Contents

Inferential Statistic

We will be starting now to cover inferential statistics. Its objective is to use sample data to obtain results about the whole population.

In a first step the goal is to describe an underlying population. Since the populations are described in form of models, that are characterized by parameters (mean μ and standard deviation σ or probability p for the event of interest) our goal will be to estimate those characteristics or target parameters.
There are two different approaches for estimating: Point Estimation and Interval Estimation.

- For Point Estimation you give one value for a characteristic, which is hopefully close to the true unknown value. We can not expect to find the precise value describing the population when only using data of a sample.
- For Interval Estimation you give an interval of likely values, where the width of the interval will depend on the confidence you require to have in this interval.

See how to use statistics (values based on sample data) to obtain reasonable values for the population characteristics.
Since we base our statements just on a sample we see later how to give a measure of accuracy or confidence for the estimate.

1 Estimation of a population mean μ

1.1 Point Estimator for μ

Definition:
A point estimate of a possible characteristic is a single number that is based on sample data and represents a plausible value of the characteristic.

Example:

- The sample mean $\bar{x}$ is a point estimate for the population mean μ.
- The sample standard deviation s is a point estimate for the population standard deviation σ.
- The sample proportion $\hat{p}$ is a point estimate of p the population probability for Success.

Example:

- To estimate the average height μ of students in this class, we take a sample of size 10 and calculate a sample mean of $\bar{x}$ =172.9cm. We estimate the mean height in this class is 172.9cm!
- The sample standard deviation s in the sample of 10 students from this class is s =9.3cm. We estimate that the population standard deviation σ of the height in this class is 9.3cm.

- To estimate the probability p to survive a specific cancer treatment we look at a sample of size 250. 200 of the patients survived the first five years. The sample proportion $\hat{p} = 200/250 = 0.8$ is the estimate for p.

A point estimate gives a single value that is supposed to be close to the true value of the characteristic but it doesn't tell how close the estimate is.
One desirable property of an estimator is that the mean of it's distribution equals the parameter it is supposed to estimate.

Definition:
An estimator is said to be *unbiased estimator* for a parameter if the mean of its distribution is equal to the true value of the parameter. Otherwise is said to be biased.

The sample mean $\bar{x}$ of n observation is an unbiased estimator for the population mean μ, since we saw that $\mu_{\bar{x}} = \mu$.

But every single observation is also an unbiased estimate!
Intuition tells us that the sample mean is the better estimator, but why?

Remark:
Given a choice between several unbiased statistics for a given population characteristic, the best statistic to choose is the one with the **smallest** standard deviation of its distribution.

Since the standard deviation of the the mean $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ and the standard deviation of a single observation is σ, this remark leads us to choose the sample mean as the better estimator (unbiased statistic). The intuition was right!

Remark:
The sample standard variance

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

is an unbiased estimator for estimating the population variance σ^2. In fact, the denominator has to be $n-1$ in order for this statistic to be unbiased. (*This statement doesn't imply that* $s = \sqrt{s^2}$ *is an unbiased estimator for* σ*, in fact it usually under estimates the true value of* σ*.*)

Definition:
The distance between an estimate and the true parameter is called *the error of estimation.*

Definition:
The *standard error* of a statistic is the standard deviation of the statistic.

Remark: For unbiased estimators, the error of estimation will be most likely (with probability 0.95 for normal distributions) less than 1.96 standard errors (SE)–(Compare Empirical Rule). But on the other hand we find that for large populations $\mathrm{P}(\bar{x} = \mu) = 0$, a frustrating result, because we are 100% certain that the value we give is wrong. We only know that $\bar{x}$ should be close to μ, but again do not know how close.
To deal with this dilemma we give an interval for estimating μ instead of just one value.

Summary:

Estimation of a population mean μ

To estimate the population mean μ, the point estimator $\bar{x}$ is unbiased, with the standard error estimated as

$$SE = \frac{s}{\sqrt{n}}$$

These claims hold as long as the sample size is large, so that the Central Limit Theorem can be applied, that is $n \geq 30$

1.2 Confidence Intervals

As an alternative to point estimation we can report not just a single value for the population characteristic, but an entire interval of reasonable values based on sample data. A measure of confidence will be connected to such an interval.
For example we could give

$$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$$

to estimate μ.
Then the chances that we capture μ with this interval is about 95% (it actually 0.9544)
If we do not use 3, but 3 as a factor in the interval this chance will increase, or if make the facto smaller the probability to capture μ will also decrease.
In general:

Definition: A $(1-\alpha)$–*confidence interval* for a population characteristic is an interval of values for the characteristic. It is constructed so that, the probability for the true value of the characteristic to be captured in the interval equals the confidence level $(1-\alpha)$.

Remark:

- The confidence level provides information on how much confidence we can have in the **method** used to construct the interval estimate.

  If we would use the method for different samples, $1-\alpha$ gives the probability, that the true value falls into the calculated intervals.

- You also can give the confidence level in percent.

- Usual choices for the confidence level are 90%, 95%, or 99%.

- Most confidence intervals are of the form

  (point estimator) ± margin of error

- Commonly used critical values

Confidence Coefficient $(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
0.90	0.1	0.05	1.645
0.95	0.05	0.025	1.96
0.99	0.01	0.005	2.58

1.2.1 Large-Sample z -Confidence Interval for a Population Mean μ

If we use the statistic $\bar{x}$ for estimating the population mean μ, we can use the following information from the Central Limit Theorem in order to obtain a confidence interval for μ.

- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ standard error of $\bar{x}$.
- If $n \geq 30$, we can assume that the sampling distribution of $\bar{x}$ is approximately normal.

This leads to the following confidence interval for the population mean μ.

The One-Sample z Confidence Interval for μ
If $n \geq 30$ and the standard deviation σ is known an $(1-\alpha)$ confidence interval for the population mean μ is given by

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

With $z_{\alpha/2}$ being the $(1-\alpha/2)$ percentile of the standard normal distribution (Table IV).

Usually σ is **unknown**. In the case, that σ is unknown, it can be approximated by the sample standard deviation s when the sample size is large ($n \geq 30$) and the approximate confidence interval is

$$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Proof:
Start with

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

which as a result from the Central Limit Theorem is standard normal distributed.
With $z_{\alpha/2}$ being the $(1-\alpha/2)$ percentile of the standard normal distribution, we get

$$P(-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

or equivalent

$$P(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

or equivalent

$$P(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

That is, when we randomly choose a sample and use above formula to find the interval, then the probability that μ falls within the interval equals $1 - \alpha$.

Example:
A scientist interested in monitoring chemical contaminants in food, and thereby the accumulation of contaminants in human diets, selected a random sample of $n = 50$ male adults. It was found that the average daily intake of dairy products was $\bar{x} = 756$grams with a standard deviation of $s = 35$grams.

An approximate 95% confidence interval for the mean daily intake of dairy products for men is then:

$$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$756 \pm 1.96\left(\frac{35}{\sqrt{50}}\right)$$

$$756 \pm 9.70$$

Hence, the 95% confidence interval for μ is from 746.30 to 765.70 grams per day.
The true mean daily intake of diary products for men is with confidence 0.95 in the interval from 746.30 to 765.70 grams per day.

Remember:
Being "95% confident" means, if you were to construct 100 95% confidence intervals from 100 different random samples. Of the 100 intervals you expect 95 to capture the true mean, and 5 not to capture the mean.
In conclusion, you can not be sure that a specific confidence interval captures the true mean μ.

1.2.2 Choosing the Sample Size

One of the important decisions, before drawing a sample, is how many experimental units from the population should be sampled. That is: what is the appropriate sample size?
The answer depends on the specific object of investigation and the precision or accuracy one wants to insure. A measure for the accuracy in estimation is the margin of error.

