Two-sample tests

2020

Two-sample tests for means

Two-sample tests match one sample against another. To such tests we come in practice when we would like to check if some binary random variable has a statistically significant effect on another quantitative random variable. For example if the gender has statistically significant effect on the grade of students in statistics.

Case 1. Independent random variables X and Y.

Let us now consider a sample $X_1, X_2, \ldots, X_{n_1}$ of independent observations on $X \in N(\mu_X, \sigma_X^2)$ and another sample $Y_1, Y_2, \ldots, Y_{n_2}$ of independent observations on $Y \in N(\mu_Y, \sigma_Y^2)$. By assumption X

and Y are independent and σ_X and σ_Y are unknown. Let us denote by s_X^2 and s_Y^2 the unbiased estimators correspondingly of σ_1^2 and σ_2^2.

✓ Equal variances

In this point we assume that we have some evidence to assume that $\sigma^2 = \sigma_X^2 = \sigma_Y^2$. The testing hypothesis about equality between the variances will be considered later on.

$H_0 : \mu_X = \mu_Y$ - the difference between the averages in the samples are not statistically significant. The binary random variable has no statistically significant effect on the quantitative random variable.

The alternative could be formulated in one of the following ways:

$$H_A : \mu_X > \mu_Y$$

Note: If $H_A : \mu_X < \mu_Y$ we can change the places of X and Y and to reduce this hypothesis to the previous one. Therefore, we will not consider it separately and we call the right-sided alternative just **one-sided alternative**.

or

$H_A : \mu_X \neq \mu_Y$. The means of the observed random variables are different. The last is the same as differences between the averages of the samples are statistically significant. Or, the binary random variable has statistically significant effect on the quantitative random variable.

We chose **type I error (error of the first kind, or significance level / ниво на значимост/)** in such a way that it is the **maximal probability which we allow to reject H_0 when H_0 is correct** , i.e.

$$\mathbb{P}(To\ reject\ H_0\,|\,H_0) \leq \alpha$$

As you already know the form of H_A determines the formula for the critical area W_α

- In case of $H_A : \mu_X > \mu_Y$
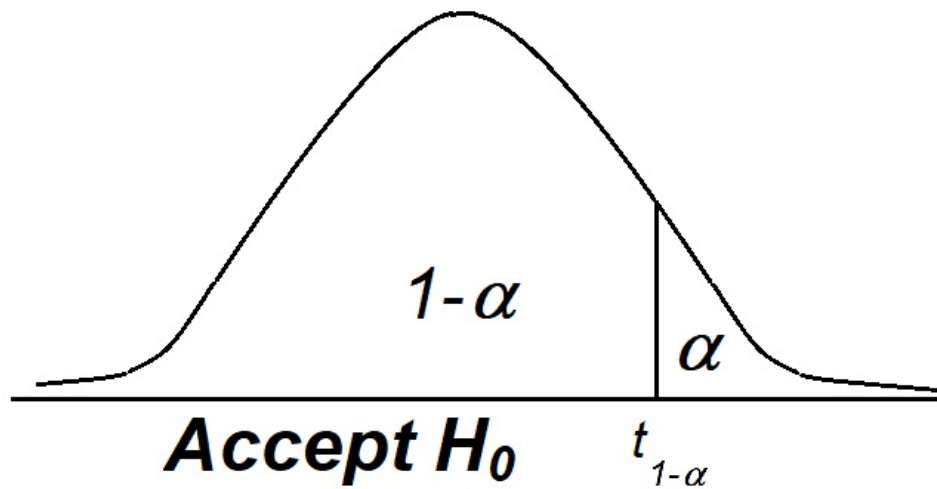
the one-sided alternative the critical area is

$$W_\alpha = \left\{ \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{S_{\overline{X}_{n_1} - \overline{Y}_{n_2}}} \geq t_{1-\alpha, t(n_1+n_2-2)} \right\}$$

where

$$S_{\overline{X}_{n_1} - \overline{Y}_{n_2}} = \sqrt{\frac{\hat{\sigma}_X^2}{n_1} + \frac{\hat{\sigma}_Y^2}{n_2}} = \sqrt{\frac{s_X^2(n_1 - 1) + s_Y^2(n_2 - 1)}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

The reason is that as far as the observed random variables are independent

$$\mathbb{D}(\overline{X}_{n_1} - \overline{Y}_{n_2}) = \mathbb{D}(\overline{X}_{n_1}) + \mathbb{D}(\overline{Y}_{n_2}) = \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}$$

1-α

α

Accept H₀ $t_{1-\alpha}$

When we compute $t_{emp} = \dfrac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{S_{\overline{X}_{n_1} - \overline{Y}_{n_2}}}$ from the data we can

compare t_{emp} with $t_{1-\alpha, t(n_1+n_2-2)}$. If $t_{emp} < t_{1-\alpha, t(n_1+n_2-2)}$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $t_{emp} \geq t_{1-\alpha, t(n_1+n_2-2)}$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable $\mathbb{E}X$ is statistically significant bigger than the mean of the observed random variable $\mathbb{E}Y$.

$$p-value = \mathbb{P}(\eta \geq t_{emp}), \eta \in t(n_1 + n_2 - 2)$$
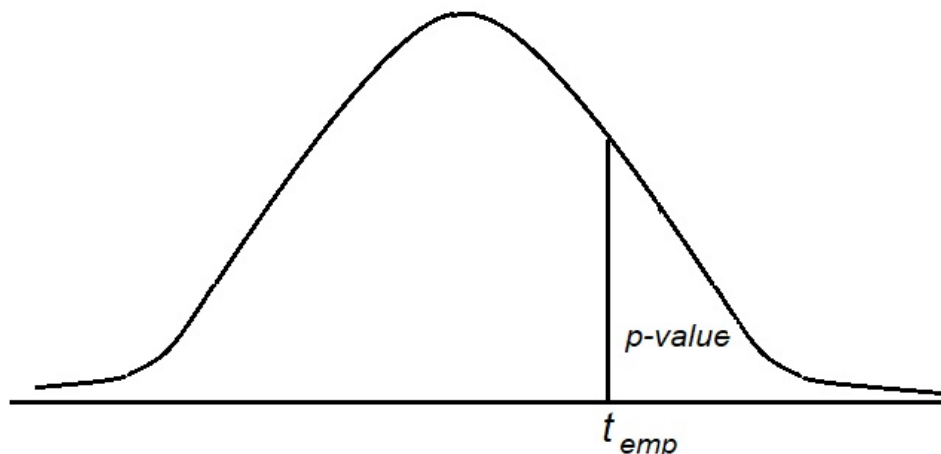
p-value

t_{emp}

If $p-value > \alpha$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $p-value < \alpha$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significant bigger than the mean of the observed random variable Y.

Example 1:

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows
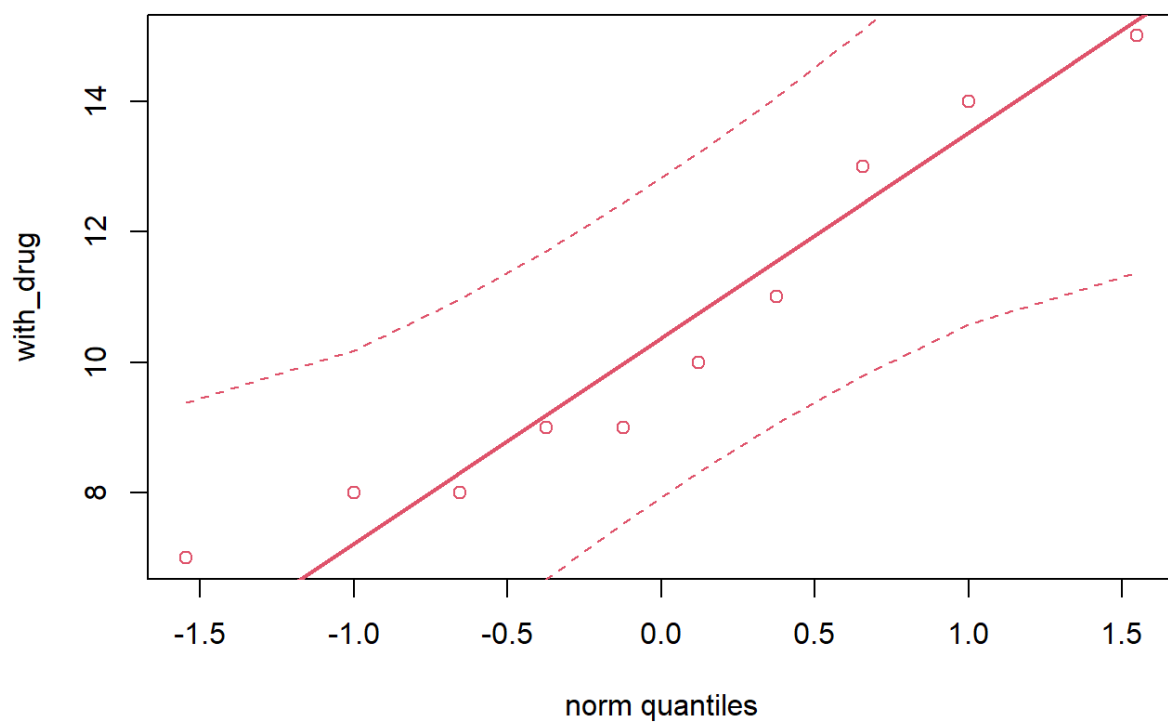
```
> with_drug <- c(15, 10, 13, 7, 9,  8, 11, 9,  14,  8)
> placebo <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
```

