# Course 02402 Introduction to Statistics

# Lecture 6: Two-sample comparisons and power/sample size

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

# Overview

1. Motivating example: Nutrition study
2. Repetition: $p$-values and hypothesis tests
3. Two-sample $t$-test and $p$-value
4. Confidence interval for the mean difference
5. Overlapping confidence intervals?
6. The paired setup
7. Checking the normality assumptions
8. Planning for wanted precision or power
    - Precision requirements
    - Power and sample size - one-sample
    - Power and sample size - two-sample
9. The pooled t-test - a possible alternative

# Overview

# Motivating example - nutrition study

### Nutrition study

In a nutrition study, the aim is to investigate whether there is a difference in the energy usage for two different types of (moderately physically demanding) work.

In the study, the energy usage of 9 nurses from hospital $A$ and 9 (other) nurses from hospital $B$ have been measured. The measurements are given in the following table in mega Joule (MJ):

Sample from each hospital, $n_1 = n_2 = 9$:

| Hospital A | Hospital B |
|---:|---:|
| 7.53 | 9.21 |
| 7.48 | 11.51 |
| 8.08 | 12.79 |
| 8.09 | 11.85 |
| 10.15 | 9.97 |
| 8.40 | 8.79 |
| 10.88 | 9.69 |
| 6.13 | 9.68 |
| 7.90 | 9.19 |

# Example - nutrition study

The hypothesis of no difference in mean energy usage is in focus:

$$H_0: \ \mu_A = \mu_B$$

# Example - nutrition study

The hypothesis of no difference in mean energy usage is in focus:

$$H_0: \ \mu_A = \mu_B$$

Sample means and standard deviations:

$$\hat{\mu}_A = \bar{x}_A = 8.293, \ (s_A = 1.428)$$
$$\hat{\mu}_B = \bar{x}_B = 10.298, \ (s_B = 1.398)$$

# Example - nutrition study

The hypothesis of no difference in mean energy usage is in focus:

$$H_0: \ \mu_A = \mu_B$$

Sample means and standard deviations:

$$\hat{\mu}_A = \bar{x}_A = 8.293, \ (s_A = 1.428)$$
$$\hat{\mu}_B = \bar{x}_B = 10.298, \ (s_B = 1.398)$$

Is data in accordance with the null hypothesis $H_0$?

Data: $\bar{x}_B - \bar{x}_A = 2.005$

Null hypothesis: $H_0: \ \mu_B - \mu_A = 0$

# Example - nutrition study

The hypothesis of no difference in mean energy usage is in focus:

$$H_0: \ \mu_A = \mu_B$$

Sample means and standard deviations:

$$\hat{\mu}_A = \bar{x}_A = 8.293, \ (s_A = 1.428)$$
$$\hat{\mu}_B = \bar{x}_B = 10.298, \ (s_B = 1.398)$$

NEW: $p$-**value for difference:**

$$p\text{-value} = 0.0083$$

(Found under the scenario that $H_0$ is true.)

Is data in accordance with the null hypothesis $H_0$?

$$\text{Data: } \bar{x}_B - \bar{x}_A = 2.005$$

$$\text{Null hypothesis: } H_0: \ \mu_B - \mu_A = 0$$

NEW: **Confidence interval for difference**:

$$2.005 \pm 1.412 = [0.59; \ 3.42]$$

# Overview

# Definition of a hypothesis test and significance (repetition)

### Definition 3.24. Hypothesis test:

When we *carry out a hypothesis test*, we decide against a null hypothesis or not, using the data.

A null hypothesis is *rejected* if the $p$-value, calculated after the data has been observed, is less than some $\alpha$ – that is, if the $p$-value $< \alpha$, where $\alpha$ is some pre-specifed (so-called) *significance level*.

If we do not reject the null hypothesis, it is said to be *accepted*.

### Definition 3.29. Statistical significance:

An *effect* is said to be *(statistically) significant* if the $p$-value is less than the significance level $\alpha$.