Argument: Suppose you want to estimate the average daily yield μ of a chemical process and you want to insure with a high level of confidence that the estimate is not more than 4 tons of the true mean yield μ.

In this situation you would require that the sampling error of $\bar{x}$ in a $(1-\alpha)100\%$ confidence interval is less than 4 tons.

This will ensure, that if you would take 100 samples the distance between the true mean and the sample mean from about $(1-\alpha)100$ samples will be at most 4 tons.

In general the researcher chooses the largest value SE that is acceptable for the error indicated by a confidence interval.

Then the researcher determines what confidence level $(1-\alpha)$ he wants to attain in his claims in the study.

From this the necessary sample size can be determined. We require that the margin of error in a $(1-\alpha)$ confidence interval is less or equal than SE.

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq SE \Leftrightarrow \left(\frac{z_{\alpha/2}\sigma}{SE}\right)^2 \leq n$$

Go back to the example. Plan to do a 95% confidence interval for μ, where we allow a margin of error not greater than $SE = 4$.

At this point we still do not know σ, the standard deviation of the daily yield of this chemical process.

If σ is unknown, what is the realistic case, you can use the best approximation available:

- An estimate s obtained from a previous sample.
- A range estimate based on knowledge of the largest and smallest possible measurement: $\sigma \approx \text{Range}/4$.

In this example assume a previous sample would have shown a sample standard deviation of $s = 21$tons. Then

$$n \geq \left(\frac{z_{\frac{\alpha}{2}}\sigma}{SE}\right)^2 = \left(\frac{1.9621}{4}\right)^2 = 105.8$$

We obtain that the sample size has to be at least 106 in order to estimate μ with a 95% confidence interval, with a margin of error smaller than 4.

Find that this result is only approximate since we had to use an approximation for σ, but this is still better than just choosing any number.

Example:
The financial aid office wishes to estimate the mean cost of textbooks per quarter for students at a particular college. For the estimate to be useful, it should be used be within \$20 of the true population mean. How large a sample should be used to be 95% confident of achieving this level of accuracy?
The financial aid knows that the amount spent varies between \$50 and \$450.
A reasonable estimate of σ is then

$$\frac{\text{range}}{4} = \frac{450 - 50}{4} = 100$$

The required sample size is

$$n \geq \left(\frac{1.96\sigma}{SE}\right)^2 = \left(\frac{1.96 \cdot 100}{20}\right)^2 = 9.8^2 = 96.04$$

So that in this case a sample size of at least 97 is required.

1.2.3 t-confidence interval for a mean μ

The problem with the large sample confidence interval for μ is that it requires us to know σ the population standard deviation. This assumption is strong and never met.
For that reason we should replace the large sample confidence interval with an alternative, that does not require σ.

So far the confidence interval was based on

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Now we will replace σ by the sample standard deviation s, which gives us

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

this score follows a t-distribution with $n-1$ degrees of freedom if the sample size is large, or the population is normally distributed. The t-distribution is described in table VI in the textbook. The t-distribution is also called "Student's t-distribution". It was introduced by a mathematician called W.S. Gosset in 1908, who used the pen name "Student".

Student's t distribution
Consider the t-score

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The distribution of the t-score only depends on one parameter, which is called the degrees of freedom (df). "Student" showed that the t-score is t distributed with $n-1$ degrees of freedom ($df = n-1$). The appendix provides a table (Table VI) with values from this distribution for different choices for the df.
The table gives uppertail areas.

t-confidence interval for μ

$$\bar{x} \pm t^{n-1}_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t^{n-1}_{\alpha/2}$ is the $1-\alpha/2$ percentile of the t-distribution with $df = n-1$.

Example:
A scientist is interested in monitoring the daily intake of dairy products in a population.
A sample of $n = 50$ people let to a sample mean of $\bar{x} = 756$g with a standard deviation of $s = 35$g.
We will find a 95% confidence interval for μ =the mean daily intake of dairy products in this population.
$\alpha = 0.05$, so $\alpha/2 = 0.025$ (upper tail area needed for finding the percentile in table VI), $df = n-1 = 49$, from table VI find $t^{40}_{(0.975)} = 2.021$ (use df=40), the largest value that is smaller than the true df.

$$756 \pm 2.021\left(\frac{35}{\sqrt{50}}\right) \to 756 \pm 10.002 \to [745.998, 766.002]$$

We are 95% confident that the mean daily intake of dairy products in this population falls between 746g and 766g.

2 Statistical Tests of Hypotheses

Previously population characteristics were described, now we will be checking if claims about the population characteristics are true, or plausible to a given degree,
Since this is statistics and decisions about the population are based on samples, we might make errors when making decisions. You will learn how to control the probabilities to make errors.

A hypothesis test is a method for using sample data to decide between two competing claims or hypotheses about a population characteristic.

Example:

$$p \leq 0.5 \quad \text{vs.} \quad p > 0.5$$

$$\mu = 100 \quad \text{vs.} \quad \mu \neq 100$$

Definition:

> The null hypothesis H_0 is a claim about a population characteristic. (We will try to disprove this hypothesis with the help of sample data)
>
> The alternative hypothesis H_a is the competing claim and logical compliment of H_0. (When we can disprove H_0, then H_a must be correct).

In testing H_0 vs. H_a:

- H_0 will be rejected only if the evidence from the sample strongly suggests that H_0 is false.
- Otherwise H_0 will not be rejected, and we will state that we could not find evidence against the claim.

So there are two possible conclusions:

- reject H_0 (accept H_a)
- do not reject H_0

Note that these decisions are not symmetric, there is no way you can say you accept H_0.

Remark:
Hypotheses should be the logical compliment of each other.
Common choices of hypotheses are

- Two-tailed Test
  H_0: population characteristic = specific value versus
  H_a: population characteristic $\neq$ specific value
- Upper-tailed Test
  H_0: population characteristic $\leq$ specific value versus
  H_a: population characteristic $>$ specific value

- Lower-tailed Test
  H_0: population characteristic $\geq$ specific value versus
  H_a: population characteristic $<$ specific value

In the text book they always choose "H_0: population characteristic = specific value", which they argue is equivalent to the other null hypotheses. The decision would be the same but not the underlying logic.

Examples:

- H_0: $p = 0.25$ versus H_a: $p \neq 0.25$
- H_0: $\mu \geq 100$ versus H_a: $\mu < 100$
- We can not test H_0: $\mu \leq \mathbf{100}$ versus H_a: $\mu > \mathbf{150}$

Be careful when choosing hypotheses, because a statistical test can only support the alternative hypothesis, by rejecting H_0.

Is H_0 not being rejected doesn't mean strong support for H_0, but lack of strong evidence against H_0.

Example:
A company is advertising that the average lifetime of their light bulbs is 1000 hours. You might question this, and want to show that in fact the lifetime is shorter.
You would test H_0: $\mu \geq 1000$ versus H_a: $\mu < 1000$.
Rejection of H_0 would then support your claim. However, nonrejection of H_0 doesn't necessarily provide strong support for the advertised claim.

The way the decisions are made, the scientist will choose H_a to contain the claim he wants to prove.

How to make the decision (reject H_0, or do not reject H_0)

The decision to reject, or not to reject H_0 is based on information contained in a sample drawn from the population of interest. This information will be given in form of

- the test statistic (a number that measures, if the sample data is in accordance with H_0), or
- the P-value (the probability for observing the value of the test statistic, if H_0 is true) Assuming that H_0 is true the P-value measures how likely it is to observe such data, as those found in the sample.

  If the P-value is small, this indicates that the assumption, that H_0 is true, is (probably) wrong. Is the P-value not small, this indicates that the sample does not provide evidence against H_0.
- use the test statistic or the P-value to make a decision.