Test the hypothesis if the effect of the drug decreases statistically significantly the recovery time.
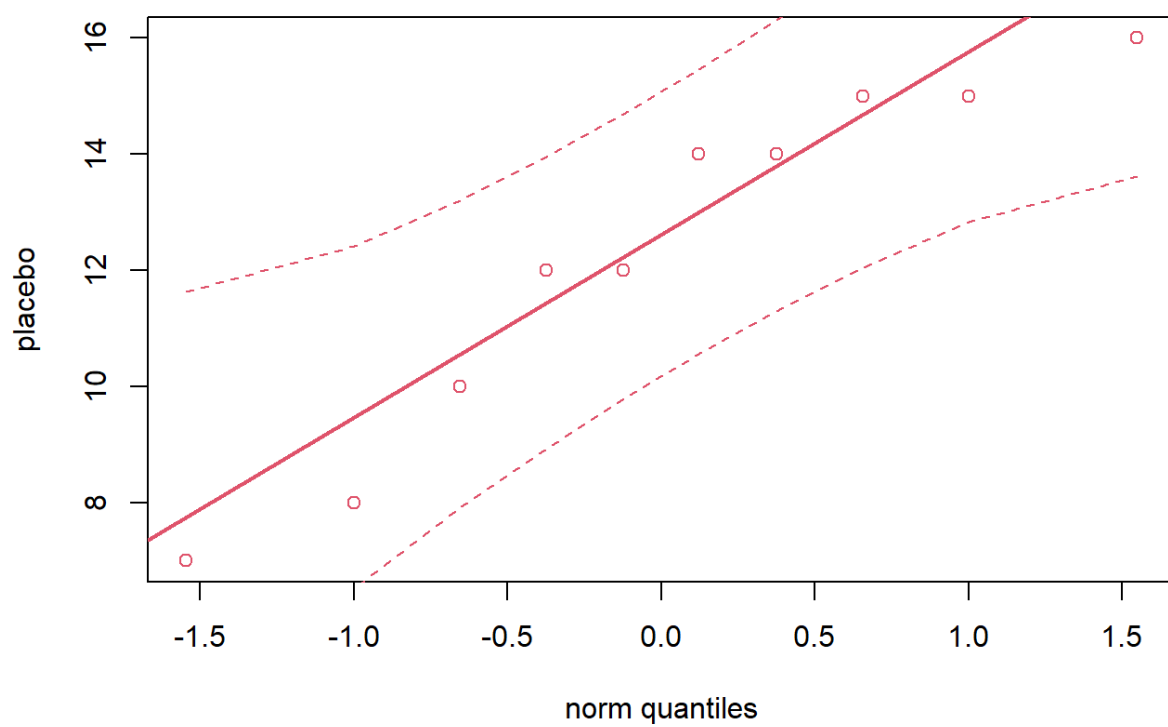
First we need to check in the observed random variables are normal. Therefore, we make normal qq-plots.

```
> library(StatDA)
Warning: package 'StatDA' was built under R version 4.0.3
Loading required package: sgeostat
Warning: package 'sgeostat' was built under R version 4.0.3
Registered S3 method overwritten by 'geoR':
  method           from
  plot.variogram sgeostat
```

```
> qqplot.das(with_drug)
```
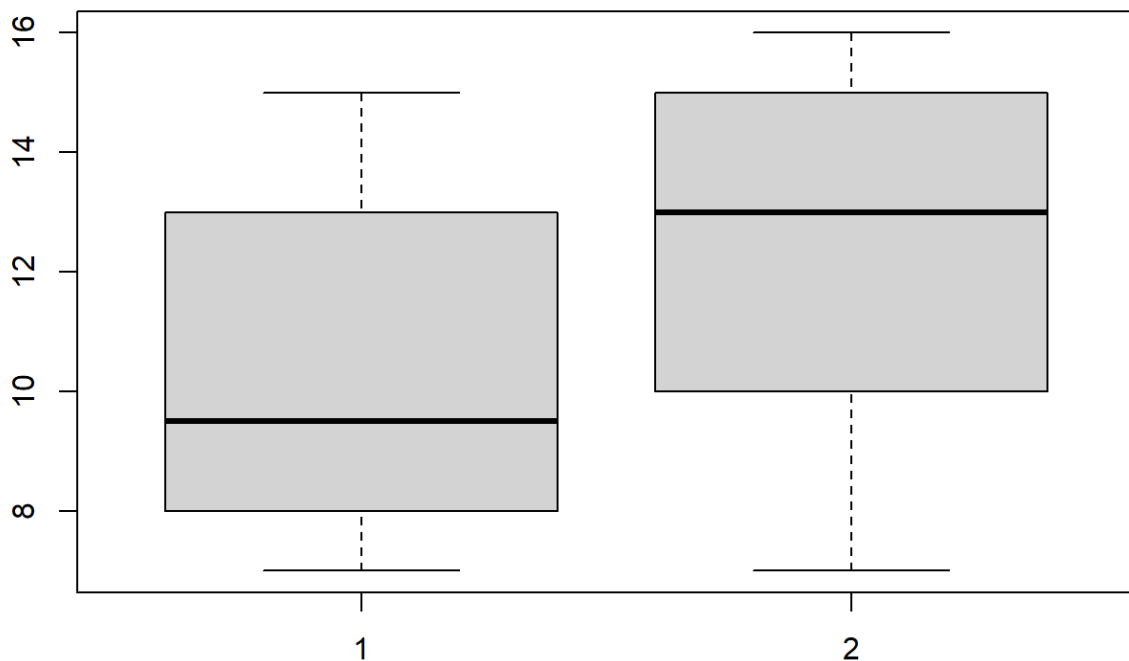


```
> qqplot.das(placebo)
```

We see that both random variables could be accepted as normal.

In order to check if the observed random variables have equal variances we compute

```
> var(with_drug)
[1] 7.6
> var(placebo)
[1] 9.566667
```

In order to observe the difference between the variances we can plot also not normalized boxplots

```
> boxplot(with_drug, placebo)
```


Now we just assume that $\mathbb{D}X = \mathbb{D}Y$. Later on we will make this via hypothesis testing.

Now we check

$H_0 : \mu_X = \mu_Y$ where X is the recovery time of the patients taking placebo and Y is the recovery time of the patients taking the new drug. The drug does not decrease the recovery time. The difference between

the averages of the recovery times in the samples is not statistically significant.

$H_A : \mu_X = \mu_Y$. Taking the drug decreases the recovery time. The recovery times when the patients take the drug are statistically significantly shorter that when they have not taken the drug.