Often (and unless otherwise mentioned) we use $\alpha = 0.05$.

# Steps of a hypothesis test - overview (repetition)

Generelly, a hypothesis test consists of the following steps:

1. Formulate the hypothesis and choose the level of significance $\alpha$ (the "risk-level").
2. Calculate, using the data, the value of the test statistic.
3. Calculate the $p$-value using the test statistic and the relevant distribution. Compare the $p$-value to the significance level $\alpha$ and make a conclusion.

   **OR**:

   Alternatively, make a conclusion based on the relevant critical value(s).

# Definition and interpretation of the $p$-value (repetition)

The $p$-value expresses the *evidence* against the null hypothesis – Table 3.1:

| | |
|---|---|
| $p < 0.001$ | Very strong evidence against $H_0$ |
| $0.001 \leq p < 0.01$ | Strong evidence against $H_0$ |
| $0.01 \leq p < 0.05$ | Some evidence against $H_0$ |
| $0.05 \leq p < 0.1$ | Weak evidence against $H_0$ |
| $p \geq 0.1$ | Little or no evidence against $H_0$ |

Definition 3.22 of the $p$-value:

**The $p$-value** is the probability of obtaining a test statistic that is at least as extreme as the test statistic that was actually observed. This probability is calculated under the assumption that the null hypothesis is true.

# Critical value, confidence interval and hypothesis test (repetition)

### Theorem 3.33: Critical value method $\approx$ Confidence interval method

We consider a $(1-\alpha)\cdot 100\%$ confidence interval for $\mu$:

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

The confidence interval corresponds to the acceptance region for $H_0$ when testing the (non-directional) hypothesis

$$H_0: \quad \mu = \mu_0$$

# Critical value, confidence interval and hypothesis test (repetition)

### Theorem 3.33: Critical value method $\approx$ Confidence interval method

We consider a $(1-\alpha)\cdot 100\%$ confidence interval for $\mu$:

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

The confidence interval corresponds to the acceptance region for $H_0$ when testing the (non-directional) hypothesis

$$H_0: \quad \mu = \mu_0$$

### (New) interpretation of the confidence interval:

The confidence interval covers those values of the parameter that we believe in given the data.

The confidence interval contains those values of the parameter that we would accept by the corresponding hypothesis test.

# Overview

# Method 3.49: Two-sample $t$-test

Computing the test statistic:

When considering the null hypothesis about the difference between the means of two *independent* samples:

$$\delta = \mu_2 - \mu_1$$

$$H_0: \ \delta = \delta_0$$

the (Welch) two-sample $t$-test statistic is

$$t_{\text{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

# Theorem 3.50: Distribution of the (Welch) $t$-test statistic

The Welch $t$-test statistic is (approximately) $t$-distributed:

Under the null hypothesis, the (Welch) two-sample statistic, seen as a random variable:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

approximately follows a $t$-distribution with $\nu$ degrees of freedom, where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

if the two population distributions are normal or if the two sample sizes are large enough.

# Method 3.51: Two-sample $t$-test

The level $\alpha$ test is:

1. Compute $t_{\text{obs}}$ and $\nu$ as given above.
2. Compute the evidence against the *null hypothesis*[a] $H_0 : \mu_1 - \mu_2 = \delta_0$ vs. the *alternative hypothesis* $H_1 : \mu_1 - \mu_2 \neq \delta_0$ by the

$$p\text{-value} = 2 \cdot P(T > |t_{\text{obs}}|)$$

   where the $t$-distribution with $\nu$ degrees of freedom is used.
3. If $p$–value $< \alpha$: We reject $H_0$, otherwise we accept $H_0$.

   OR

   The rejection/acceptance conclusion could alternatively, but equivalently, be made based on the critical value(s) $\pm t_{1-\alpha/2}$:

   If $|t_{\text{obs}}| > t_{1-\alpha/2}$ we reject $H_0$, otherwise we accept $H_0$.

