Course 02323 Introduction to Statistics 02402 Statistics (Polytechnical Foundation)

Lecture 6: Analysis of Two Samples

DTU Compute Technical University of Denmark 2800 Lyngby – Denmark

Overview

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- t-test with pooled variance An alternative
- Sonfidence interval for the difference
- Overlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Overview

Summary from last week

- Motivating example: Nutrition study
- I t-test with two samples
- I t-test with pooled variance An alternative
- 6 Confidence interval for the difference
- 6 Overlapping confidence intervals?
- *t*-test with two paired samples (paired *t*-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Hypothesis testing



t-test "cook-book":

- Null hypothesis $H_0: \mu = \mu_0 (= 0?)$
- Calculate: \bar{x} , s, and $s.e._{\bar{x}} = s/\sqrt{n}$
- Calculate test statistic: $t_{obs} = \frac{\bar{x} \mu_0}{s.e.\bar{x}}$
- Calculate p-value: $2 \cdot P(T > |t_{obs}|)$ (lookup in t-distribution, draw a diagram)
- Compare p-value with desired lpha

- Is the value μ_0 outside the $(1-\alpha) \cdot 100\%$ confidence interval? (e.g., 95%-Cl)

- Is the calculated $|t_{obs}|$ greater than $t_{1-\alpha/2}$? (e.g., $t_{0.975}$)
- Is the p-value smaller than α ? (e.g., 0.05)

Yes? - then we reject the null hypothesis at significance level lpha.

The p-value corresponds to the significance level (α) that would just include the null hypothesis in the confidence interval.

(DTU Compute)

Today



Today

- We often want to know if there is a ("significant") difference between two different things/groups/interventions.
- We have seen examples where data reflect measured differences—and we test whether this difference in means is zero.
- But what if, for example, we want to know if there is a difference between two different groups where we cannot collect data for individual differences? (the difference cannot be measured on a single statistical unit, and the groups may also be of different sizes)
- In that case, we need to look at the t-test for two samples.

Overview

Summary from last week

Motivating example: Nutrition study

- It-test with two samples
- I t-test with pooled variance An alternative
- 5 Confidence interval for the difference
- Overlapping confidence intervals?
- *t*-test with two paired samples (paired *t*-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Motivating example: Nutrition study

Difference in Energy Expenditure?

In a nutrition study, we want to investigate whether there is a difference in energy expenditure for different types of (moderately physically demanding) work.

In the study, 9 nurses from hospital A and 9 (other) nurses from hospital B had their energy expenditure measured. The measurements are shown in the following table (in the unit megajoules, MJ):

	Hospital A	Hospital B
Sample from each hospital:	7.53	9.21
$n_1 = n_2 = 9$:	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



- Go to today's Python notebook in VS Code
 - "Example: Nutrition study 1"





Overview

- Summary from last week
- Motivating example: Nutrition study

I t-test with two samples

- I t-test with pooled variance An alternative
- 5 Confidence interval for the difference
- Overlapping confidence intervals?
- *t*-test with two paired samples (paired *t*-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Difference between two groups



What is the null hypothesis? How would you formulate it? (discuss for 2 minutes)

(board)

Difference between two groups



Considerations for the null hypothesis:

- There is no difference between groups "A" and "B"
- The two samples come from the *exact same* underlying distribution (?)
- The underlying distributions have the same mean: $\mu_A = \mu_B$
- The underlying distributions have the same variance: $\sigma_{A} = \sigma_{B}$ (?)