Example:
Do students learn statistics online as well as face-to-face?
A standardized test shows has been built such that students who are taught face-to-face score on average 100 on this test with a standard deviation of $\sigma = 15$
A random sample of 49 students from an online statistics course has taken the standardized test and they scored on average $\bar{x} = 96$. Do these data indicate that the mean performance of online students (population) score lower on the standardized test?
Let μ be the mean score of online students on this test.
The we want to decide between the following hypotheses
$H_0 : \mu \geq 100$ versus $H_a : \mu < 100$.
The *sample* mean from 49 students is smaller than 100, but is this enough evidence to conclude that this is true in general (for the population)? Or can this be explained by the sampling variability?
To find out, we calculate the test statistic, that will "compare" the sample value $\bar{x}$ with the value from the**null hypothesis** $\mu_0 = 100$.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{96 - 100}{15/\sqrt{49}} = -1.8667$$

Is this value indicating that a mean of 96 in a sample of 49 is unlikely to occur if $mu \geq 100$? The P-value will give the answer!

$$\begin{aligned} \mathrm{P-value} &= P(z \leq -1.8667) \text{ (probability statement)} \\ &\approx 0.0301 \qquad \text{(NormalTable)} \end{aligned}$$

If H_0 is true, i.e. students learning statistics online score on average at least as high as face-to-face students, then only 3% of samples of size 49 would on average score 96.
Since this is a very small chance we will probably decide to reject H_0, and conclude that online students do NOT learn as much as face-to-face students.

(I made the numbers up, and the question remains open!!!!)

The decision would be easy if the P-value equals 0, the question remains for which P-values should the null hypothesis be rejected?
In order to answer this question have a look at the different types of errors that may occur while testing.

2.1 Errors in Hypothesis Testing

As there are in criminal trials, there are two different types of errors you can make in statistical testing:
In a trial the jury might convict an innocent person, and the other error is to set a guilty person free.

Definition:

type I error – the error of rejecting H_0 even though H_0 is true

type II error – the error of failing to reject H_0 even though H_0 is false

		Truth	
		H_0 is true	H_0 is false
Test	reject H_0	type I error	OK
	do not reject H_0	OK	type II error

The only way to guarantee that neither type of error will occur is to make such decisions on the basis of a census of the entire population. The risk of error is introduced when we try to make an inference on a sample.

Definition:
The probability of a type I error is denoted by α and is called the level of significance of the test.
The probability of a type II error is denoted by β.

We would like to ensure with the choice of the method, telling us how to make a decision, that both error probabilities are small.
But a mathematical analysis shows that how ever we are making the decision between H_0 and H_a the error probabilities behave like a seesaw. When we force one to be small the other goes up.
Due to this relationship between the error probabilities, one had to choose to control one and let the other go. It was decided to make sure with the choice for a hypothesis that the P(error of type I) will be be small.

Remark: After assessing the consequences of type I and type II errors identify the largest α that is tolerable for the problem. Don't use a too small level of significance, because the smaller α the greater β.

Decision Rule: A decision as to whether H_0 should be rejected results now from comparing the P-value to the chosen α.

- H_0 should be rejected if P-value $\leq \alpha$.
- H_0 should not be rejected if P-value $> \alpha$.

Example:
A drug is proposed to lengthen the survival time after a specific cancer treatment.
To show the efficacy of the new drug a study has to be designed to test the following hypotheses for μ the mean survival time under the new treatment.

$$\begin{gathered} H_0 : \mu \leq \text{mean survival time without new treatment} \\ \text{versus} \\ H_a : \mu > \text{mean survival time without new treatment} \end{gathered}$$

An error of type I would mean to conclude the drug is lengthening the survival time, even though this is not the case.

An error of type II would mean to conclude the drug not efficient even though it is.
The scientist doing the study, wants to make sure, that this drug is only used if it is really efficient, so she has to limit the probability for the error of type I. she chooses $\alpha = 0.01$.

2.2 A Large Sample Test for a Population Mean, when σ is known

The Test

1. **Hypotheses**:
   - two tailed: $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
   - lower tailed: $H_0 : \mu \geq \mu_0$ versus $H_a : \mu < \mu_0$
   - upper tailed: $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$

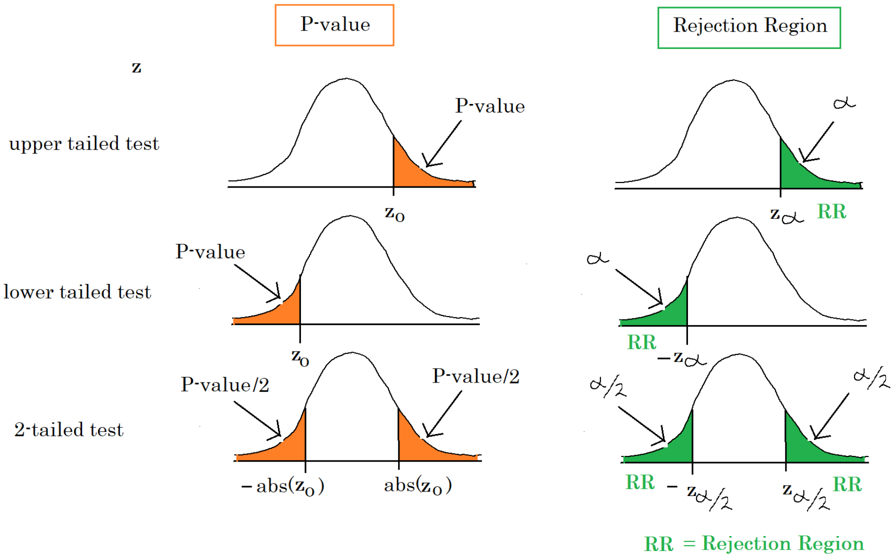
   Choose α.

2. **Assumption**: (1) The data represents a random sample, (2) the sample size is large OR the data origins from a normal population, and (3) σ is known.

3. **Test statistic:** $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ estimated by $z_0 \approx \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

4. **P-value/Rejection Region:** – to make a decision you only need one of the two

Test type	**P-value**	**Rejection Region**
Upper tail	$P(z > z_0)$	$z_0 > z_\alpha$
Lower tail	$P(z < z_0)$	$z_0 < -z_\alpha$
Two tail	$2 \cdot P(z > abs(z_0))$	$abs(z_0) > z_{\alpha/2}$



5. **Decision:**

   If $p - value \leq \alpha$ or equivalently the value of the test statistic falls into the rejection region, then reject H_0.

   If $p - value > \alpha$ or equivalently the value of the test statistic does not fall into the rejection region, then do not reject H_0.

6. **Context:** Put the result into context.

   At significance level of α the data provide/do not provide sufficient evidence that H_0 is wrong.

   - provide = test is significant, i.e. H_0 is rejected
   - do not provide = test is not significant, i.e. H_0 is not rejected
   - declare what it means that H_0 is wrong

Example: Assume you have a random sample with $n = 50$, $\bar{x} = 871$ and $\sigma = 21$.
Test at a significance level of $\alpha = 0.05$ the hypotheses:

1. Hypotheses/α:

   $H_0 : \mu = 880$ versus $H_a : \mu \neq 880$, $\alpha = 0.05$

   – two-tailed with $\mu_0 = 880$

2. Assumptions: (1) Random sample (stated above), (3) we know σ, and (2)we find that the sample size is large.

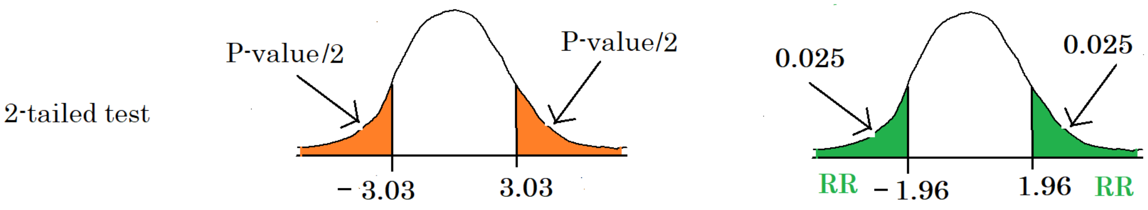
3. Test statistic: $z_0 \approx \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{871 - 880}{21/\sqrt{50}} = -3.03$

4. Rejection Region/Pvalue:(for illustration we do both, but we only need one or the other (your choice))

   Since this is a two-tailed test:

   Rejection Region: Using $\alpha = 0.05$ the rejection region is equal to $abs(z_0) > z_{\alpha/2} = 1.96$.