Choose $\alpha = 0.05$ and use the function t.test with parameters alternative = "greater" and var.equal = TRUE.

```
> t.test(with_drug, placebo, alternative = "greater",
var.equal = TRUE)

    Two Sample t-test

data:  with_drug and placebo
t = -1.4501, df = 18, p-value = 0.9179
alternative hypothesis: true difference in means is
greater than 0
95 percent confidence interval:
 -4.171998        Inf
sample estimates:
mean of x mean of y
     10.4      12.3
```

The $p-value = 0.9179 > 0.05 = \alpha$, we have no evidence to reject H_0. According to the data, taking the drug does not statistically significantly decrease the recovery time.
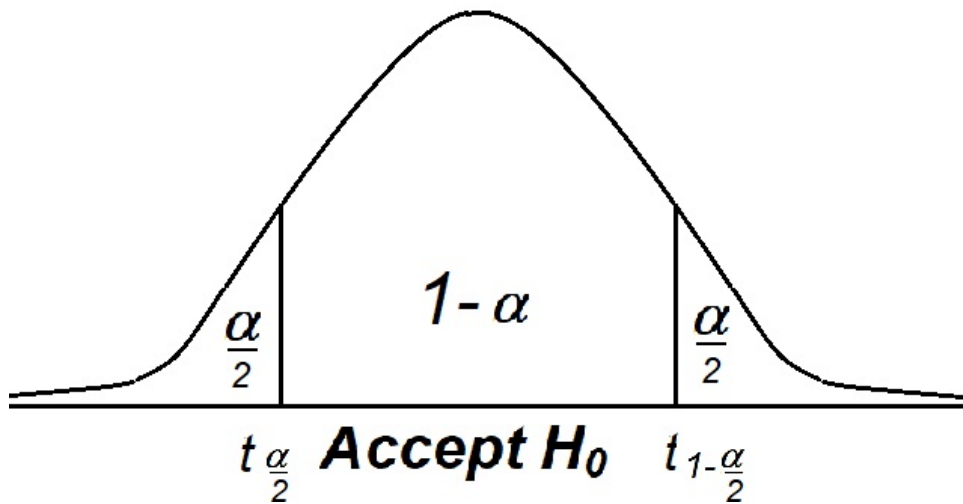
- In case of

$H_A : \mu_X \neq \mu_Y$

we have two-sided alternative and the critical area is

$$W_\alpha = \left\{ \frac{|\bar{X}_{n_1} - \bar{Y}_{n_2}|}{S_{\bar{X}_{n_1} - \bar{Y}_{n_2}}} \geq t_{1-\frac{\alpha}{2},t(n_1+n_2-2)} \right\}$$

When we compute $t_{emp} = \dfrac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{S_{\overline{X}_{n_1} - \overline{Y}_{n_2}}}$ from the data we can

compare t_{emp} with $t_{\frac{\alpha}{2},t(n_1+n_2-2)} = -t_{1-\frac{\alpha}{2},t(n_1+n_2-2)}$ and $t_{1-\frac{\alpha}{2},t(n_1+n_2-2)}$.

If $-t_{1-\frac{\alpha}{2},t(n_1+n_2-2)} < t_{emp} < t_{1-\frac{\alpha}{2},t(n_1+n_2-2)}$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $t_{emp} \leq -t_{1-\frac{\alpha}{2},t(n_1+n_2-2)}$ or $t_{1-\frac{\alpha}{2},t(n_1+n_2-2)} \leq t_{emp}$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significantly different from the mean of the observed random variable Y.

$$p-value = \mathbb{P}(|\eta| \geq t_{emp}) = \mathbb{P}(\eta \leq -t_{emp} \cup \eta \geq t_{emp}), \eta \in t(n_1 + n_2 - 2)$$
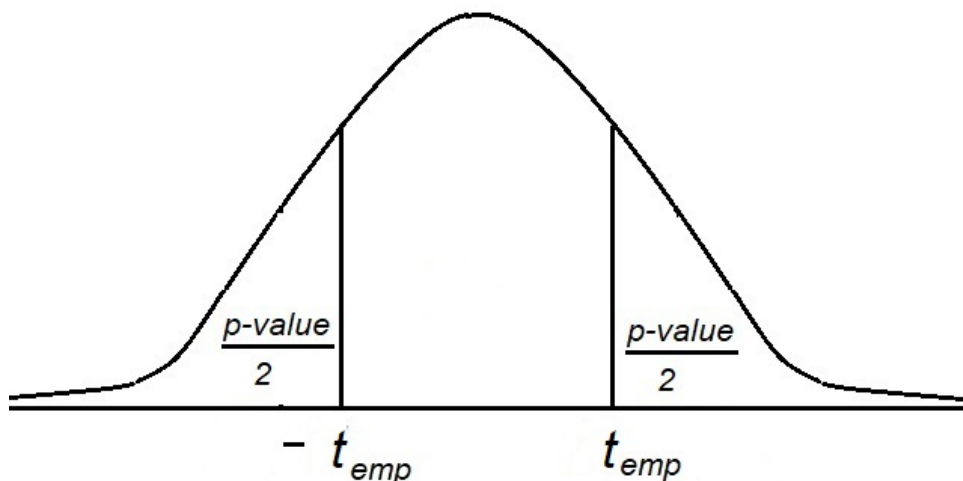
If $p-value > \alpha$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $p-value < \alpha$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significantly different from the mean of the observed random variable Y.

Example 2:

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows

```
> with_drug <- c(15, 10, 13, 7, 9,  8, 11, 9,  14,  8)
> placebo <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
```

Test the hypothesis if there is some statistically significant effect of the drug on the recovery time.

We already assumed that the observed random variables are normal and have equal variances. Now we check

$H_0 : \mu_X = \mu_Y$ where X is the recovery time of the patients taking placebo and Y is the recovery time of the patients taking the new drug. The drug does not decrease the recovery time. The difference between the averages of the recovery times in the samples is not statistically significant.

$H_A : \mu_X \neq \mu_Y$. Taking the drug changes the recovery time. The difference between the averages of the recovery times in the samples is statistically significant.

Choose $\alpha = 0.05$ and use the function t.test with parameters alternative = "two.sided" and var.equal = TRUE.

```
> t.test(with_drug, placebo, alternative = "two.sided",
var.equal = TRUE)


	Two Sample t-test


data:  with_drug and placebo
t = -1.4501, df = 18, p-value = 0.1642
```

```
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -4.6526622  0.8526622
sample estimates:
mean of x mean of y
      10.4       12.3
```

The $p-value = 0.1642 > 0.05 = \alpha$, we have no evidence to reject H H_0. According to the data, taking the drug does not statistically significantly change the recovery time.

✓ Unequal variances

Let us now assume that we have some evidence to conclude that $\sigma_X^2 \neq \sigma_Y^2$. As we already mentioned the testing hypothesis about equality between the variances will be considered later on.

The possibilities for hypothesis are the same as in the previous case. Only the variance in the denominator in the critical area and the degrees of freedom of the quantile change. Again as far as the observed random variables are independent we have

$$\mathbb{D}(\overline{X}_{n_1} - \overline{Y}_{n_2}) = \mathbb{D}(\overline{X}_{n_1}) + \mathbb{D}(\overline{Y}_{n_2}) = \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}$$

However, in this case we cannot replace these different variances with their weighted average.

$H_0 : \mu_X = \mu_Y$ - the difference between the averages in the samples are not statistically significant. The binary random variable has no statistically significant effect on the quantitative random variable.

The alternative could be formulated in one of the following ways:

$H_A : \mu_X > \mu_Y$

or

$H_A : \mu_X \neq \mu_Y$

We chose **type I error (error of the first kind, or significance level / ниво на значимост/)** in such a way that it is the **maximal probability which we allow to reject H_0 when H_0 is correct** , i.e.
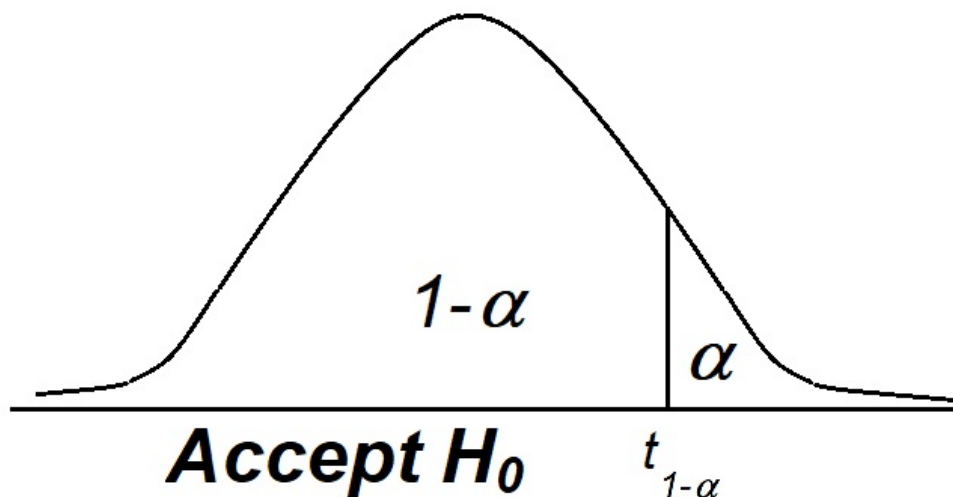
$$\mathbb{P}(To\ reject\ H_0\,|\,H_0) \le \alpha$$

- In case of

$$H_A : \mu_X > \mu_Y$$

$$W_\alpha = \left\{ \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\dfrac{s_X^2}{n_1} + \dfrac{s_Y^2}{n_2}}} \ge t_{1-\alpha,t(m)} \right\}, \quad m = \left[\frac{\left(\dfrac{S_{n_1}^2}{n_1} + \dfrac{S_{n_2}^2}{n_2} \right)^2}{\dfrac{\left(\dfrac{s_{n_1}^2}{n_1} \right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_{n_2}^2}{n_2} \right)^2}{n_2 - 1}} \right]$$

with $[x]$ meaning the integer part of x.