The hypothesis of no difference (in average energy expenditure) is to be tested:

 $H_0: \ \mu_A = \mu_B \quad \text{or} \quad H_0: \ \mu_A - \mu_B = 0$

The hypothesis of no difference (in average energy expenditure) is to be tested:

 $H_0: \ \mu_A = \mu_B \quad \text{or} \quad H_0: \ \mu_A - \mu_B = 0$

Sample Means and Standard Deviations: $\hat{\mu}_A = \bar{x}_A = 8.293 \ (s_A = 1.428)$ $\hat{\mu}_B = \bar{x}_B = 10.298 \ (s_B = 1.398)$

The hypothesis of no difference (in average energy expenditure) is to be tested:

 $H_0: \ \mu_A = \mu_B \quad \text{or} \quad H_0: \ \mu_A - \mu_B = 0$

Sample Means and Standard Deviations:

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

Are the data consistent with the null hypothesis H_0 ? Data: $\bar{x}_B - \bar{x}_A = 2.005$

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

The hypothesis of no difference (in average energy expenditure) is to be tested:

 $H_0: \ \mu_A = \mu_B \quad \text{or} \quad H_0: \ \mu_A - \mu_B = 0$

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 the difference:

p = 0.0083

(Calculated under the assumption that H_0 is true.)

Are the data consistent with the null hypothesis H_0 ?

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

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

NEW: Confidence Interval for the Difference:

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

Theory: Difference between two means

What is our null hypothesis?

$$H_0: \quad \delta = \mu_X - \mu_Y$$

Notation: δ is the "true" difference (μ_X and μ_Y are the "true" means)

What is our stochastic variable?

$$D = \bar{X} - \bar{Y}$$

Theory: Difference between two means

Mean of D:

$$\mathbf{E}[D] = \mathbf{E}[\bar{X} - \bar{Y}] = \mathbf{E}[\bar{X}] - \mathbf{E}[\bar{Y}] = \mu_X - \mu_Y$$

Variance of D:

$$\mathbf{V}[D] = \mathbf{V}[\bar{X} - \bar{Y}] = 1^2 \mathbf{V}[\bar{X}] + (-1)^2 \mathbf{V}[\bar{Y}] = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$$

Standard deviation of *D*:

$$\sigma_D = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \quad \text{or} \quad "s.e._D" = \sqrt{(s.e._{\bar{X}})^2 + (s.e._{\bar{Y}})^2}$$

Theory: Notation (from last week: one sample)

- μ_X : the "true" mean for X
- σ_X^2 : the variance (the "true" variance) for X

 \bar{X} : stochastic variable describing the mean in a (theoretical) sample of size n_X .

 \bar{x} : sample mean calculated from actual data (the sample).

Last time, we derived the theory on how \bar{X} is distributed:

$$\mathbf{E}[\bar{X}] = "\mu_{\bar{X}}" = \mu_X \qquad \mathbf{V}[\bar{X}] = "\sigma_{\bar{X}}^2" = \frac{\sigma_X^2}{n_X}$$

And we discussed the "standard error"

$$s.e._{\bar{X}} = \sqrt{\frac{\sigma_X^2}{n_X}}$$
 $s.e._{\bar{X}} = \sqrt{\frac{s_X^2}{n_X}}$ Kahoot!

02402 Statistics (Polytechnical Foundation)

Theory: Notation (two samples)

δ : the "true" difference

D: stochastic variable that describes the difference between means in two (theoretical) samples of size n_X and n_Y .

d: observed difference calculated from real data for two samples.

We have just derived the theory of how D is distributed:

$$\mathbf{E}[D] = "\mu_D" = \mu_X - \mu_Y$$
 $\mathbf{V}[D] = "\sigma_D^2" = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$

We can now define several "standard errors":

$$s.e._D = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \qquad s.e._d = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

Method 3.49: The test statistic in a (Welch) *t*-test with two unpaired samples

Assumptions:

The test is valid when both samples are large (CLT), or when both samples come from normally distributed populations.

Calculation of the observed test statistic:

We consider the following null hypothesis about the difference in means between two *independent* and *unpaired* samples: (Note the error in the book)

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

 $H_0: \ \delta = \delta_0,$

then the observed test statistic is:

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

Theorem 3.50: The distribution of the test statistic

The test statistic is (approximately) *t*-distributed:

Under the null hypothesis, the test statistic is:

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

It is approximately distributed as a t-distribution with v degrees of freedom, where

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

if the two populations are normally distributed or the sample sizes are sufficiently large.