P-value: p-value$= 2 \cdot P(z > abs(z_0)) = 2 \cdot P(z > 3.03) = 2 \cdot (1 - 0.9988) = 0.0024$, using the Normal Table.



5. Decision:

   Rejection Region: It is $abs(z_0) = abs(3.03) = 3.03 > 1.96$, reject H_0

   P-value: p-value$= 0.0024 < 0.05 = \alpha$, reject H_0

   (same decision using both approaches, as we should have expected, they always lead to the same decision, that is why we only need one)

6. At significance level of 5% the data provide sufficient evidence that $\mu \neq 880$.

Definition:
The p-value of a statistical test is the probability to observe the value of the test statistic (or one more extreme) if in fact H_0 is true.
When you look as the test statistic as representing the sample data, one can also say that the P-value measures how like the sample in front of us would have occurred if H_0 were true.

Decision Rule:

1. Find the Rejection Region: If the value of the test statistic falls into this region, reject H_0.

   or

2. Find the p-value: If $p - value \leq \alpha$ holds, reject H_0.

The assumption, that we know σ is very strong, since we already assume that we do not know μ. How come we do not know the mean but the standard deviation for the population of interest? For this reason we need a different tool, based on the t-distribution.

2.3 A test for a mean μ, when σ is unknown

The test introduces in the section above is based on the z-score, which uses the population standard deviation σ. In most situations σ is unknown and has to be replaced by the sample standard deviation s. Resulting in a procedure that then is only approximate (does not give the true error probability).

Student's t distribution
Consider the t-score

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

is t-distributed with$df = n - 1$, if the sample is large or the population follows a normal distribution.

The distribution of the *t*-score only depends on one parameter, which is called the degrees of freedom (df). "Student" showed that the *t*-score is *t* distributed with $n-1$ degrees of freedom ($df = n-1$). The appendix provides a table (Table VI) with values from this distribution for different choices for the *df*.

t-Test for a Population Mean μ

1. **Hypotheses:**

Test type	
Upper tail	$H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$
Lower tail	$H_0 : \mu \geq \mu_0$ versus $H_a : \mu < \mu_0$
Two tail	$H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$

   Choose α.

2. **Assumption:** The sample is a random sample and the population has a normal distribution or the sample is large.

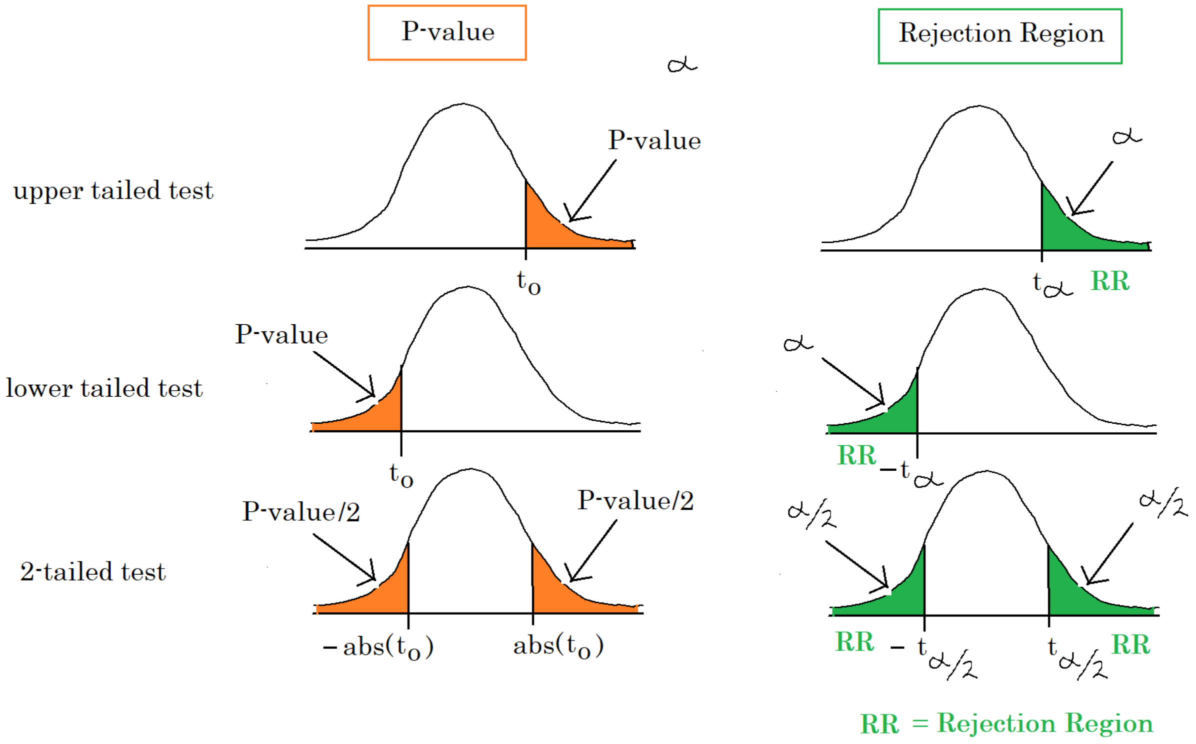
3. **Test statistic:**

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

   with $df = n-1$ degrees of freedom.

4. **P-value and Rejection Region:**

Test type	**P-value**	**Rejection Region**
Upper tail	$P(t > t_0)$	$t_0 > t_\alpha$
Lower tail	$P(t < t_0)$	$t_0 < -t_\alpha$
Two tail	$2 \cdot P(t > abs(t_0))$	$abs(t_0) > t_{\alpha/2}$



5. **Decision:**

   If p-value$\leq \alpha$ then reject H_0
   If p-value$> \alpha$ then do not reject H_0

   Or:
   If the value of the test statistic **falls** into the rejection region, then reject H_0.
   If the value of the test statistic **does not fall** into the rejection region, then do not reject H_0.

6. **Context**

$(1-\alpha)$ t-Confidence Interval for a Population Mean μ

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

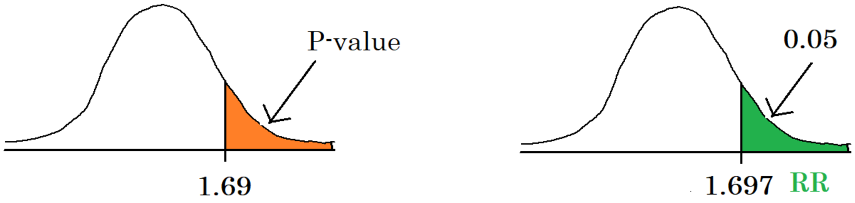
where $t_{\alpha/2}^{n-1}$ is the $(\alpha/2)$ critical value of a t-distribution with $df = n-1$.

Comparing with the z-interval that changes is that you will have to use the critical value of the t-distribution (table VI) and you may use the sample standard deviation instead of pretending you know σ.