1-α

α

Accept H₀ $t_{1-\alpha}$

When we compute $t_{emp} = \dfrac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\dfrac{s_X^2}{n_1} + \dfrac{s_Y^2}{n_2}}}$ from the data we can

compare with $t_{1-\alpha,t(m)}$. If $t_{emp} < t_{1-\alpha,t(m)}$, then the sample does not belong to the critical area of H_0 and we have no reason to reject H_0. In this case we assume H_0. Conversely if $t_{emp} \ge z_{1-\alpha,t(m)}$, then the sample

is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable $\mathbb{E}X$ is statistically significantly bigger than the mean of the observed random variable $\mathbb{E}Y$.
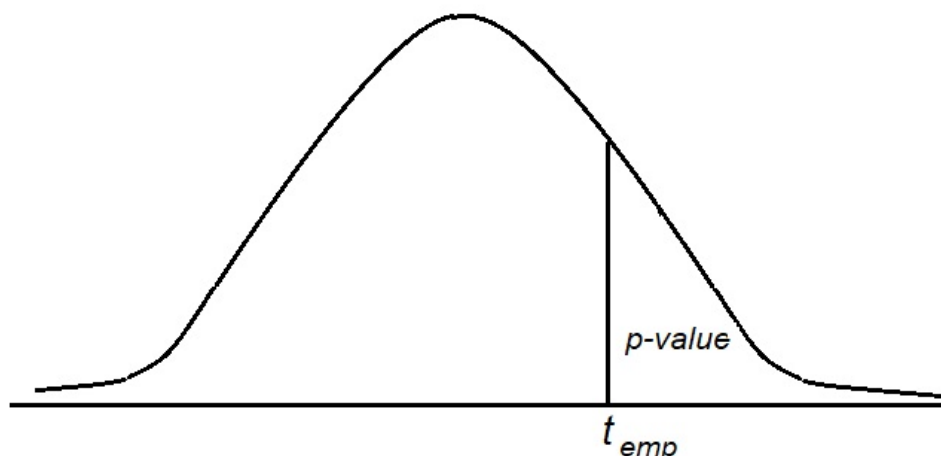
$$p-value = \mathbb{P}(\eta \geq t_{emp})$$

If $p-value > \alpha$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $p-value < \alpha$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significant bigger than the mean of the observed random variable Y.

Example 3:

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows

```
> with_drug <- c(15, 10, 13, 7, 9,  8, 21, 9,  14,  8)
> placebo <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
```
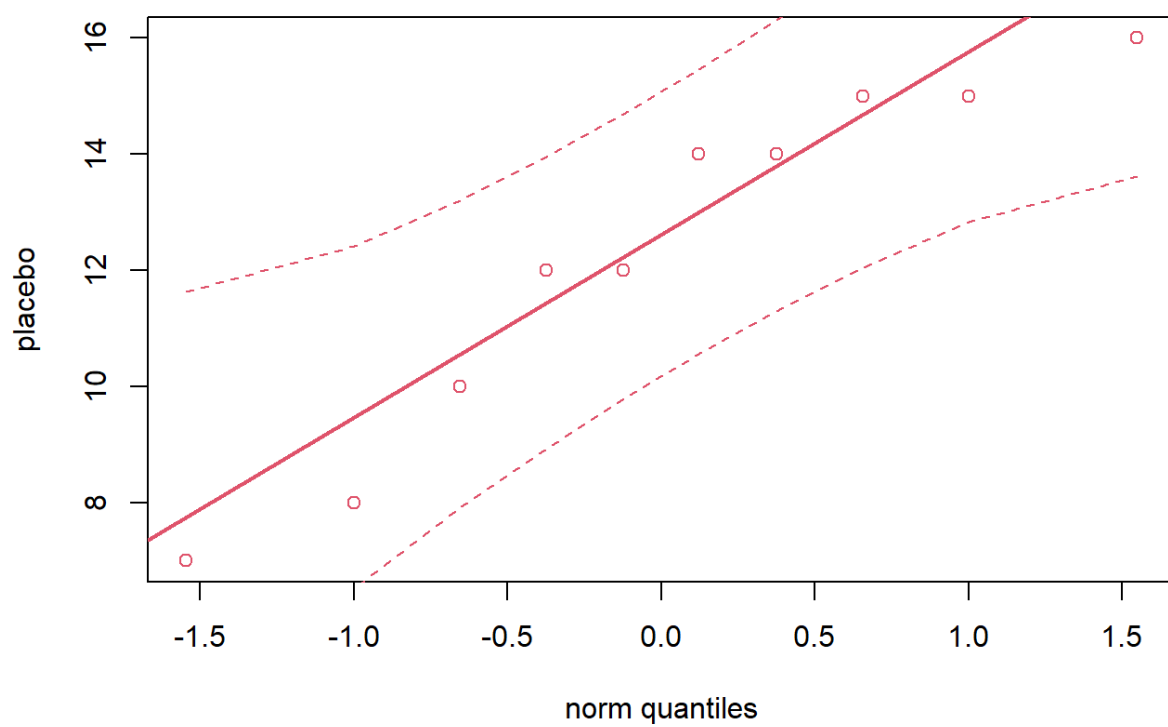
Test the hypothesis if the effect of the drug decreases statistically significantly the recovery time.

First we need to check in the observed random variables are normal. Therefore, we make normal qq-plots.

```
> qqplot.das(with_drug)
```
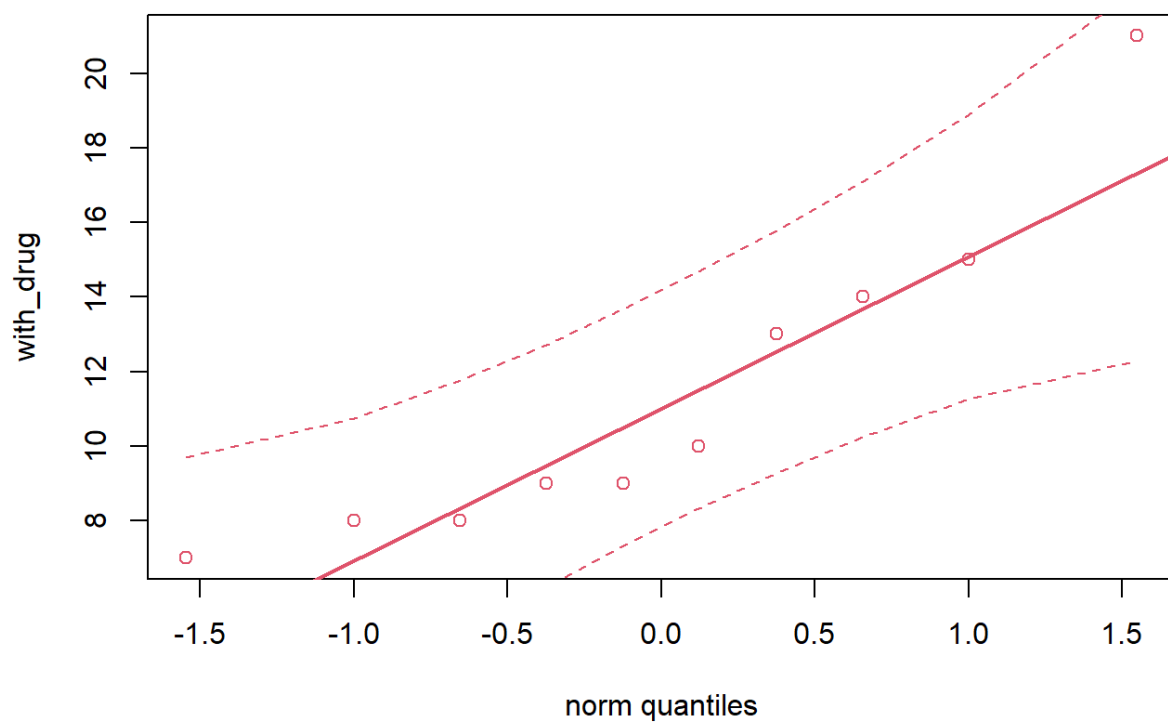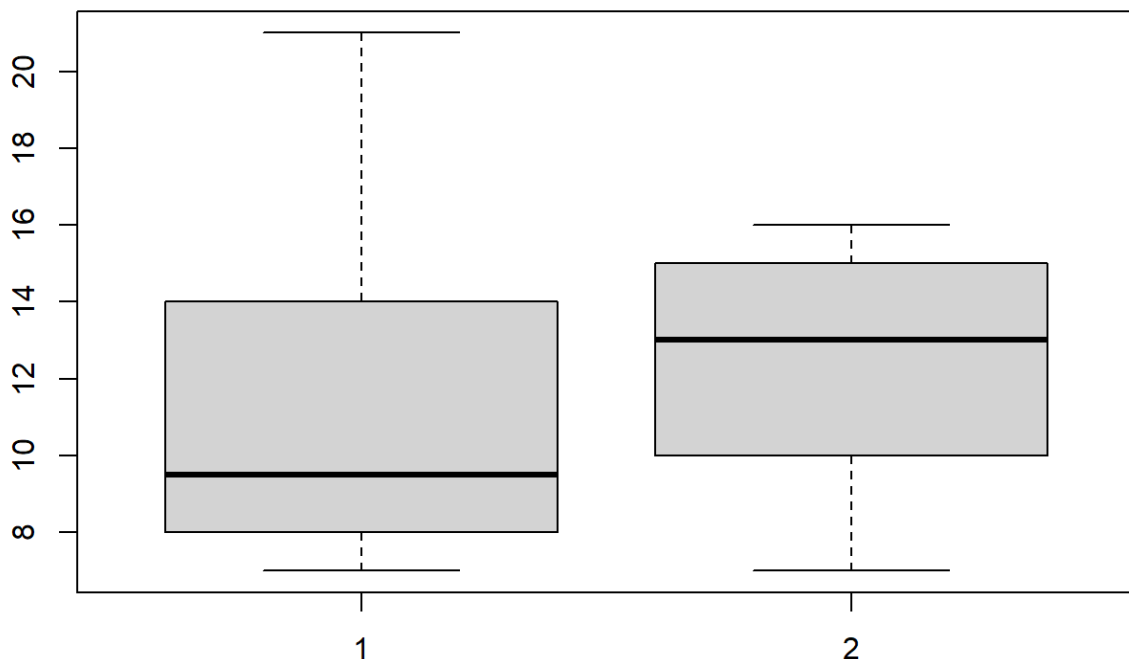


```
> qqplot.das(placebo)
```