[a] We are often interested in the test where $\delta_0 = 0$

# Example - nutrition study

The hypothesis of no difference is in focus:

$$H_0: \ \delta = \mu_B - \mu_A = 0$$

versus the non-directional (= two-sided) alternative:

$$H_1: \ \delta = \mu_B - \mu_A \neq 0$$

First, the computations of $t_{\text{obs}}$ and $\nu$:

$$t_{\text{obs}} = \frac{10.298 - 8.293}{\sqrt{2.0394/9 + 1.954/9}} = 3.01$$

and

$$\nu = \frac{\left(\frac{2.0394}{9} + \frac{1.954}{9}\right)^2}{\frac{(2.0394/9)^2}{8} + \frac{(1.954/9)^2}{8}} = 15.99$$

# Example - nutrition study

Next, the $p$-value is found:

$$p\text{–value} = 2 \cdot P(T > |t_{\text{obs}}|) = 2P(T > 3.01) = 2 \cdot 0.00415 = 0.0083$$

```
## Nutrition study example: P(T > 3.01)
1 - pt(3.01, df = 15.99)

## [1] 0.004154
```

# Example - nutrition study

Next, the $p$-value is found:

$$p\text{–value} = 2 \cdot P(T > |t_{\text{obs}}|) = 2P(T > 3.01) = 2 \cdot 0.00415 = 0.0083$$

```r
## Nutrition study example: P(T > 3.01)
1 - pt(3.01, df = 15.99)

## [1] 0.004154
```

Evaluate the evidence (Table 3.1):

There is strong evidence against the null hypothesis.

# Example - nutrition study

Next, the $p$-value is found:

$$p\text{–value} = 2 \cdot P(T > |t_{\text{obs}}|) = 2P(T > 3.01) = 2 \cdot 0.00415 = 0.0083$$

```r
## Nutrition study example: P(T > 3.01)
1 - pt(3.01, df = 15.99)

## [1] 0.004154
```

Evaluate the evidence (Table 3.1):

There is strong evidence against the null hypothesis.

Conclude based on $\alpha = 0.05$:

We reject the null hypothesis. There is a significant difference between the two groups. Nurses in Hospital $B$ can be said to have a larger (mean) energy usage than nurses in Hospital $A$.

# Overview

# Method 3.47: Confidence interval for $\mu_1 - \mu_2$

**The confidence interval for the mean difference:**

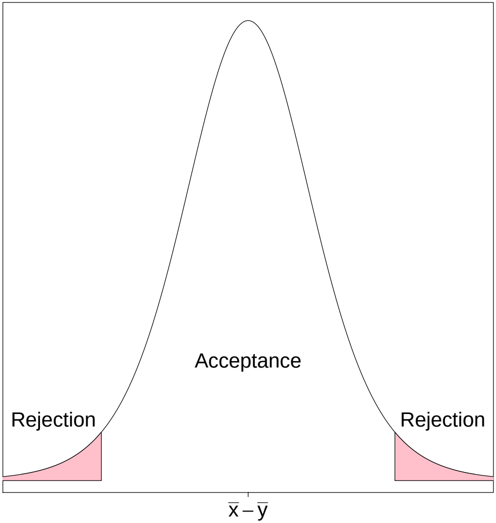
For two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ the $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $t_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$-quantile from the $t$-distribution with $\nu$ degrees of freedom given by Theorem 3.50 (see above).

# Confidence interval and hypothesis test (repetition)

The acceptance region consists of the potential values for $\mu_1 - \mu_2$ that are not too far away from the data:



# Example - nutrition study:

Let's find the $95\%$ confidence interval for $\mu_B - \mu_A$. Using $\nu = 15.99$, the relevant $t$-quantile is

$$t_{0.975} = 2.120$$

and the confidence interval becomes:

$$10.298 - 8.293 \pm 2.120 \cdot \sqrt{\frac{2.0394}{9} + \frac{1.954}{9}}.$$

This gives the confidence interval also shown above:

$$[0.59;\ 3.42]$$

# Example - nutrition study - everything in R:

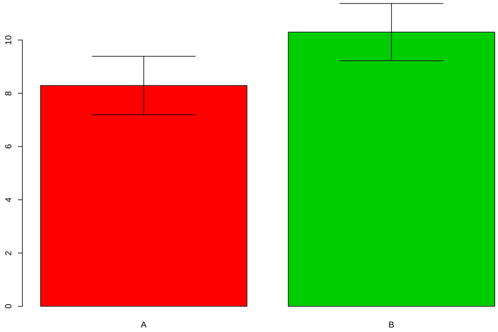
```
# Read the two samples into R
xA = c(7.53, 7.48, 8.08, 8.09, 10.15, 8.4, 10.88, 6.13, 7.9)
xB = c(9.21, 11.51, 12.79, 11.85, 9.97, 8.79, 9.69, 9.68, 9.19)

# Perform Welch two-sample t-test
t.test(xB, xA)

##
##  Welch Two Sample t-test
##
## data:  xB and xA
## t = 3, df = 16, p-value = 0.008
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5923 3.4166
## sample estimates:
## mean of x mean of y
##    10.298     8.293
```

# Overview

# Example - nutrition study - presentation of result

Barplots with *error bars* are often seen:

A grouped barplot with some "error bars" - below, the 95%-confidence intervals for the mean of each group are shown:



# Be careful about using "overlapping confidence intervals"

The approach is actually using an incorrect variation for evaluation of the difference:

$$\sigma_{(\bar{X}_A - \bar{X}_B)} \neq \sigma_{\bar{X}_A} + \sigma_{\bar{X}_B}$$

$$\mathrm{Var}(\bar{X}_A - \bar{X}_B) = \mathrm{Var}(\bar{X}_A) + \mathrm{Var}(\bar{X}_B)$$

Assume that the two standard-errors are 3 and 4: The sum is 7, but $\sqrt{3^2 + 4^2} = 5$

# Be careful about using "overlapping confidence intervals"

The approach is actually using an incorrect variation for evaluation of the difference:

$$\sigma_{(\bar{X}_A - \bar{X}_B)} \neq \sigma_{\bar{X}_A} + \sigma_{\bar{X}_B}$$

$$\text{Var}(\bar{X}_A - \bar{X}_B) = \text{Var}(\bar{X}_A) + \text{Var}(\bar{X}_B)$$

Assume that the two standard-errors are 3 and 4: The sum is 7, but $\sqrt{3^2 + 4^2} = 5$

The correct relation between the standard deviations is:

$$\sigma_{(\bar{X}_A - \bar{X}_B)} < \sigma_{\bar{X}_A} + \sigma_{\bar{X}_B}$$

# Be careful about using "overlapping confidence intervals"

**Remark 3.59. Rule for using "overlapping confidence intervals":**

When two CIs do NOT overlap: The two groups are significantly different.

When two CIs DO overlap: We do not know what the conclusion is – but we could, e.g., make a CI for the mean *difference* instead, to investigate.

# Overview

# Motivating example - sleep medicine

### Difference between sleep medicines?

In a study, the aim is to compare two kinds of sleep medicine, A and B. 10 test persons tried both kinds of medicine, and the following results were obtained, in terms of prolonged sleep length (in hours) for each medicine type:

### Sample, $n = 10$:

| Person | $A$ | $B$ | $D = B - A$ |
|---|---|---|---|
| 1 | $+0.7$ | $+1.9$ | $+1.2$ |
| 2 | -1.6 | $+0.8$ | $+2.4$ |
| 3 | -0.2 | $+1.1$ | $+1.3$ |
| 4 | -1.2 | $+0.1$ | $+1.3$ |
| 5 | -1.0 | -0.1 | $+0.9$ |
| 6 | $+3.4$ | $+4.4$ | $+1.0$ |
| 7 | $+3.7$ | $+5.5$ | $+1.8$ |
| 8 | $+0.8$ | $+1.6$ | $+0.8$ |
| 9 | 0.0 | $+4.6$ | $+4.6$ |
| 10 | $+2.0$ | $+3.4$ | $+1.4$ |