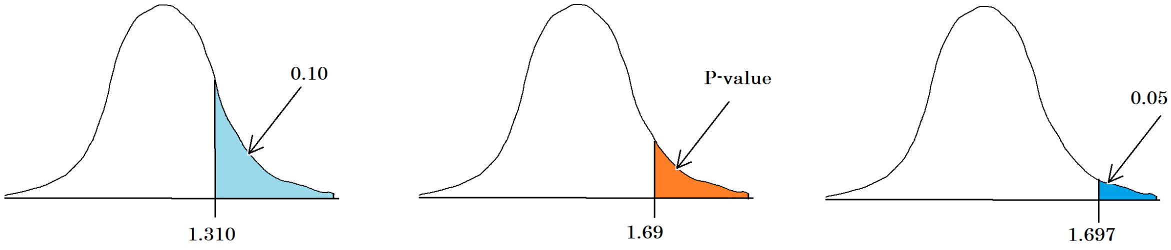
Example: In recent decades, the mean weight of human males, aged 18 to less than 75, has been 78.1 kg with a standard deviation of 13.5 kg.
In a study wether weights are changing, a researcher samples 40 males in that age group and obtains a mean of 82.3 kg with a standard deviation of 15.7 kg.
At significance level of 5% can the researcher conclude that the mean weight has increased?

1. $H_0 : \mu \leq 78.1$ versus $H_a : \mu > 78.1$, where μ is the mean weight of males aged 18 to 75. $\alpha = 0.05$.
2. The sample size is large enough, and we will assume that the participants were randomly chosen.
3. $$t_0 = \frac{82.3 - 78.1}{15.7/\sqrt{40}} = 1.69, df = 39$$
4. This is an upper tail test:

   upper tailed test

   P-value: p-value= $P(t > t_0 = 1.69)$, the upper tail probability. Using the t-table use $df = 39$.
   Observe that 1.69 falls between 1.685 and 2.023, giving that: 0.025 <p-value< 0.05 (from the column labels).

   => 0.05<P-value<0.10

   Rejection Region: The rejection region is in the upper tail (upper tail test), with $t_\alpha^{39} = 1.685$, i.e. $t_0 > 1.685$.
5. P-value approach: Since the p-value is less than $\alpha = 0.05$, do reject H_0.

   Or Rejection Region approach: $t_0 = 1.69 > 1.685$, and falls into the rejection region, so reject H_0.
6. At significance level of 5% that data do provide sufficient evidence that the mean weight of males aged between 18 and 75 increased lately.

Estimate the mean weight of human males aged 10 to 75 based on the sample with a 95% confidence interval

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

$n = 40, \bar{x} = 82.3, s = 15.7, \alpha = 0.05, \alpha/2 = 0.025, df = 39$, from table using df=30 $t_{0.025} = 2.023$

$$82.3 \pm 2.023 \frac{15.7}{\sqrt{40}} \quad \leftrightarrow \quad 82.3 \pm 5.069 \quad \leftrightarrow \quad [77.231; 87.369]$$

Based on the sample data we are 95% confident that the mean weight of males from the population falls somewhere between 77.231kg and 87.369kg.

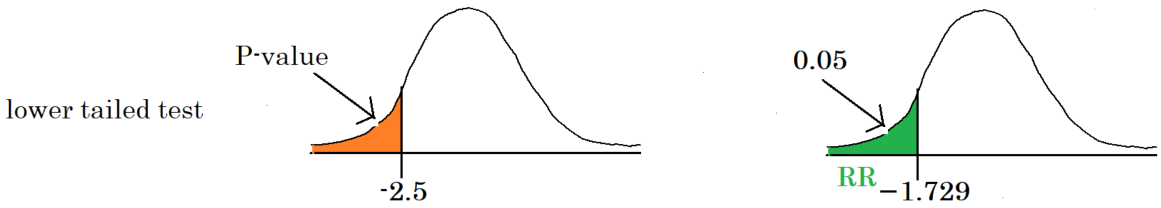
Example:
An industrial plant has to demonstrate that they in compliance with the law and provide evidence that they discharge on average less than 1000 liters of waste water per hour. They sample 25 hours in a week, and observe that on average they discharge $\bar{x} = 950l/h$ with a standard deviation of $s = 100l/h$.
Do the data provide sufficient evidence that they discharge less than 1000l of waste water per hour?

1. $H_0 : \mu \geq 1000$ versus $H_a : \mu < 1000$, where μ is the mean amount of waste water discharged per hour. Choose $\alpha = 0.05$.

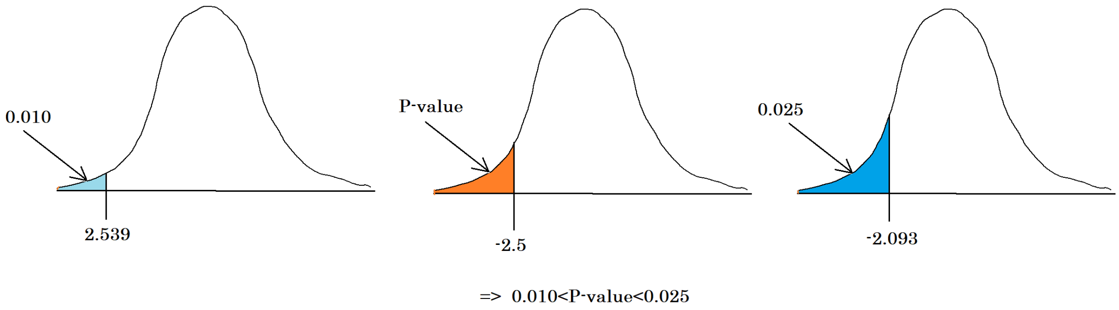
2. They randomly chose the times (random sample), and know from previous experiments that they can assume the discharge to be normally distributed.

3. $$t_0 = \frac{950 - 1000}{100/\sqrt{25}} = \frac{-50}{20} = -2.5, df = 19$$

4. This is a lower tail test:



P-value: p-value= $P(t < t_0 = -2.5) = P(t > 2.5)$, the lower tail probability (using symmetry gives the upper tail probability). According to the t-table using $df = 19$ we observe that 2.5 falls between 2.093 and 2.539, giving that: 0.010 <p-value< 0.025 (from the column labels).



Rejection Region: The rejection region is in the lower tail (lower tail test), with $t_\alpha^{19} = 1.729$, i.e. $t_0 < -1.729$.

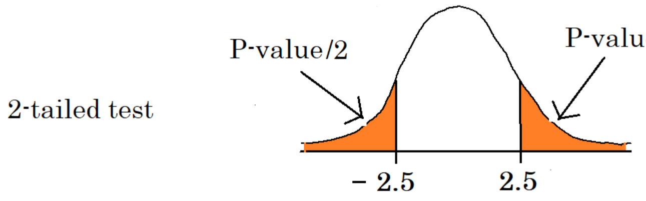
5. P-value approach: Since the p-value $< 0.025 < 0.05 = \alpha$, and therefore less than $\alpha = 0.05$, reject H_0.

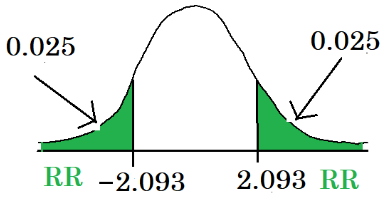
   Or Rejection Region approach: $t_0 = -2.5 < -1.729$, and falls into the rejection region, reject H_0.

6. At significance level of 5% that data provide sufficient evidence that the mean amount of waste water discharged per hour falls below 1000l.