We see that both random variables could be accepted as normal.

In order to check if the observed random variables have equal variances we compute

```
> var(with_drug)
[1] 18.93333
> var(placebo)
[1] 9.566667
```

In order to observe the difference between the variances we can plot also not normalized boxplots

```
> boxplot(with_drug, placebo)
```


Now we just assume that $\mathbb{D}X \neq \mathbb{D}Y$. Later on we will make this via hypothesis testing.

Now we check

$H_0 : \mu_X = \mu_Y$ where X is the recovery time of the patients taking placebo and Y is the recovery time of the patients taking the new drug. The drug does not decrease the recovery time. The difference between

the averages of the recovery times in the samples is not statistically significant.

$H_A : \mu_X > \mu_Y$. Taking the drug decreases the recovery time. The recovery times when the patients take the drug are statistically significantly shorter that when they have not taken the drug.

Choose $\alpha = 0.05$ and use the function t.test with parameters alternative = "greater" and var.equal = FALSE.

```
> t.test(with_drug, placebo, alternative = "greater",
var.equal = FALSE)

    Welch Two Sample t-test

data:  with_drug and placebo
t = -0.53311, df = 16.245, p-value = 0.6994
alternative hypothesis: true difference in means is
greater than 0
95 percent confidence interval:
 -3.844664        Inf
sample estimates:
mean of x mean of y
     11.4      12.3
```

The $p-value = 0.6994 > 0.05 = \alpha$, we have no evidence to reject H_0. According to the data, taking the drug does not statistically significantly decrease the recovery time.

- In case of

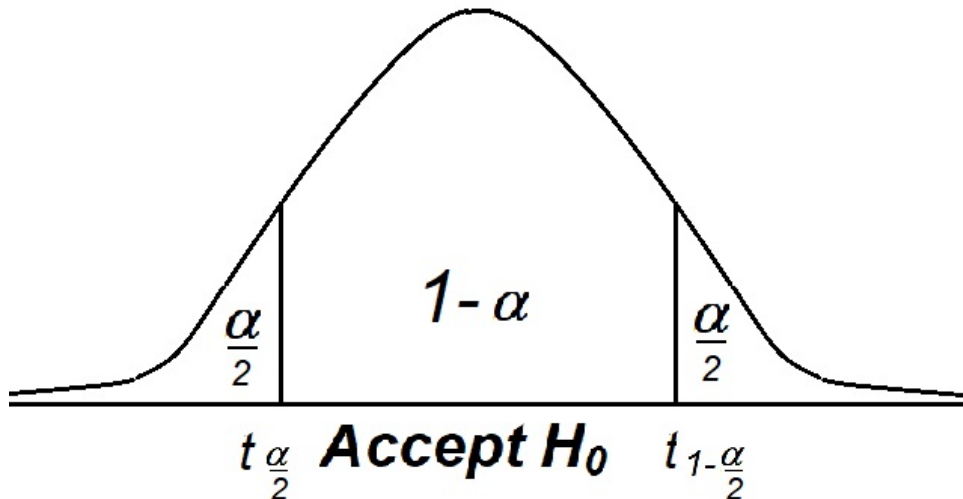$H_A : \mu_X \neq \mu_Y$

$$
W_\alpha = \left\{ \frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \geq t_{1-\frac{\alpha}{2}, t(m)} \right\}, \quad m = \left[\frac{\left(\frac{s_{n_1}^2}{n_1} + \frac{s_{n_2}^2}{n_2} \right)^2}{\frac{\left(\frac{s_{n_1}^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_{n_2}^2}{n_2} \right)^2}{n_2 - 1}} \right]
$$

When we compute $t_{emp} = \dfrac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\dfrac{s_X^2}{n_1} + \dfrac{s_Y^2}{n_2}}}$ from the data we can

compare t_{emp} with $t_{\frac{\alpha}{2},t(m)} = -t_{1-\frac{\alpha}{2},t(m)}$ and $t_{1-\frac{\alpha}{2},t(m)}$

$$t_{\frac{\alpha}{2}} \quad \textbf{Accept } H_0 \quad t_{1-\frac{\alpha}{2}}$$

If $-t_{1-\frac{\alpha}{2},t(m)} < t_{emp} < t_{1-\frac{\alpha}{2},t(m)}$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $t_{emp} \leq -t_{1-\frac{\alpha}{2},t(m)}$ or $t_{1-\frac{\alpha}{2},t(m)} \leq t_{emp}$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significantly different from the mean of the observed random variable Y.

$$p-value = \mathbb{P}(|\eta| \geq t_{emp}) = \mathbb{P}(\eta \leq -t_{emp} \cup \eta \geq t_{emp}), \eta \in t(m)$$
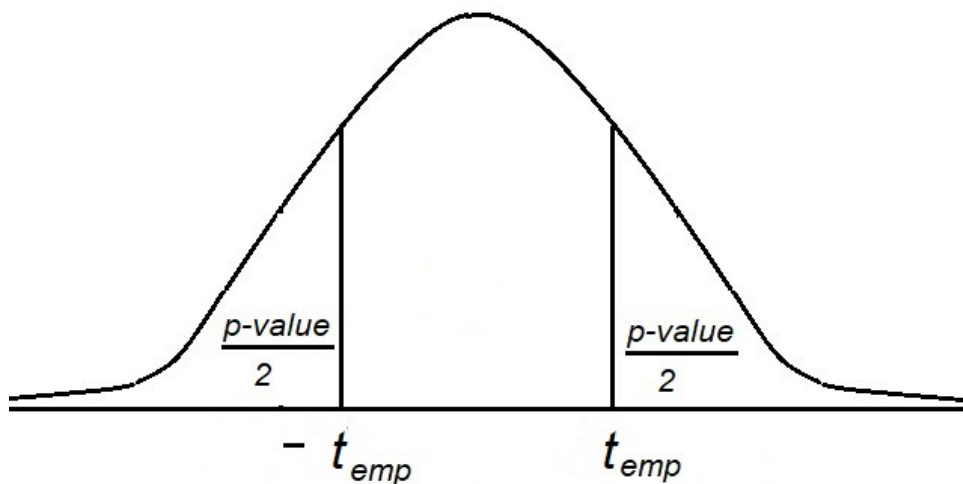
$$-t_{emp} \qquad\qquad t_{emp}$$

If $p-value > \alpha$, then the sample does not belong to the critical area of H_0 and there is no reason to reject H_0. In this case we assume H_0. Conversely if $p-value < \alpha$, then the sample is in the critical area of H_0. In this case we reject H_0. The mean of the observed random variable X is statistically significantly different from the mean of the observed random variable Y.

