ▶ significance level of the test;
- ▶ power of the test.

We can plot the graph of  $\beta$  (the probability of type II error) versus  $\delta$  - the true difference in means. This curve is called Operating Characteristic curve (OC curve).

## Choice of sample size

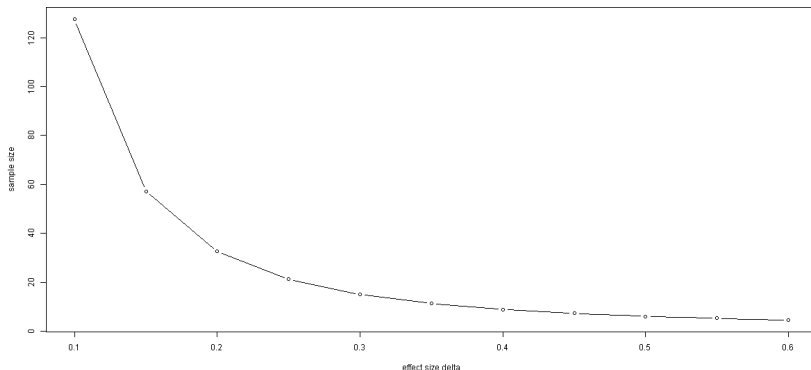
Operating characteristic curve (OC curve) for two-sided t-test, where  $d = \frac{|\delta|}{2\sigma}$ .



**Figure 2-12** Operating characteristic curves for the two-sided  $t$ -test with  $\alpha = 0.05$ . (Reproduced with permission from "Operating Characteristics for the Common Statistical Tests of Significance," C. L. Ferris, F. E. Grubbs, and C. L. Weaver, *Annals of Mathematical Statistics*, June 1946.)

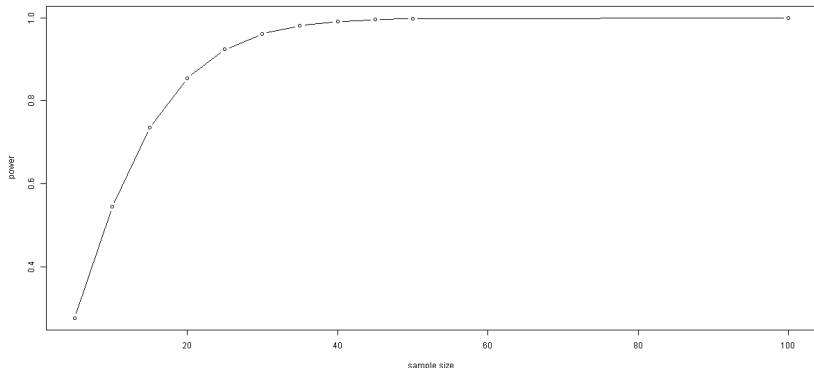
## Choice of sample size - Sample curve

```
for (i in c(.1,.15,.2,.25,.3,.35,.4,.45,.5,0.55,0.6))  
{  
  pwrt<-power.t.test(sd=0.284,delta=i,power=.8,sig.level=.05,  
    type="two.sample",alternative="two.sided")  
  ptab<-rbind(ptab, cbind(pwrt$d, pwrt$n))  
}  
plot(ptab[,1],ptab[,2],type="b",xlab="effect size",ylab="sample size")
```



## Choice of sample size - Power curve

```
pwrt<-power.t.test(delta=0.278, sd=0.284,  
                    n=c(5,10,15,20,25,30,35,40,45,50,100),  
                    sig.level=.05,  
                    type="two.sample",  
                    alternative="two.sided")
```



## Next Lecture and Today Exercise

### Next Lecture: Comparing several treatment means, linear regression

- ▶ Review basic statistical concepts and analysis of variance (ANOVA).
- ▶ Design and analyze single factor experiments using ANOVA.
- ▶ Analysis of the fixed effects model.
- ▶ Interpretation of results.
- ▶ Model diagnostic checking, residual analysis  
(normal probability plot, plot of residuals, residuals versus fitted values).
- ▶ Multiple linear regression model.
- ▶ Submodel, Stepwise regression.

### Today Exercise:

- ▶ Run and familiarize with R.
- ▶ Solve problems 2.20, 2.26, 2.30 from following slides.  
(originally from Montgomery - Design and Analysis of Experiments).



## Exercise 2.20

The shelf life of a carbonated beverage is of interest. Ten bottles are randomly selected and tested, and the following results are obtained:

| Days |     |
|------|-----|
| 108  | 138 |
| 124  | 163 |
| 124  | 159 |
| 106  | 134 |
| 115  | 139 |

1. We would like to demonstrate that the mean shelf life exceeds 120 days. Set up appropriate hypotheses for investigating this claim.
2. Test these hypotheses using significant level  $\alpha = 0.01$ . Find the P-value for the test. What are your conclusions?
3. Construct a 99 percent confidence interval on the mean shelf life.

## Exercise 2.26

The following are the burning times (in minutes) of chemical flares of two different formulations. The design engineers are interested in both the mean and variance of the burning times.

| Type 1 |    | Type 2 |    |
|--------|----|--------|----|
| 65     | 82 | 64     | 56 |
| 81     | 67 | 71     | 69 |
| 57     | 59 | 83     | 74 |
| 66     | 75 | 59     | 82 |
| 82     | 70 | 65     | 79 |

1. Test the hypothesis that the two variances are equal. Use  $\alpha = 0.05$ .
2. Using the results of part 1), test the hypothesis that the mean burning times are equal. Use  $\alpha = 0.05$ . What is the P-value for this test?
3. Discuss the role of the normality assumption in this problem. Check the assumption of normality for both types of flares.

## Exercise 2.30

Front housings for cell phones are manufactured in an injection molding process. The time the part is allowed to cool in the mold before removal is thought to influence the occurrence of a particularly troublesome cosmetic defect, flow lines, in the finished housing. After manufacturing, the housings are inspected visually and assigned a score between 1 and 10 based on their appearance, with 10 corresponding to a perfect part and 1 corresponding to a completely defective part. An experiment was conducted using two cool-down times, 10 and 20 seconds, and 20 housings were evaluated at each level of cool-down time. All 40 observations in this experiment were run in random order.

| 10 seconds |   |   |   | 20 seconds |   |   |   |
|------------|---|---|---|------------|---|---|---|
| 1          | 3 | 2 | 6 | 7          | 6 | 8 | 9 |
| 1          | 5 | 3 | 3 | 5          | 5 | 9 | 7 |
| 5          | 2 | 1 | 1 | 5          | 4 | 8 | 6 |
| 5          | 6 | 2 | 8 | 6          | 8 | 4 | 5 |
| 3          | 2 | 5 | 3 | 6          | 8 | 7 | 7 |

1. Is there evidence to support the claim that the longer cool-down time results in fewer appearance defects? Use  $\alpha = 0.05$ .
2. What is the P-value for the test conducted in the previous part?
3. Find a 95 percent confidence interval on the difference in means. Provide a practical interpretation of this interval.
4. Draw dot diagrams to assist in interpreting the results from this experiment.