# The paired setup and analysis = one-sample analysis

```r
# Read the two samples into R
x1 = c(.7,-1.6,-.2,-1.2,-1,3.4,3.7,.8,0,2)
x2 = c(1.9,.8,1.1,.1,-.1,4.4,5.5,1.6,4.6,3.4)

# Compute differences to get a paired t-test
dif = x2 - x1

# Perform paired t-test
t.test(dif)

##
##  One Sample t-test
##
## data:  dif
## t = 4.7, df = 9, p-value = 0.001
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.8613 2.4787
## sample estimates:
## mean of x
##      1.67
```

# The paired setup and analysis = one-sample analysis

```
# Another way to perform the paired t-test
t.test(x2, x1, paired = TRUE)

##
##  Paired t-test
##
## data:  x2 and x1
## t = 4.7, df = 9, p-value = 0.001
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.8613 2.4787
## sample estimates:
## mean of the differences
##                    1.67
```

# Paired vs. independent experiment

### Completely randomized (independent samples)

20 patients are used and allocated, completely at random, to one of the two treatments (usually making sure to have 10 patients in each group). That is, there are different patients in the different treatment groups.

### Paired (dependent samples)

10 patients are used, and each of them tests both of the treatments. Usually this will involve some time in between treatments, to make sure that the experiment is meaningful, and also one would typically make sure that some patients try A before B and others try B before A (order allocated at random). That is, the same patients are included in both treatment groups.

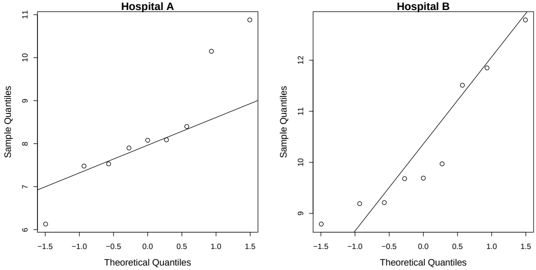
# Example - sleep medicine - WRONG analysis

```r
# WRONG analysis
t.test(x1, x2)

##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## t = -1.9, df = 18, p-value = 0.07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4854  0.1454
## sample estimates:
## mean of x mean of y
##      0.66      2.33
```

# Overview

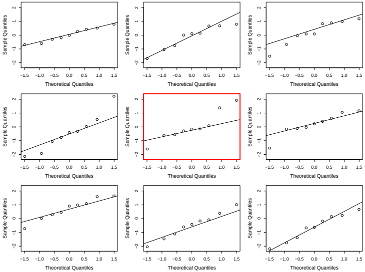
# Example - Q-Q plot for EACH sample:

```r
# Q-Q plots separately for each sample
qqnorm(xA, main = "Hospital A")
qqline(xA)
qqnorm(xB, main = "Hospital B")
qqline(xB)
```



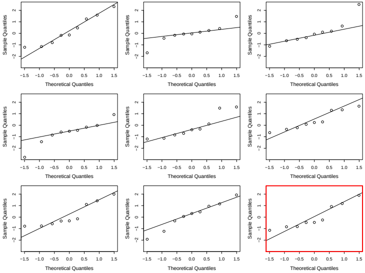
# Example - Comparing with simulated data, A

```
# Multiple (simulated) Q-Q plots and sample A
require(MESS)
fitA <- lm(xA ~ 1)
qqnorm.wally <- function(x, y, ...) { qqnorm(y, ...); qqline(y, ...)}
wallyplot(fitA, FUN = qqnorm.wally, main = "")
```



# Example - Comparing with simulated data, B

```r
# Multiple (simulated) Q-Q plots and sample B
fitB <- lm(xB ~ 1)
qqnorm.wally <- function(x, y, ...) { qqnorm(y, ...); qqline(y, ...)}
wallyplot(fitB, FUN = qqnorm.wally, main = "")
```



# Overview

# Planning of study with requirements to the precision

The one-sample $100 \cdot (1-\alpha)\%$ CI: $\bar{x} \pm t_{1-\alpha/2} \cdot s/\sqrt{n}$.