- Null hypothesis $H_0: \delta = \delta_0 (= 0?)$
- Calculate: $d = \bar{x} \bar{y}$ and $s.e._d$
- Calculate test statistic: $t_{obs} = \frac{d \delta_0}{s.e.d}$
- Calculate p-value: $2 \cdot P(T > |t_{obs}|)$ (lookup in *t*-distribution)
- Compare p-value with desired lpha

- Go to today's Python notebook in VS Code
 - "Example: Nutrition study 2"



Overview

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- I-test with pooled variance An alternative
- Confidence interval for the differenceOverlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Over and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

t-test with pooled variance for two unpaired samples

The pooled variance estimate (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 test statistic in a *t*-test with pooled variance, Method 3.53

Consider the null hypothesis about the difference in means between two *independent* samples:

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

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

then the test statistic in a *t*-test with the pooled variance is given by:

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

Theorem 3.54: The distribution of the test statistic

Result

The test statistic in a *t*-test with pooled variance:

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

follows under the null hypothesis (and under the assumption $\sigma_1^2 = \sigma_2^2$) a *t*-distribution with $n_1 + n_2 - 2$ degrees of freedom, if the two populations are normally distributed.

We always use the "Welch" version

Somewhat (idiot)proof to always use the Welch version:

- If $s_1^2 = s_2^2$, then the two tests are the same. If the variances are very different, the two test statistics can be significantly different. In such cases, we do not necessarily prefer the test with the pooled variance, as the assumption of equal variances may be highly questionable.
- Only if the two variances are very different, can it happen that the two tests yield very different results? If the variances seem very different, the assumption of equal variances is likely violated.
- In cases with a small sample size in at least one of the groups, the Welch version is a more "cautious" approach.

Overview

- Summary from last week
- Motivating example: Nutrition study
- It-test with two samples
- t-test with pooled variance An alternative
- Confidence interval for the difference
 Overlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

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

Confidence interval for the difference in means:

For two samples (x_1, \ldots, x_{n_1}) and (y_1, \ldots, y_{n_2}) , the $(1 - \alpha)$ -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 $(1-\alpha/2)$ -quantile in the *t*-distribution with *v* degrees of freedom (given in Theorem 3.50).

$$d \pm t_{1-\alpha/2} \cdot s.e._d$$

It is the same "recipe" as last time.

Be sure to calculate v to insert as the degrees of freedom in the *t*-distribution.

- Go to today's Python notebook in VS Code
 - "Example: Nutrition study 3"



Overview

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- t-test with pooled variance An alternative
- 6 Confidence interval for the difference
- Overlapping confidence intervals?
- *t*-test with two paired samples (paired *t*-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Overlapping confidence intervals

Do you remember the overlapping confidence intervals?



Even though the two confidence intervals overlap slightly, we concluded that there is a significant difference.

Alternative presentation from the book

Bar Charts with error bars are commonly seen:

A bar chart with some error bars: Below are the 95% confidence intervals for each group:



Be cautious with "overlapping confidence intervals"

Note 3.59 – Rule for using "Overlapping confidence intervals": When two confidence intervals DO NOT overlap: The two groups are significantly different.

When two confidence intervals overlap: No conclusion can be drawn without examining the confidence interval for the *difference* between the groups.

Overview

- Summary from last week
- Motivating example: Nutrition study
- It-test with two samples
- t-test with pooled variance An alternative
- Confidence interval for the difference
- Overlapping confidence intervals?
- *t*-test with two paired samples (paired *t*-test)
- Normality assumptions
- Over and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Motivating example: Sleep medication

Difference in Sleep medication?

In a study, the aim is to compare two sleep aids, A and B. From 10 test subjects. The following results were obtained, indicating the extended sleep time in hours (the difference in the effect of the two drugs is noted):

Sample with n = 10:

Person	Α	В	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

Motivating example: Sleep medication

Difference in Sleep medication?