Example 4:

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows

```
> with_drug <- c(15, 10, 13, 7, 9,  8, 21, 9,  14,  8)
> placebo <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
```

Test the hypothesis if there is some statistically significant effect of the drug on the recovery time.

We already assumed that the observed random variables are normal and have different variances. Now we check

$H_0 : \mu_X = \mu_Y$ where X is the recovery time of the patients taking placebo and Y is the recovery time of the patients taking the new drug. The drug does not decrease the recovery time. The difference between the averages of the recovery times in the samples is not statistically significant.

$H_A : \mu_X \neq \mu_Y$. Taking the drug changes the recovery time. The difference between the averages of the recovery times in the samples is statistically significant.

Choose $\alpha = 0.05$ and use the function t.test with parameters alternative = "two.sided" and var.equal = FALSE.

```
> t.test(with_drug, placebo, alternative = "two.sided",
var.equal = FALSE)

    Welch Two Sample t-test

data:  with_drug and placebo
t = -0.53311, df = 16.245, p-value = 0.6012
```

```
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -4.474425  2.674425
sample estimates:
mean of x mean of y
     11.4      12.3
```

The $p-value = 0.6012 > 0.05 = \alpha$, we have no evidence to reject H_0. According to the data, taking the drug does not statistically significantly change the recovery time.

Case 2. Dependent random variables X and Y. /Matched or paired samples test/

Matched or paired t-test assumes that the two samples share common traits.

Suppose that we have observes two random variables X and Y on one and the same unit. Therefore, we assume that X and Y are dependent(paired) and the samples have equal sizes $n_1 = n_2 = n$.

The basic model is that

$$Y = X + \varepsilon$$

where ε is the randomness.
We transform the sample to the one of the differences $\varepsilon = Y - X$.
For $\mu_\varepsilon := \mathbb{E}\varepsilon$ we want to test if

$H_0 : \mu_\varepsilon = 0$

The alternative could be

$H_A : \mu_\varepsilon > 0$ or $H_A : \mu_\varepsilon \neq 0$

In this way we reduce this test to the one about the difference between the mean of one random variable and a constant.

Therefore, in order to solve the problem we subtract the X's from the Y's and then perform a regular one-sample t-test.

Example 5

In order to promote fairness in grading papers, each application was graded twice by different graders. The data is

Grader 1: 3 0 5 2 5 3 5 3 4 5

Grader 2: 2 1 4 1 4 3 3 2 3 5

a. Based on the grades, can we see if there is a statistically significant difference between the $grader_1$ and $grader_2$?

b. Check if the $grader_1$ gives lower grades than the $grader_2$.

c. Let us check in the $grader_1$ gives higher grades than the $grader_2$.

The data for X the grades of the $grader_1$ are **not independent** of Y - those of $grader_2$ as far as $grader_1$ and $grader_2$ the same papers. So, the samples are **matched (paired)**.
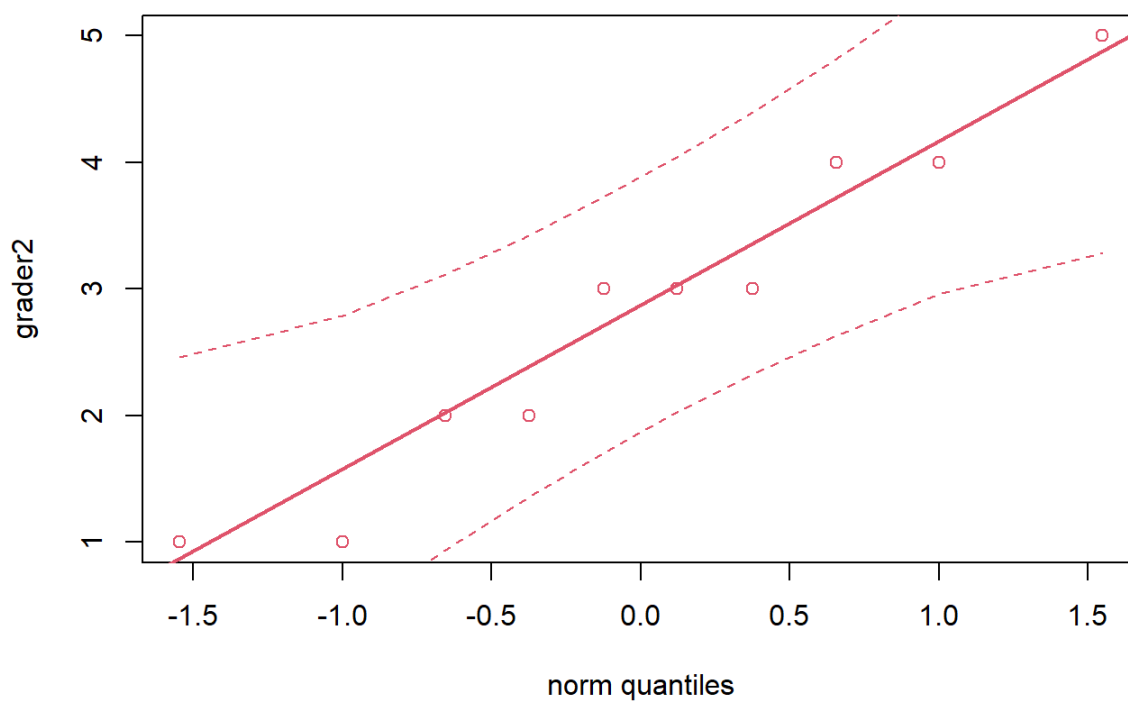
```
> grader1 <- c(3, 0, 5, 2, 5, 3, 5, 3, 4, 5)
> grader2 <- c(2, 1, 4, 1, 4, 3, 3, 2, 3, 5)
```

First we need to check the assumption of normality

```
> library(StatDA)
> qqplot.das(grader1)
```
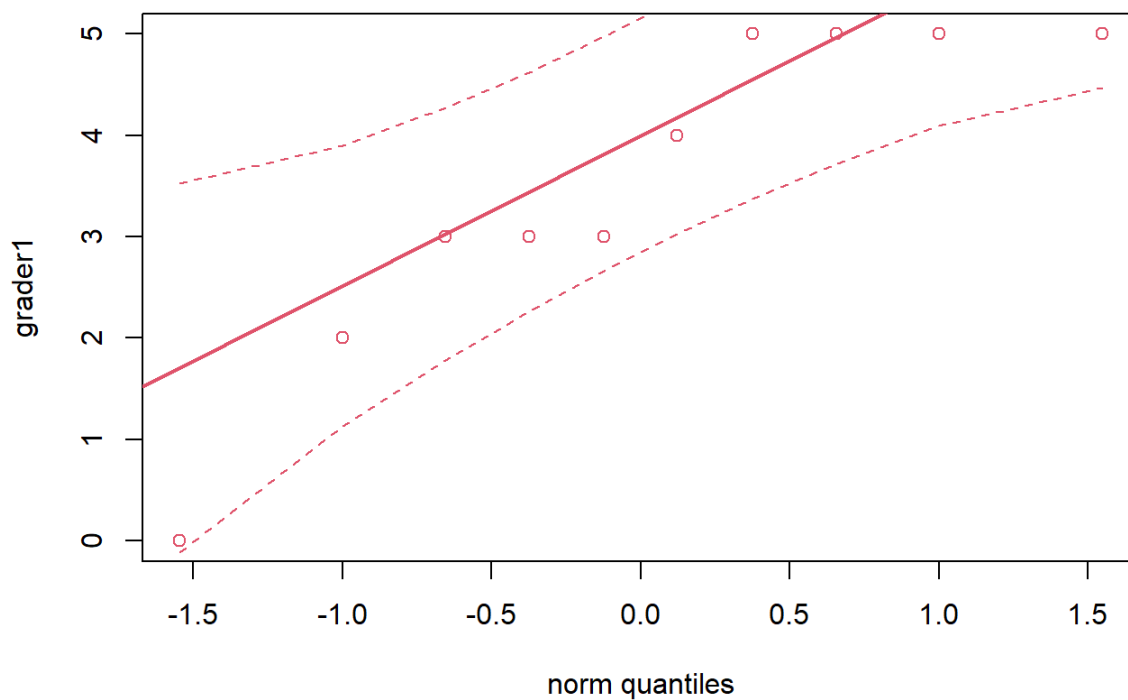


```
> qqplot.das(grader2)
```

```
> shapiro.test(grader1)

    Shapiro-Wilk normality test

data:  grader1
W = 0.85239, p-value = 0.06202
> shapiro.test(grader2)

    Shapiro-Wilk normality test

data:  grader2
W = 0.94191, p-value = 0.5745
```