The *margin of error (ME)* is defined as

$$t_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

## Method 3.63: The one-sample CI sample size formula:

When $\sigma$ is known, or guessed to be some value, we can calculate the sample size $n$ needed to achieve a given margin of error, $ME$, with probability $1-\alpha$, as:

$$n = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{ME}\right)^2$$

# Example, height data again

Sample mean og standard deviation:

$$\bar{x} = 178$$
$$s = 12.21$$

Estimate the population mean and standard deviation:

$$\hat{\mu} = 178$$
$$\hat{\sigma} = 12.21$$

# Example, height data again

Sample mean og standard deviation:

$$\bar{x} = 178$$
$$s = 12.21$$

Estimate the population mean and standard deviation:

$$\hat{\mu} = 178$$
$$\hat{\sigma} = 12.21$$

If we want $ME = 3$ cm with $95\%$ confidence, how large should $n$ be?

$$n = \left(\frac{1.96 \cdot 12.21}{3}\right)^2 = 63.64 \approx 64$$

# Planning, power

## What is the power of a future study/experiment:

- The probability of detecting an (assumed) effect.
- $P(\text{reject } H_0)$ when $H_1$ is true.
- Probability of correct rejection of $H_0$.
- Challenge: The null hypothesis can be wrong in many ways!
- Practically: Scenario-based approach
    - E.g. "What if $\mu = 86$, how good will my study be to detect this?"
    - E.g. "What if $\mu = 84$, how good will my study be to detect this?"
    - etc.

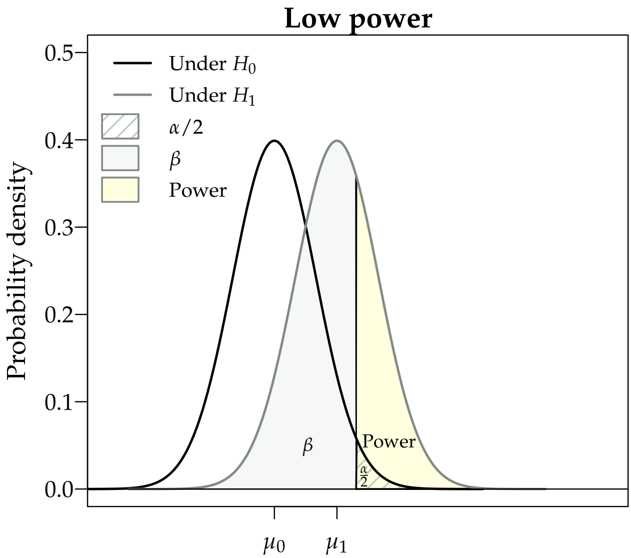
# Planning and power

When the null hypothesis has been decided on:

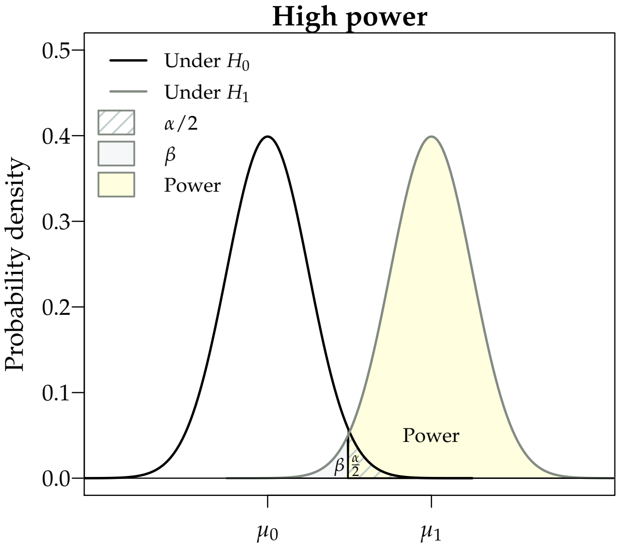
If you know (or set/guess) four out of the following five pieces of information, you can find the fifth:

- The sample size $n$.
- Significance level $\alpha$ of the test.
- A difference in mean that you would want to detect (effect size) $\mu_0 - \mu_1$.
- The population standard deviation, $\sigma$.
- The power $(1-\beta)$.

