Hypothesis Testing

2020

Hypothesis Testing /Проверка на хипотези/

Hypothesis testing is mathematically related to the problem of finding confidence intervals. However, the approach is different.
For **confidence interval**, you use the data to tell you where the unknown parameters should lie. For **hypothesis testing**, you make a hypothesis about the value of the unknown parameter and then calculate how likely it is that you observed the data according to the hypothesis.

Let $X_1, X_2, \ldots, X_n$ be n independent observations on a random variable X. Under "hypothesis" in statistics we understand a concept which trueness is contained in some way in the type or in the parameters of the distribution of the sample. In this way any hypothesis is equivalent to an assumption about the probability distribution of the sample.

General algorithm

1. We formulate
   H_0 - **testing hypothesis**
   H_A - **alternative hypothesis**

The **maximal probability with which we allow to reject H_0 when H_0 is correct is denoted by** α and is called **type I error /грешка от първи род/ (significance level /ниво на значимост/)**

$$T(To\ reject\ H_0 \,|\, H_0) \leq \alpha$$

$1 - \alpha$ is called **confidence level /ниво на доверие/**

2. Obligatory we chose $\alpha \in [0,1]$. Usually $\alpha \in \{0.01;\ 0.05;\ 0.1\}$.

3. For this fixed α we recognize a specific set $W_a \in \mathbb{R}^n$ where if the sample falls we reject H_0. Otherwise we have no reason to reject H_0. W_a is called **critical area for H_0 /критична област за нулевата хипотеза/**.
4. We make a conclusion. More precisely, we check if the sample falls in W_a (the critical area for H_0). If “Yes” we reject H_0. Otherwise we have no reason to reject H_0.

The last means that the following are equivalent:

- To reject H_0;
- $(X_1, X_2, \ldots, X_n) \in W_a$;

This set W_a is not unique. It is natural to chose the set W_a for with the **type II error** β is the **smallest** one.

“To reject H_A” when it is correct is called **type II error** and is usually the risk of such an error is denoted by β,

$$\mathbb{P}(To\ reject\ H_A \,|\, H_A) = \beta$$

It can be computed depending on α.

When α decreases to 0, β increases to 1 and vice versa.

The value

$$1 - \beta = \mathbb{P}(To\ reject\ H_0 \,|\, H_A)$$

is called **power of the criterion /мощност на критерия/**.

The task to find the best formulas for W_a (the critical area for H_0) is object of **mathematical statistics**. We just need to recognize them and to use them.

Another frequently used characteristic of the sample is its $p\!-\!value$. This is the probability to observe current or more extreme test statistics if H_0 is correct.

When the $p\!-\!value$ gets smaller we reject more and more convincingly H_0. More precisely this happens when $p\!-\!value < \alpha$.

Let us repeat the **algorithm** for testing hypothesis.

1. We formulate

H_0 - the tested hypothesis and H_A - alternative hypothesis.

2. We choose $\alpha \in [0,1]$.

3. We recognize the critical area for H_0 which we denoted by W_a.

4A. Compute the **test statistic** (empirical characteristic) from the data and compare it with the corresponding **critical value** (quantile) which participates in W_a.

or

4B. Compute the $p\!-\!value$ of the **test statistic** (empirical characteristic) and compare it with α.

5. Make a conclusion. If the sample is in the critical area for H_0, i.e. if the sample satisfies the inequality of W_a, or with is the same if $p\!-\!value < \alpha$ we reject H_0. Otherwise we have no reason to reject H_0.

Hypothesis testing for the parameters of one population

Let us assume that the probability type of the observed random variable is known, however some of its parameters are unknown and we test their values. Such hypothesis are called **parametric**.

Hypothesis testing of probability for "success"

Example 1:

A coin is tossed 20 times. 13 of them it lends on the tail side. Perform a hypothesis test at a $5\,\%$ significance level to see if the tails are more favourable.

Let us denote by X the number of tails and by p the probability of the event "to have a tail". Then $X \in Bi(20, p)$.

$H_0 : p = \frac{1}{2}$. The coin is symmetric.

$H_A : p > \frac{1}{2}$. The coin is biased in favour of tails.

We only need a **one-tailed test** as the alternative hypothesis says "in favour of tails".

We have $\alpha = 0.05$ and have to determine the critical area for H_0 in such a way that **the maximal** probability