As far as $p-values > \alpha$ we can assume that the data is normally distributed.

a. We perform the two-sided t-test for paired samples

$$H_0 : \mu_Y - \mu_X = 0$$
$$H_A : \mu_Y - \mu_X \neq 0$$

```
> t.test(grader1, grader2, paired = TRUE)

    Paired t-test

data:  grader1 and grader2
t = 2.6888, df = 9, p-value = 0.02485
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 0.1110663 1.2889337
sample estimates:
mean of the differences
                    0.7
```

The $p-value = 0.02485 < 0.05 = \alpha$, so we reject H_0.

b. Let us check if the $grader_1$ gives lower grades than the $grader_2$
$$H_0 : \mu_Y - \mu_X = 0$$
$$H_A : \mu_Y - \mu_X > 0$$

```
> t.test(grader1, grader2, paired = TRUE, alternative =
"greater")

      Paired t-test

data:  grader1 and grader2
t = 2.6888, df = 9, p-value = 0.01242
alternative hypothesis: true difference in means is
greater than 0
95 percent confidence interval:
 0.2227643        Inf
sample estimates:
mean of the differences
                   0.7
```

The $p-value = 0.01242 < 0.05 = \alpha$, so we reject H_0.

c. Now let us check in the $grader_1$ gives higher grades than the $grader_2$

$H_0 : \mu_Y - \mu_X = 0$
$H_A : \mu_Y - \mu_X < 0$

```
> t.test(grader1, grader2, paired = TRUE, alternative =
"less")

      Paired t-test

data:  grader1 and grader2
t = 2.6888, df = 9, p-value = 0.9876
alternative hypothesis: true difference in means is less
than 0
95 percent confidence interval:
    -Inf 1.177236
sample estimates:
mean of the differences
                   0.7
```

The $p-value = 0.9876 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Two-sample test for proportions

Let $p_1 \in [0,1]$ and $p_2 \in [0,1]$ be the proportions of A in two populations which we call correspondingly "first" and "second". Suppose we have a sample of n_1 independent observations on the first population and n_2 independent populations on the second population. We denote by X_1 and X_2 the number of realizations of A correspondingly in the first and the second population. Then $X_1 \in Bi(n_1, p_1)$ and $X_2 \in Bi(n_2, p_2)$. By the properties of the Binomial distribution

$$\mathbb{E}X_i = n_i p_i, \ \mathbb{D}X_i = n_i p_i(1 - p_i), \ i = 1,2.$$

Consider the proportions $p_{n,1} = \dfrac{X_1}{n_1}$ and $p_{n,2} = \dfrac{X_2}{n_2}$ of A correspondingly in the first and the second population. We suppose that the samples are **independent**. Then,

$$\mathbb{E}(p_{n,1} - p_{n,2}) = p_1 - p_2, \ D(p_{n,1} - p_{n,2}) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}.$$

If the samples are large enough we can apply the Central Limit Theorem,

$$\frac{p_{n,1} - p_{n,2} - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1 - p_1)}{n_1} + \dfrac{p_2(1 - p_2)}{n_2}}} \to \eta, \ n \to \infty, \eta \in N(0,1)$$

If $p_1 = p_2 = p$, i.e. if H_0 is true

$$\frac{p_{n,1} - p_{n,2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \to \eta, \; n \to \infty, \; \eta \in N(0,1)$$

If we would like to apply this limit relation we first have to estimate p. As far as given H_0, $p_1 = p_2 = p$ and $X_1 + X_2 \in Bi(n_1 + n_2, p)$, it's estimator is

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1}{n_1 + n_2}\frac{X_1}{n_1} + \frac{n_2}{n_1 + n_2}\frac{X_2}{n_2} = \frac{n_1}{n_1 + n_2}p_{n,1} + \frac{n_2}{n_1 + n_2}p_{n,2}$$

These considerations lead us to the following test.

$H_0 : p_1 = p_2$, i.e. $p_1 - p_2 = 0$ - the differences between the proportions of A in both samples are not statistically significant. If the critical area is left-sided

$$H_A : p_1 < p_2, \text{ i.e. } p_1 - p_2 < 0$$

$$W_\alpha = \left\{ \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \leq t_{\alpha, t(n_1 + n_2 - 2)} \right\},$$

If the critical area is right-sided

$$H_A : p_1 > p_2, \text{ i.e. } p_1 - p_2 > 0$$

then

$$W_\alpha = \left\{ \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \leq t_{1-\alpha, t(n_1 + n_2 - 2)} \right\}$$

If the critical area is two-sided

$H_A : p_1 \neq p_2$, i.e. $p_1 - p_2 \neq 0$

$$W_\alpha = \left\{ \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \leq t_{1-\frac{\alpha}{2}, t(n_1+n_2-2)} \right\}$$

Example 6

A survey is taken two times over the course of two weeks. The pollsters wish to see if there is a difference in the results as there has been a new advertising campaign run.

	Week1	Week2
Favorable:	45	56
Unfavorable:	35	47

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

```
> prop.test(c(45, 56), c(45+35, 56+47))

    2-sample test for equality of proportions with
continuity correction

data:  c(45, 56) out of c(45 + 35, 56 + 47)
X-squared = 0.010813, df = 1, p-value = 0.9172
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1374478  0.1750692
sample estimates:
   prop 1    prop 2
0.5625000 0.5436893
```

The $p-value = 0.9172 > 0.05 = \alpha$, so we have no evidence to reject H_0. We can assume that $p_1 = p_2$.

Example 7

"Should the federal tax on cigarettes be raised to pay for health care reform?"

	Week1	Week2
Yes:	351	41
No:	254	154

p_1 - the proportion of the non-smoker population who reply "yes"

p_2- the proportion of the smoker population who reply "yes"

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Let us compute the **critical area**

```
> alpha <- 0.05
> alpha/2
[1] 0.025
> qnorm(0.025, 0, 1)
[1] -1.959964
```

-1.96

```
> qnorm(1-0.025, 0, 1)
[1] 1.959964
```

1.96

Therefore,

$$W_\alpha = \left\{ \frac{|p_1 - p_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq 1.96 \right\},$$

We reject H_0 if $t_{emp} \leq -1.96$ or if $t_{emp} \geq 1.96$

Let us now check if the sample is in the critical area

```
> n1 <- 351 + 254
> p1_hat <- 351/n1; p1_hat
[1] 0.5801653
> n2 <- 41 + 154
> p2_hat <- 41/n2; p2_hat
[1] 0.2102564
> p_hat <- (41 + 351)/(n1 + n2);p_hat
[1] 0.49
> z <- ((p1_hat - p2_hat) - 0)/sqrt(p_hat*(1 - p_hat)*(1/
n1 + 1/n2));z
[1] 8.985901
```

$z = 8.9859$ is much greater than 1.96, so the sample is in the critical area for H_0 and we reject H_0.

The p-value approach

```
> pnorm(z, 0, 1, lower.tail = FALSE)
[1] 1.283115e-19
```

The $p-value = 0.01283 - 17 < 0.05 = \alpha$, so we reject H_0.

We can also use the `prop.test` function in R

```
> prop.test(c(351, 41), c(351+254, 41+154))

    2-sample test for equality of proportions with
continuity correction

data:  c(351, 41) out of c(351 + 254, 41 + 154)
X-squared = 79.273, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.2971087 0.4427091
sample estimates:
   prop 1    prop 2
```

`0.5801653 0.2102564`

The $p-value = 0.001e-14 < 0.05 = \alpha$, so we reject H_0.

Two-sample test for variances

Suppose that the observed random variables $X \in N(\mu_1, \sigma_1^2)$ and $Y \in N(\mu_2, \sigma_2^2)$ are independent and we have a sample $X_1, X_2, \ldots, X_{n_1}$ of independent observations on X and a sample $Y_1, Y_2, \ldots, Y_{n_2}$ of independent observations on Y.