In a study, the aim is to compare two sleep aids, A and B. From 10 test subjects. The following results were obtained, indicating the extended sleep time in hours (the difference in the effect of the two drugs is noted):

Sample with n = 10:

Person	Α	В	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	$\bar{x} = 1.67$
6	+3.4	+4.4	+1.0	s = 1.13
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	

Example: Sleep medication

- Go to today's Python notebook in VS Code
 - "Example: Sleep medicine (Paired t-test)"



Experimental setup: Paired and independent samples

Completely Randomized (Independent Samples):

We have 20 patients, randomly assigned to two groups (typically equal numbers in each group). This means there are different (independent) patients in the two groups.

Paired Observations (Dependent Samples):

We have 10 patients, all of whom receive both treatments (typically with some time between and random order of treatments).

This means the same patients appear in both groups.

Overview

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- I t-test with pooled variance An alternative
- Sonfidence interval for the difference
- Overlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Normality assumptions



We revisit data on energy consumption among nurses at hospital A and hospital B.

QQ-plot

- Go to today's Python notebook in VS Code
 - "QQ-plots"



Overview

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- t-test with pooled variance An alternative
- 5 Confidence interval for the difference
- 6 Overlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples

Example - Light rail (letbane)

Let's revisit the Light Rail example. The data measures the voltage drop across a point where the voltage drop should be zero on average.



From the sample (n=10):
$$\bar{x} = 1.23$$

 $s = 1.62$

Confidence interval: $\bar{x} \pm t_{1-\alpha/2} \cdot (s/\sqrt{n}) = 1.23 \pm 1.16$ [0.07; 2.9]

In this case, we have: $\hat{\mu} = 1.23 \pm 1.16$ volts ($\alpha = 0.05$). We state that the **"Margin of Error" (ME)** is 1.16 (volts).

The conclusion from the sample was that we reject the null hypothesis — thus we conclude that the voltage drop is NOT zero (on average).

But the manager is not satisfied and says we must investigate the matter better.

Our margin of error of ± 1.16 volts is too large.

We plan to take a new sample.

How can we do better than the first time?

What can we do to improve our estimate of μ ?

Kahoot! (×1)

Experimental planning with precision requirements

Margin of Error (ME) is defined as

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

For sufficiently large samples, we can approximate $t_{1-\alpha/2}$ (from the t-distribution) with $z_{1-\alpha/2}$ (from the normal distribution).

Method 3.63: Sample Size for Confidence Interval Based on a Sample If σ is known, or estimated to be a certain value, we can calculate the required sample size to achieve a given margin of error, with probability $1 - \alpha$.

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

Let's estimate the sample size if we want ME to be at most 0.5 volts. We assume that s = 1.62 is a reasonable guess for the true σ .



Since we cannot take 40.3 measurements, we round up to 41 (in the new sample, n = 41).

Planning (Power and sample size)

We will now talk about experimental planning and "power".

To do this, we need to consider the two types of errors that we introduced last time:

Type I: Rejecting H_0 when H_0 is true.

 $P(\text{Type I error}) = \alpha$

Type II: Accepting (not rejecting) H_0 when H_1 is true.

 $P(\mathsf{Type II error}) = \beta$

The manager is now becoming a bit more demanding. She says that if the true average voltage drop is ± 0.5 volts (or more), we want to know it.

How should we go about it?

We will now go through some different scenarios.



Scenario 1: The null hypothesis is true

- If the true average voltage drop is zero
- and we take another round of measurements this time with a larger n
- then the risk of incorrectly rejecting the null hypothesis will be α (e.g., 0.05 or 5%)



The values of \bar{x} that we potentially accept as being "not significantly different from zero" lie in the range $0 \pm ME$. The width of this range depends on the size of the sample (n).

> Kahoot! (x2)

02402 Statistics (Polytechnical Foundation)

Scenario 2: The alternative hypothesis is true

- If the true average voltage drop is something other than zero the true value is μ_1
- and we take another round of measurements this time with a larger n
- What is then the risk (probability β) of accepting (not rejecting) the null hypothesis?