Demonstration:
If we would have used the data in the previous example to test if the data provide sufficient evidence that the mean amount of waste water discharged per hour is **different** from 1000l. We should have done the following test

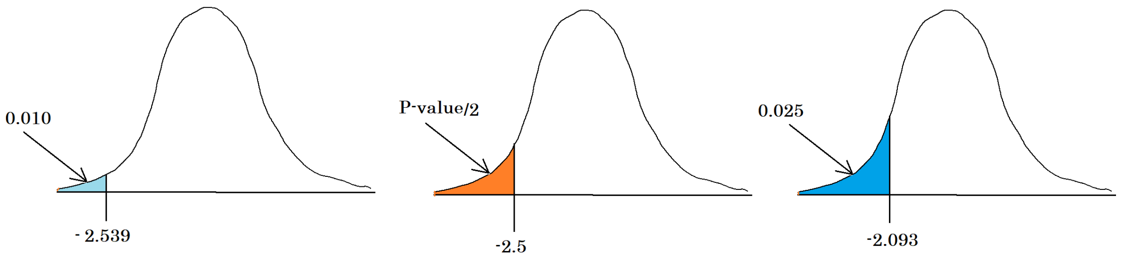
1. $H_0 : \mu = 1000$ versus $H_a : \mu \neq 1000$ Choose $\alpha = 0.05$.
2. (no change) They randomly chose the times (random sample), and know from previous experiments that they can assume the discharge to be normally distributed.
3. (no change)
$$t_0 = \frac{950 - 1000}{100/\sqrt{25}} = \frac{-50}{20} = -2.5, df = 19$$
4. This is a 2-tailed test:





P-value: p-value= $2 \times P(t > abs(t_0) = 2.5) = 2 \times P(t > 2.5)$, two times the upper tail probability for the absolute value.

According to the t-table using $df = 19$ we observe that 2.5 falls between 2.093 and 2.539, giving for half the P-value: $0.010 <$p-value/2< 0.025 (from the column labels), and therefore $0.020 <$p-value< 0.05.



=> 0.010 < P-value/2 <0.025 => 0.020 < P-value < 0.05

Rejection Region: The rejection region is both tails (2-tailed test), with $t^{19}_{\alpha/2} = 2.093$, i.e. $abs(t_0) > 2.093$.

5. P-value approach: Since the p-value $< 0.05 = \alpha$, i.e. less than $\alpha = 0.05$, reject H_0.

   Or Rejection Region approach: $abs(t_0) = 2.5 > 2.093$, and falls into the rejection region, reject H_0.

6. At significance level of 5% that data provide sufficient evidence that the mean amount of waste water discharged per hour is different from 1000l.

Wording, when a test results in not rejecting H_0:
Assume that in the last example we would not have rejected $H_0 : \mu < 1000$, then we would have come to the following conclusion:
At significance level of 5% the data do not provide sufficient evidence that the mean amount of waste water falls below 1000l per hour.

3 Comparison of Two Populations

In this section we will study how to use inferential statistics for comparing two populations. Again we will consider being interested in a mean of a population and in the proportion of a population.

3.1 Inference for the difference between two population means, $\mu_1 - \mu_2$

Consider a study for comparing the effect of two drugs on the blood pressure. If μ_1 and μ_2 are the mean changes in the blood pressure for drug one and two, respectively, then here one might want to compare the two means based on sample data from the two populations including people being treated with one or the other drug. This would call for a method using the information in the two samples and estimating and testing properties for the two means.

The sample data will give us:

	Sample 1	Sample 2
sample size	n_1	n_2
mean	$\bar{x}_1$	$\bar{x}_2$
s.d.	s_1	s_2

When comparing two population means we need to distinguish the following two situations describing the type of samples we will base the comparison on.

1. independent samples – example: height of males and females, where the samples of males and females were independently obtained.

   The measurements from the two samples under investigation are not associated with each other, they are independent.

2. paired samples – example: blood pressure measured in the morning and evening in the same group of people

   The measurements from two samples are connected. For every measurement in sample one there is a corresponding measurement in sample two (in the example: for every individual, their measurement in sample one corresponds to their measurement in sample two)

3.1.1 t-Procedure for Two Paired Samples

In order to control extraneous factors in some studies paired samples are used. In this case for every individual in the sample from population 1 you find a matching individual from population 2. And the decision is made based on the resulting sample data. In these cases the sample sizes are always the same, $n = n_1 = n_2$.

Example:

- Compare the resting pulse and pulse after exercise.

  To control for all other influences, take two measurements for every individual resulting in two samples (before and after exercise).

To compare the two population mean we use $\mu_d = \mu_1 - \mu_2$, the difference between the two population means, as parameter of interest. This one number can tell us everything:

- $\mu_d = \mu_1 - \mu_2 = 0$ means $\mu_1 = \mu_2$
- $\mu_d = \mu_1 - \mu_2 < 0$ means $\mu_1 < \mu_2$
- $\mu_d = \mu_1 - \mu_2 > 0$ means $\mu_1 > \mu_2$

In case of paired sample, to conduct statistical inference about μ_d the differences of the paired observations are used

$$\text{sample 1 value} - \text{sample 2 value}$$

are used. Which then will create **one** sample of size n of measurements of pairwise differences.

For example consider paired samples of size 5:

x_1	x_2	$x_1 - x_2$
5	9	-4
7	3	4
4	4	0
9	7	2
7	9	-2

$\bar{x}_d$ and s_d denote the sample mean and sample standard deviation for those differences, respectively.

For the distribution of $\bar{X}_d$, we get equivalent results as seen in the Central Limit Theorem for $\bar{X}$:

1. $\mu_{\bar{X}_d} = \mu_1 - \mu_2$, $\bar{X}_d$ is an unbiased estimator for $\mu_1 - \mu_2$
2. $\sigma_{\bar{X}_d} = \sigma_d/\sqrt{n}$, where σ_d is the population standard deviations of the pairwise differences.
3. If n is large than $\bar{X}_d$ is normally distributed.

Resulting in the following t-score standardizing $\bar{X}_d$:

$$t = \frac{\bar{X}_d - (\mu_1 - \mu_2)}{s_d/\sqrt{n}}$$

is t-distributed with $df = n - 1$, if either n is large or the population of pairwise differences is normally distributed.

These facts results to the following

Paired t-Test for Comparing Two Population Means

1. **Hypotheses:**

Test type	
Upper tail	$H_0 : \mu_d \leq d_0 \Leftrightarrow \mu_1 - \mu_2 \leq d_0$ versus $H_a : \mu_d > d_0 \Leftrightarrow \mu_1 - \mu_2 > d_0$
Lower tail	$H_0 : \mu_d \geq d_0 \Leftrightarrow \mu_1 - \mu_2 \geq d_0$ versus $H_a : \mu_d < d_0 \Leftrightarrow \mu_1 - \mu_2 < d_0$
Two tail	$H_0 : \mu_d = d_0 \Leftrightarrow \mu_1 - \mu_2 = d_0$ versus $H_a : \mu_d \neq d_0 \Leftrightarrow \mu_1 - \mu_2 \neq d_0$

   Choose α.

2. **Assumption:** Random sample of differences, and n is large or the population distribution of differences is approximately normal.