Let us remind that if $\sigma_1 = \sigma_2$, then

$$\frac{S_X^2}{S_Y^2} \in F(n_1 - 1; n_2 - 1).$$

It allows us to test

$H_0 : \sigma_X^2 = \sigma_Y^2$ - i.e. the differences between the sample variances are not statistically significant.

As far as F distribution is asymmetric and concentrated on the positive real line usually we work with right-sided critical area.

$H_A : \sigma_X^2 > \sigma_Y^2$

For a chosen α

$$W_\alpha = \left\{ \frac{S_X^2}{S_Y^2} \geq x_{1-\alpha, F(n_1-1, n_2-1)} \right\},$$

Analogously if the critical area is two sided.

Example 8:

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows

```
> with_drug <- c(15,10,13,7,9,8,11,9,14,8)
> placebo <- c(15,14,12,8,14,7,16,10,15,12)
```

Test the hypothesis if the variances of the observed random variables are equal.

```
> var(with_drug)
[1] 7.6
> var(placebo)
[1] 9.566667
```

Now we denote by Y - "with_drug" and X - "placebo" and check

$$H_0 : \sigma_X = \sigma_Y$$
$$H_A : \sigma_X > \sigma_Y$$

Choose $\alpha = 0.05$ and use the function var.test with parameter alternative = "greater".

```
> var.test(with_drug, placebo, alternative = "greater")

    F test to compare two variances

data:  with_drug and placebo
F = 0.79443, num df = 9, denom df = 9, p-value = 0.6313
alternative hypothesis: true ratio of variances is
greater than 1
95 percent confidence interval:
 0.2499062         Inf
sample estimates:
ratio of variances
         0.7944251
```

The $p-value = 0.6316 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Example 9

Recovery time for patients taking a new drug is measured in days. A placebo group is also used. The data are as follows

```
> with_drug<-c(15,2,13,7,9,18,25,9,14,8)
> placebo<-c(15,14,12,8,14,7,16,10,15,12)
```

Test the hypothesis if the variances of the observed random variables are equal.

```
> var(with_drug)
[1] 42
> var(placebo)
[1] 9.566667
```

Now we denote by Y - "with_drug" and X - "placebo" and check

$$H_0 : \sigma_X = \sigma_Y$$
$$H_A : \sigma_X \neq \sigma_Y$$

Choose $\alpha = 0.05$ and use the function var.test with parameter alternative = "two.sided".

```
> var.test(with_drug, placebo, alternative = "two.sided")

    F test to compare two variances

data:  with_drug and placebo
F = 4.3902, num df = 9, denom df = 9, p-value = 0.03817
alternative hypothesis: true ratio of variances is not
equal to 1
95 percent confidence interval:
  1.090474 17.675096
sample estimates:
ratio of variances
          4.390244
```

The $p-value = 0.03817 < 0.05 = \alpha$, so we reject H_0.

Two-sample tests for medians

If we do not know if the means exist then we can compare the medians.

Example 10

Let's compare taxi out times at Newark airport for American and Northwest Airlines.

```
> library(UsingR)
Warning: package 'UsingR' was built under R version 4.0.3
Loading required package: MASS
Loading required package: HistData
Loading required package: Hmisc
Loading required package: lattice
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2


Attaching package: 'Hmisc'
The following objects are masked from 'package:base':

    format.pval, units


Attaching package: 'UsingR'
The following object is masked from 'package:survival':

    cancer
> head(ewr)
  Year Month  AA  CO  DL   HP  NW   TW  UA  US inorout
1 2000   Nov 8.6 8.3 8.6 10.4 8.1  9.1 8.4 7.6      in
2 2000   Oct 8.5 8.0 8.4 11.2 8.2  8.5 8.5 7.8      in
3 2000   Sep 8.1 8.5 8.4 10.2 8.3  8.6 8.2 7.6      in
4 2000   Aug 8.9 9.1 9.2 14.5 9.0 10.3 9.2 8.7      in
5 2000   Jul 8.3 8.9 8.2 11.5 8.8  9.1 9.2 8.2      in
6 2000   Jun 8.8 9.0 8.8 14.9 8.4 10.8 8.9 8.3      in
> ewr_out <- subset(ewr, inorout == "out", select =
c("AA","NW"))
> head(ewr_out)
     AA   NW
24 22.7 19.7
```

```
25 24.8 21.6
26 26.4 22.4
27 33.7 26.1
28 29.3 23.2
29 33.4 24.8
> ewr_AA <- ewr_out$AA; ewr_AA
 [1] 22.7 24.8 26.4 33.7 29.3 33.4 26.4 26.2 24.2 26.3
26.3 23.2 22.9 25.6 26.6
[16] 29.3 35.2 30.2 29.7 27.0 24.0 24.1 24.8
> ewr_NW <- ewr_out$NW; ewr_NW
 [1] 19.7 21.6 22.4 26.1 23.2 24.8 22.8 21.9 21.4 22.0
21.2 19.9 21.3 22.7 23.0
[16] 25.5 31.9 25.0 24.5 24.1 21.8 20.7 21.4
```

In order to see if the distributions of the observed random variables are skewed we plot the boxplots of the samples.

```
> boxplot(ewr_AA, ewr_NW)
```
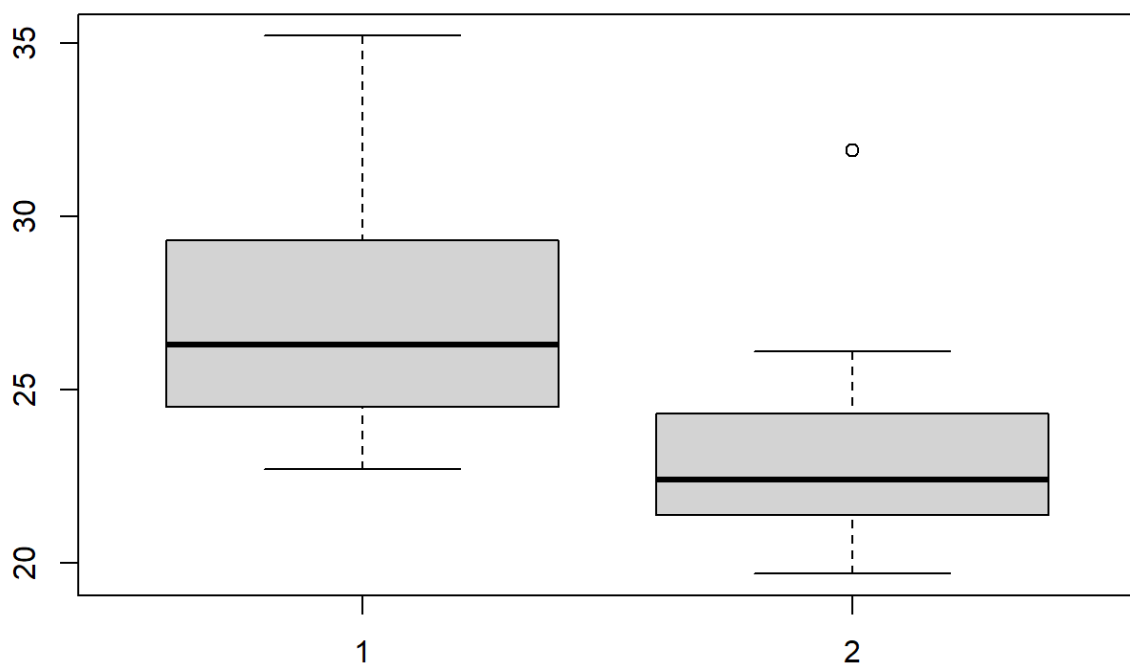

The boxplots show that the distributions are skewed. So a test for the medians is more appropriate.

$$H_0 : Me(ewr_{AA}) = Me(ewr_{NW})$$
$$H_A : Me(ewr_{AA}) \neq Me(ewr_{NW})$$

```
> wilcox.test(ewr_AA, ewr_NW, conf.level = 0.95,
alternative = "two.sided")
Warning in wilcox.test.default(ewr_AA, ewr_NW, conf.level
= 0.95, alternative =
"two.sided"): cannot compute exact p-value with ties

    Wilcoxon rank sum test with continuity correction

data:  ewr_AA and ewr_NW
W = 460.5, p-value = 1.736e-05
alternative hypothesis: true location shift is not equal
to 0
```

The $p-value = 1.73e - 0.5 < 0.05 = \alpha$, so we reject H_0.