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

We need to compute **the minimal** $b_{1-\alpha}$ such that

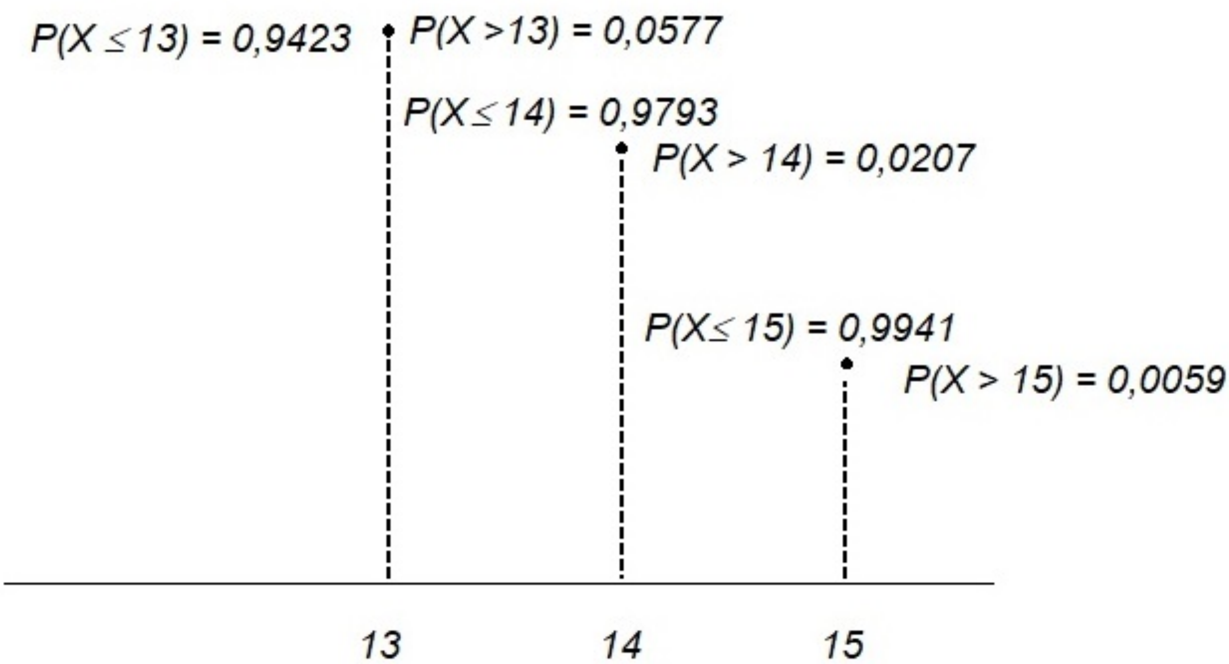
$$\mathbb{P}(X > b_{1-\alpha} \,|\, H_0) \leq 0.05$$

$$1 - \mathbb{P}(X > b_{1-\alpha} \,|\, H_0) \geq 1 - 0.05$$

$$\mathbb{P}(X \leq b_{1-\alpha} \,|\, H_0) \geq 0.95$$

By the definition of the quantiles it is the smallest number such that $\mathbb{P}(X \leq b_{1-\alpha}) \geq 1 - \alpha$. Therefore, $b_{1-\alpha}$ is the $1 - \alpha = 0.95$ quantile of $Bi\left(20, \frac{1}{2}\right)$.

```
> alpha <- 0.05
> qbinom(1 - alpha, 20, 1/2)
[1] 14
> pbinom(14, 20, 1/2)
[1] 0.9793053
> pbinom(13, 20, 1/2)
[1] 0.9423409
> pbinom(15, 20, 1/2)
[1] 0.994091
> pbinom(14, 20, 1/2, lower.tail = FALSE)
[1] 0.02069473
> pbinom(13, 20, 1/2, lower.tail = FALSE)
[1] 0.05765915
> pbinom(15, 20, 1/2, lower.tail = FALSE)
[1] 0.005908966
> pValue <- pbinom(12, 20, 1/2, lower.tail = FALSE)
> pValue
[1] 0.131588
```



By the definition of the **critical area** and as far as 14 is the smallest number such that $\mathbb{P}(X > 14) \leq \alpha$ the **critical area** is

$$W_a = \{X > 14\}$$

If we have 15 or more tails we can reject H_0 and we can say that the tails are more favourable than the heads. The coin will be biased in favour of tails.

By the condition of the task we have 13 tails. So, the sample does not belong to the critical area for H_0 and we cannot reject H_0. We can't say the coin is biased in favour of tails.

We can make the same by

```
> binom.test (13, 20, p = 0.5, alternative = "greater")

	Exact binomial test

data:  13 and 20
number of successes = 13, number of trials = 20, p-value
= 0.1316
alternative hypothesis: true probability of success is
greater than 0.5
95 percent confidence interval:
 0.4419655 1.0000000
sample estimates:
probability of success
                  0.65
```

Now let's solve the task if the alternative is "The coin is biased". In this case we have to use **two-tailed test**.

Again let X be the number of tails and p be the probability of the event "to have a tail" on the observed coin.

Then, $X \in Bi(20, p)$.

$H_0 : p = \dfrac{1}{2}$. The coin is symmetric.

$H_A : p \neq \dfrac{1}{2}$. The coin is biased.