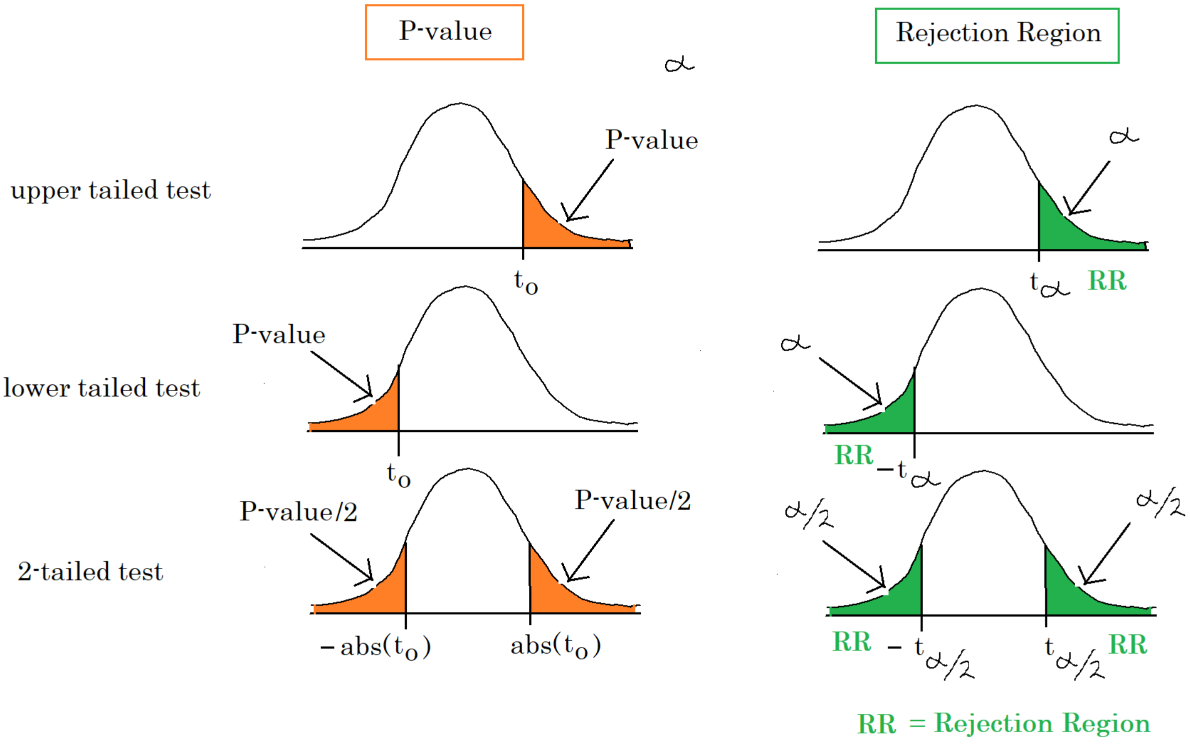
3. **Test statistic:**

$$t_0 = \frac{\bar{x}_d - d_0}{\frac{s_d}{\sqrt{n}}}$$

   with $n - 1$ df.

4. **P-value/Rejection Region:**

Test type	**P-value**	**Rejection Region**
Upper tail	$P(t > t_0)$	$t_0 > t_\alpha$
Lower tail	$P(t < t_0)$	$t_0 < -t_\alpha$
Two tail	$2 \cdot P(t > abs(t_0))$	$abs(t_0) > t_{\alpha/2}$



5. **Decision:**

   Either:
   If the P-value$\leq \alpha$ reject H_0
   If the P-value$> \alpha$ do not reject H_0

   Or:
   If t_0 falls into the Rejection Region reject H_0
   If t_0 falls outside the Rejection Region do not reject H_0

6. **Context:**

d_0 is the value we are comparing the difference against. Very often $d_0 = 0$ is used.

- Do the data indicate that the mean before is greater than the mean after?

  For $\mu_d = \mu_{before} - \mu_{after}$ test $H_0 : \mu_d \leq 0$ versus $H_a : \mu_d > 0$ $(d_0 = 0)$

- Do the data indicate that there is a difference between the two means?

  For $\mu_d = \mu_{before} - \mu_{after}$ test $H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$ $(d_0 = 0)$

- But: Do the data indicate that the difference between the two means is less than 20? (or the mean before exceeds the mean after is less than 20?)

  For $\mu_d = \mu_{before} - \mu_{after}$ test $H_0 : \mu_d \geq 20$ versus $H_a : \mu_d < 20$ $(d_0 = 20)$

Beside *testing* for difference of the means it is always insightful to estimate $\mu_d = \mu_1 - \mu_2$ with a confidence interval:

Paired t-Confidence Interval for $\mu_1 - \mu_2$

Assumption: n is large or the population distribution of differences is approximately normal.

The $(1-\alpha)$ **Confidence Interval for** μ_d**:**

$$\bar{x}_d \pm t^{n-1}_{(\alpha/2)} \frac{s_d}{\sqrt{n}}$$

and $t^{n-1}_{(\alpha/2)}$ is the critical value of the t-distribution with $n-1$ degrees of freedom (Table 4).

Example:
The effect of exercise on the amount of lactic acid in the blood was examined.
Blood lactate levels were measured in eight males before and after playing three games of racquetball.

Player	Before	After	Difference
1	13	18	-5
2	20	37	-17
3	17	40	-23
4	13	35	-22
5	13	30	-17
6	16	20	-4
7	15	33	-18
8	16	19	-3

This data results in $\bar{x}_d = -13.63$, $s_d = 8.28$, $n = 8$
Lets test whether these data indicate a significant increase in mean lactate levels at a significance level of 0.05. That is

1. $$H_0 : \mu_b - \mu_a \geq 0 \ (\mu_d \geq 0) \ vs. \ \ H_a : \mu_b - \mu_a < 0 \ (\mu_d < 0)$$

   .

   where μ_b (μ_a) is the mean lactate level before (after) three games of racquetball.

   $\alpha = 0.05$.

2. **Assumption:** The sample is a random sample and it is appropriate to assume that the difference in lactate level is normally distributed.

3. **Test statistic:** with $d_0 = 0$

$$t = \frac{\bar{x}_d - d_0}{\frac{s_d}{\sqrt{n}}} = \frac{-13.63}{\frac{8.28}{\sqrt{8}}} = -4.65597$$

   with $df = n - 1 = 7$.

4. **P-value:**
   Since we perform a lower tail test the P-value$=P(t < t_0) = P(t > abs(t_0)) = P(t > 4.65597)$.

   Use table VI from the text book. Focus on row with $df = 7$ find the 2 numbers that enclose $abs(t_0) = 4.656$.

   Here: $3.499 < 4.656 < 4.785$. Then the P-value falls between the two upper tail probabilities corresponding to these values.

   Here: $0.001 <$P-value< 0.005.

5. **Decision:** Since P-value$< 0.005 < \alpha = 0.05$, we reject H_0 and accept H_a.

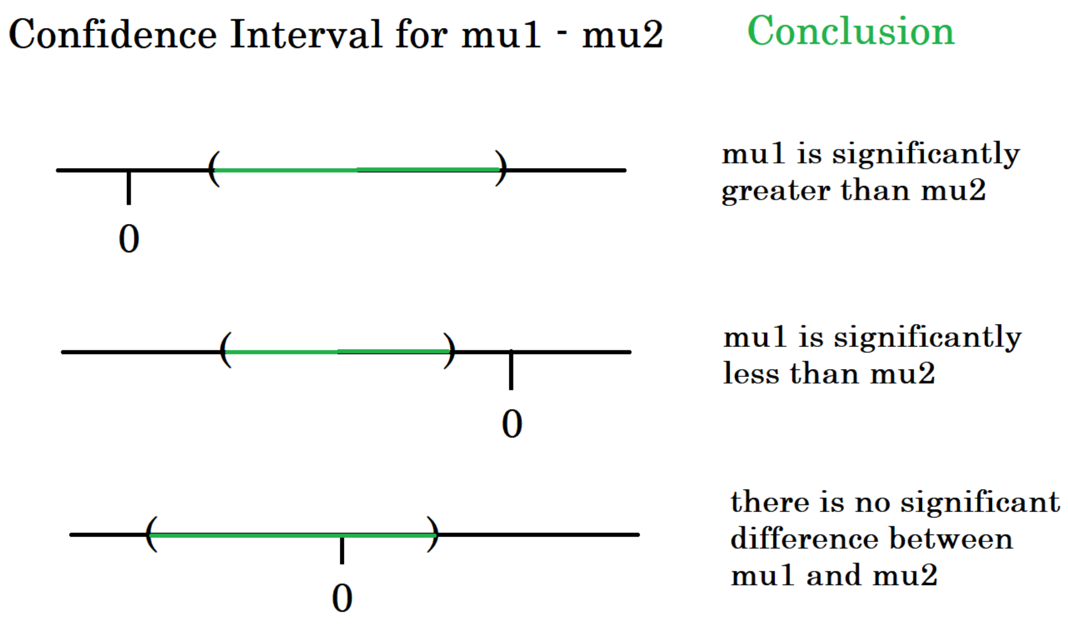
6. **Result:** The data provide sufficient evidence that the lactate level after three games of racquet ball is higher than before at significance level of 0.05.