Yes, it depends on... ...how large μ_1 actually is.

Kahoot! (x2)

Spring 2025

Type I and Type II errors

Red Curve: Represents the null hypothesis (H0), which assumes no effect or no difference.

Green Curve: Represents the alternative hypothesis (H1), which assumes there is an effect or difference.



 α : the red area

(DTU Compute)

Experimental planning: Power



- Power = "Probability of detecting a (proposed) effect."
- Typical values for desired power are 80%, 90%, etc. (something relatively large)
- In practice: Use a scenario-based approach

Kahoot! (x3)

Experimental planning: Power

If we know (or assume) four out of the five following quantities, we can find the missing one:

- Sample size, *n*.
- Significance level, α , at which we are testing.
- Difference in means (effect size), $\mu_0 \mu_1$.
- Population standard deviation, σ .
- Power, 1β .

(whiteboard)

Experimental planning: Sample size *n*

The Big Question: How large should n be?

We need enough observations to detect a relevant effect with high power $1 - \beta$ (typically at least 80%):

Experimental planning: Sample size *n*

The Big Question: How large should n be?

We need enough observations to detect a relevant effect with high power $1-\beta$ (typically at least 80%):

Method 3.65: Formula for Sample Size with One Sample

For a t-test with one sample, where α , β , and σ are given:

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

Here, $\mu_0 - \mu_1$ is the difference in means that we want to measure, while $z_{1-\beta}$ and $z_{1-\alpha/2}$ are quantiles of the standard normal distribution.

We now tell the boss that we can design the next sample such that we will have an 80% probability of detecting a voltage drop of 0.5 volts (or more) at a significance level of $\alpha = 0.05$.

To calculate the necessary sample size, we use (as an informed guess): $\sigma = 1.62$, corresponding to the spread we observed in our previous measurements.



We find that the next sample needs to have 83 measurements:

$$n = \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1}\right)^2$$
$$= \left(1.62 \frac{z_{0.80} + z_{0.975}}{0.5}\right)^2$$
$$= 82.4$$

"83 measurements are too expensive," says the boss.

"Can we settle for 60 measurements?"

Yes... - but then our experiment will have lower power.



New power calculation:

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

$$\Rightarrow$$

$$z_{1-\beta} = \sqrt{60 \frac{0.5^2}{1.62^2}} - z_{0.975}$$

$$= 0.43$$

corresponding to a power of 0.67



- Go to today's Python notebook in VS Code
 - "Example: Power calculations"



Experimental planning: Two samples

The big question: How large should n_1 and n_2 be?

We can generalize the experimental planning to the situation where we compare two samples. The sample sizes are now given by n_1 and n_2 . We will assume that the variance (σ^2) is equal in the two populations.

Experimental planning: Two samples

The big question: How large should n_1 and n_2 be?

We can generalize the experimental planning to the situation where we compare two samples. The sample sizes are now given by n_1 and n_2 . We will assume that the variance (σ^2) is equal in the two populations.

Formula for Sample Size with Two Samples

For a t-test with two samples, where α , β , and σ are given:

$$n_1 = (k+1) \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_1 - \mu_2}\right)^2$$

Here, $\mu_1 - \mu_2$ is the difference in means that we want to measure, while $z_{1-\beta}$ and $z_{1-\alpha/2}$ are quantiles of the standard normal distribution. Additionally, $k = n_1/n_2$ is the ratio of the two sample sizes, so we have $n_2 = n_1/k$.



- Go to today's Python notebook in VS Code
 - "Example: Power calculations, 2 samples"



Agenda

- Summary from last week
- Motivating example: Nutrition study
- I t-test with two samples
- t-test with pooled variance An alternative
- Sonfidence interval for the difference
- Overlapping confidence intervals?
- t-test with two paired samples (paired t-test)
- Normality assumptions
- Power and sample size Experimental design
 - Precision requirements
 - Power and sample size One sample
 - Power and sample size Two samples