# Low power example



# High power example



# Planning, sample size $n$

The big practical question: What should $n$ be?

The experiment should be large enough to detect a relevant effect with high power $1-\beta$ (usually at least $80\%$):

# Planning, sample size $n$

The big practical question: What should $n$ be?

The experiment should be large enough to detect a relevant effect with high power $1-\beta$ (usually at least $80\%$):

Metode 3.65: The one-sample sample size formula:

For the one-sample t-test for given $\alpha$, $\beta$ and $\sigma$:

$$n = \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1}\right)^2$$

Here, $\mu_0 - \mu_1$ is the difference in mean that we would like to detect, and $z_{1-\beta}$, $z_{1-\alpha/2}$ are quantiles of the standard normal distribution.

# Example - The power for $n = 40$

```r
# Power calculation (one-sample)
power.t.test(n = 40, delta = 4, sd = 12.21, type = "one.sample")

##
##      One-sample t test power calculation
##
##               n = 40
##           delta = 4
##              sd = 12.21
##       sig.level = 0.05
##           power = 0.5242
##     alternative = two.sided
```

# Example - The sample size for power $= 0.80$

```r
# Sample size calculation (one-sample)
power.t.test(power = .80, delta = 4, sd = 12.21, type = "one.sample")

##
##      One-sample t test power calculation
##
##               n = 75.08
##           delta = 4
##              sd = 12.21
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
```

# Power and sample size - two-sample

Finding the power of detecting a group difference of $2$ with $\sigma = 1$ for $n = 10$:

```
# Power calculation (two-sample)
power.t.test(n = 10, delta = 2, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 2
##              sd = 1
##       sig.level = 0.05
##           power = 0.9882
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Power and sample size - two-sample

Finding the sample size for detecting a group mean difference of $2$ with $\sigma = 1$ and power $= 0.9$:

```
# Sample size calculation (two-sample)
power.t.test(power = 0.90, delta = 2, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 6.387
##           delta = 2
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Power and sample size - two-sample

Finding the detectable effect size (delta) with $\sigma = 1$, $n = 10$ and power $= 0.9$:

```
## Detectable effect size (two-sample)
power.t.test(power = 0.90, n = 10, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 1.534
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Overview

# The pooled two-sample $t$-test statistic

The *pooled* estimate of variance (assuming $\sigma_1^2 = \sigma_2^2$)

Method 3.52

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

The pooled t-test statistic, Method 3.53

When considering the null hypothesis about the difference between the means of two *independent* samples:

$$\delta = \mu_2 - \mu_1$$

$$H_0: \ \delta = \delta_0$$

the pooled two-sample $t$-test statistic is

$$t_{\text{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}$$

# Theorem 3.54: The distribution of the pooled test-statistic

... is a $t$-distribution:

The pooled two-sample statistic seen as a random variable:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_p^2/n_1 + S_p^2/n_2}}$$

follows, under the null hypothesis and under the assumption that $\sigma_1^2 = \sigma_2^2$, a $t$-distribution with $n_1 + n_2 - 2$ degrees of freedom if the two population distributions are normal.

# We always use the "Welch" t-test

Almost (fool)proof to always use the Welch-version:

- If $s_1^2 = s_2^2$, the Welch and the pooled t-test statistics are the same.
- Only when the two variances become really different, the two test-statistics may differ in an important way. Furthermore, if this is the case, we would not tend to favour the pooled version, since the assumption of equal variances appears questionable then.
- Only for cases with a small sample size in at least one of the two groups, the pooled approach may provide slightly higher power (if you believe in the equal variance assumption). For these cases, the Welch approach is then a somewhat cautious approach.

# Overview