Let us find an estimate (95% Confidence interval) for the increase in the mean lactate level through three games of racquetball in males.

$$\bar{x}_d \pm t^{n-1}_{(\alpha/2)} \frac{s_d}{\sqrt{n}} = -13.63 \pm 2.365 \frac{8.28}{\sqrt{8}} = -13.63 \pm 6.938$$

or $(-20.568; -6.696)$. $t^{n-1}_{(\alpha/2)} = t^{7}_{0.025} = 2.365$.

Complete Interpretation:
Based on the sample data we can be 95 % confident that the mean increase in lactate level is between 6.692 and 20.568 after three racquetball games. Since zero does not fall within the confidence interval we are 95% confident that the mean lactate levels before and after the games are not the same. Finding that the confidence interval falls entirely below 0, we can conclude the mean before is significantly smaller than the lactate level after.



Confidence Interval for mu1 - mu2

Conclusion

mu1 is significantly
greater than mu2

mu1 is significantly
less than mu2

there is no significant
difference between
mu1 and mu2

3.1.2 Independent Random Samples

In this section we will see how to perform statistical tests and calculate confidence intervals for the difference between two means, $\mu_1 - \mu_2$, based on two independent samples.

The point estimator for $\mu_1 - \mu_2$ that seems an intuitive choice is the difference between the sample means, $\bar{X}_1 - \bar{X}_2$.
In order to do inferential statistics using this difference we have to investigate the distribution of this statistic.

Sample Distribution of $\bar{X}_1 - \bar{X}_2$ from two independent samples.

- For the mean: $\mu_{\bar{X}_1 - \bar{x}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$, so that $\bar{X}_1 - \bar{X}_2$ is an unbiased estimator for $\mu_1 - \mu_2$.
- For the variance:
$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$
- For the standard deviation:
$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
- If n_1 and n_2 are both large or both populations are normal distributed, then is the sampling distribution of $\bar{X}_1 - \bar{X}_2$ (approximately) normal.

The t-statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with

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

For hand calculations this can be approximated by

$$\text{df} = min(n_1 - 1, n_2 - 1)$$

$min(n_1 - 1, n_2 - 1)$ is the smaller of the two numbers $n_1 - 1$ and $n_2 - 1$.
This is all the information we need to put together a

Two–sample t-Test for Comparing Two Population Means

1. **Hypotheses**

Test type	
Upper tail	$H_0 : \mu_1 - \mu_2 \leq d_0$ versus $H_a : \mu_1 - \mu_2 > d_0$
Lower tail	$H_0 : \mu_1 - \mu_2 \geq d_0$ versus $H_a : \mu_1 - \mu_2 < d_0$
Two tail	$H_0 : \mu_1 - \mu_2 = d_0$ versus $H_a : \mu_1 - \mu_2 \neq d_0$

   Choose α

2. **Assumption:** random samples, n_1 and n_2 are large or both populations are approximately normal distributed.

3. **Test statistic:**

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{with} \quad \text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right)}$$

   and for hand calculations (exams) use df $= min(n_1 - 1, n_2 - 1)$

4. **P-value/Rejection Region:**

Test type	**P-value**	**Rejection region**
Upper tail	$P(t > t_0)$	$t_0 > t_\alpha$
Lower tail	$P(t < t_0)$	$t_0 < -t_\alpha$
Two tail	$2 \cdot P(t > abs(t_0))$	$abs(t_0) > t_{\alpha/2}$

5. **Decision**

   Either:
   If the P-value$\leq \alpha$ reject H_0
   If the P-value$> \alpha$ do not reject H_0

   Or:
   If t_0 falls into the Rejection Region reject H_0
   If t_0 falls outside the Rejection Region do not reject H_0

6. **Context**

Example:
A company wants to show that a vitamin supplement decreases the recover time from the common cold. They selected randomly 70 adults with a cold. 35 of those were randomly selected to receive the vitamin supplement. The data on the recover time for both samples is shown below.

population	1	2
	no vitamin	vitamin
sample size	35	35
sample mean	6.9	5.8
sample standard deviation	2.9	1.2

Now test the claim of the company: $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_a : \mu_1 - \mu_2 > 0$ at a significance level of α=0.05.

Assumption: We are using random samples of patients and the sample sizes are large enough.

Test statistic: with $d_0 = 0$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{6.9 - 5.8}{\sqrt{\frac{2.9^2}{35} + \frac{1.2^2}{35}}} = \frac{1.1}{0.53} = 2.07$$

and df = 66.28 (using the long formula) and df = $min(35 - 1, 35 - 1) = 34$ (using the quick formula)

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right)} = \frac{\left(\frac{2.9^2}{35} + \frac{1.2^2}{35}\right)^2}{\left(\frac{(2.9^2/35)^2}{34} + \frac{(1.2^2/35)^2}{34}\right)} = \frac{0.0784}{0.001698 + 0.00004978} = 44.856$$

Rejection Region: Since this is an upper tail test the rejection region is $t_0 > t_\alpha$ for a t-distribution with 44 df and $\alpha = 0.05$. From the t-Table we get $t^4_{0.05}4 = 1.680$.

Decision: Since the value of the test statistic falls into the rejection region, we reject H_0 and accept H_a.

P-value approach: We found $t_0 = 2.07$ for df=44, we find that t_0 falls between $t^{44}_{0.025} = 2.015$ and $t^{44}_{0.01} = 2.414$, so the p-value falls between 0.01 and 0.025 and

$$p - value \leq 0.025 < 0.05 = \alpha$$

so the p-value is less than $\alpha = 0.05$. The test is significant at significance level 0.05, we can reject H_0.

Context: At significance level of 5% the data provide sufficient evidence that the mean recovery time taking the vitamins is shorter than the mean recovery time taking the placebo.

Use a confidence interval to estimate the difference in the mean recovery time for patients taking placebo and those taking vitamins.

Two–sample t-Confidence Interval for Comparing Two Population Means

Assumption: n_1 and n_2 are large or both populations are approximately normally distributed.

The $(1 - \alpha)$ **Confidence Interval for** $\mu_1 - \mu_2$**:**

$$(\bar{x}_1 - \bar{x}_2) \pm t^{df}_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with

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

and for hand calculations (exams) use df $= min(n_1 - 1, n_2 - 1)$

and $t^{df}_{(\alpha/2)}$ is the critical value of the t-distribution with the given number of degrees of freedom (Table D).

Continue Example:
Calculate a 95% Confidence Interval to estimate the difference in recovery time $\mu_1 - \mu_2$.
The degrees of freedom are 66, from the table get $t^{df}_{(\alpha/2)} = t^{44}_{(0.025)} = 2.015$

$$\bar{x}_1 - \bar{x}_2 \pm t^{df}_{(\alpha/2)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$6.9 - 5.8 \pm 2.015 \cdot 0.53$$

$$1.1 \pm 1.068$$

or (0.032 ; 2.168). The 95% confidence interval lies entirely above 0. So that 0 is with a confidence of 95% less than $\mu_1 - \mu_2$. We can state at confidence level of 0.95 that the mean recovery time using placebo is greater than the mean recovery time using vitamins. But the time gained might be small as 0.032 days= 0.768 hours, or about 45 minutes. But it is also possible that one gains on average more than 2 days (now we are talking). Unfortunately we do not know which one.