We chose α and have to determine the **critical area** for H_0 in such a way that

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

We need to compute the the largest $b_{\frac{\alpha}{2}}$ and the smallest $b_{1-\frac{\alpha}{2}}$ such that

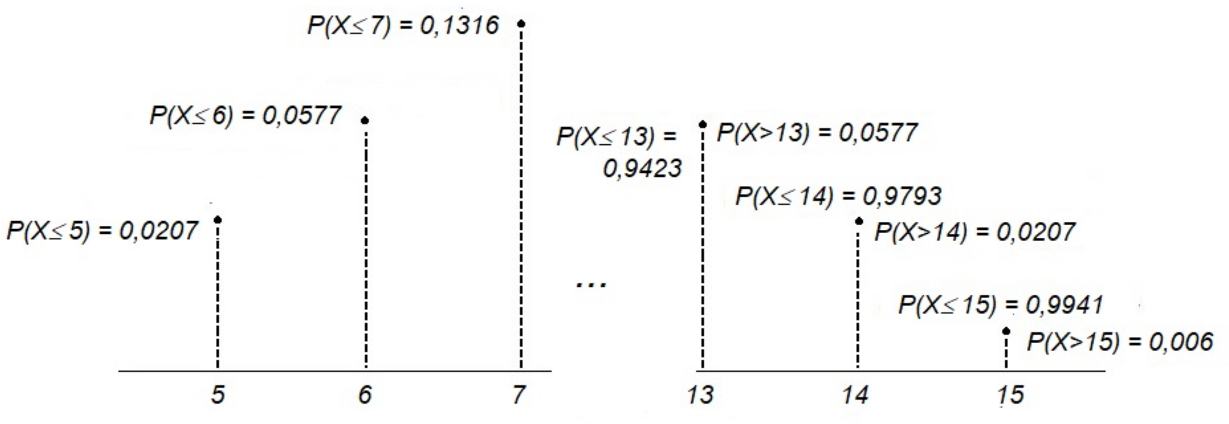
$$\mathbb{P}(To\ reject\ H_0 | H_0) \leq \alpha$$

By the definition of the quantiles $b_{\frac{\alpha}{2}}$ is the smallest number such that $\mathbb{P}(X \leq b_{\frac{\alpha}{2}}) \geq \frac{\alpha}{2}$.

Analogously $b_{1-\frac{\alpha}{2}}$ is the smallest number such that $\mathbb{P}(X \leq b_{1-\frac{\alpha}{2}}) \geq 1 - \frac{\alpha}{2}$.

```
> qbinom(alpha / 2, 20, 1/2)
[1] 6
> qbinom(1 - alpha/2, 20, 1/2)
[1] 14
> pbinom(6, 20, 1/2)
[1] 0.05765915
> pbinom(5, 20, 1/2)
[1] 0.02069473
> pbinom(7, 20, 1/2)
[1] 0.131588
> pbinom(14, 20, 1/2)
[1] 0.9793053
> pbinom(13, 20, 1/2)
[1] 0.9423409
> pbinom(15, 20, 1/2)
[1] 0.994091
> pbinom(14, 20, 1/2, lower.tail = FALSE)
[1] 0.02069473
> pbinom(13, 20, 1/2, lower.tail = FALSE)
[1] 0.05765915
> pbinom(15, 20, 1/2, lower.tail = FALSE)
[1] 0.005908966
> pValue <- pbinom(7, 20, 1/2) + pbinom(12, 20, 1/2,
lower.tail = FALSE)
```

```
> pValue
[1] 0.263176
```



By the definition of the **critical area** and as far as the largest number such that $\mathbb{P}(X < 6) \leq \dfrac{\alpha}{2}$ is 6, and the smallest number such that $\mathbb{P}(X > 14) \leq \dfrac{\alpha}{2}$ is 14, the **critical area** is

$$W_\alpha = \{X < 6 \ or \ X > 14\}$$

Then if the tails are < 6 or > 14 we reject H_0 and conclude that the coin is not fair.

However by the condition of the task we have 13 tails. So, the sample does not belong to the critical area for H_0 and we cannot reject H_0. We cannot conclude that the coin is not fair.

We can make the same by

```
> binom.test (13, 20, p = 0.5)

    Exact binomial test

data:  13 and 20
number of successes = 13, number of trials = 20, p-value
= 0.2632
alternative hypothesis: true probability of success is
not equal to 0.5
95 percent confidence interval:
 0.4078115 0.8460908
```

```
sample estimates:
probability of success
                  0.65
```

or

```
> prop.test(13, 20, p = 0.5)

    1-sample proportions test with continuity correction

data:  13 out of 20, null probability 0.5
X-squared = 1.25, df = 1, p-value = 0.2636
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.4094896 0.8369133
sample estimates:
   p
0.65
```

The $p-value = 0.2636 > 0.05 = \alpha$, so the sample is not in the critical area for H_0 and we can't reject H_0. So we have no evidence to say that the coin is not symmetric.

Example 2:

Suppose you have a die and suspect that it is biased towards the number six. We throw the die 25 times and count that the number six comes up 7 times. Perform a hypothesis test at a $5\ \%$ significance level to see whether the die is biased and six is more favourable.

Let us denote by X the number of sixes and by p the probability of the event "to have six points" on the observed die.

Then on this die $X \in Bi(25, p)$.

$H_0 : p = \frac{1}{6}$. The die is not biased.

$H_A : p > \frac{1}{6}$. The die is biased in favour of six points.

We need a **one-tailed test** as the alternative hypothesis says “in favour of six points”.

We have $\alpha = 0.05$ and have to determine the critical area for H_0 in such a way that

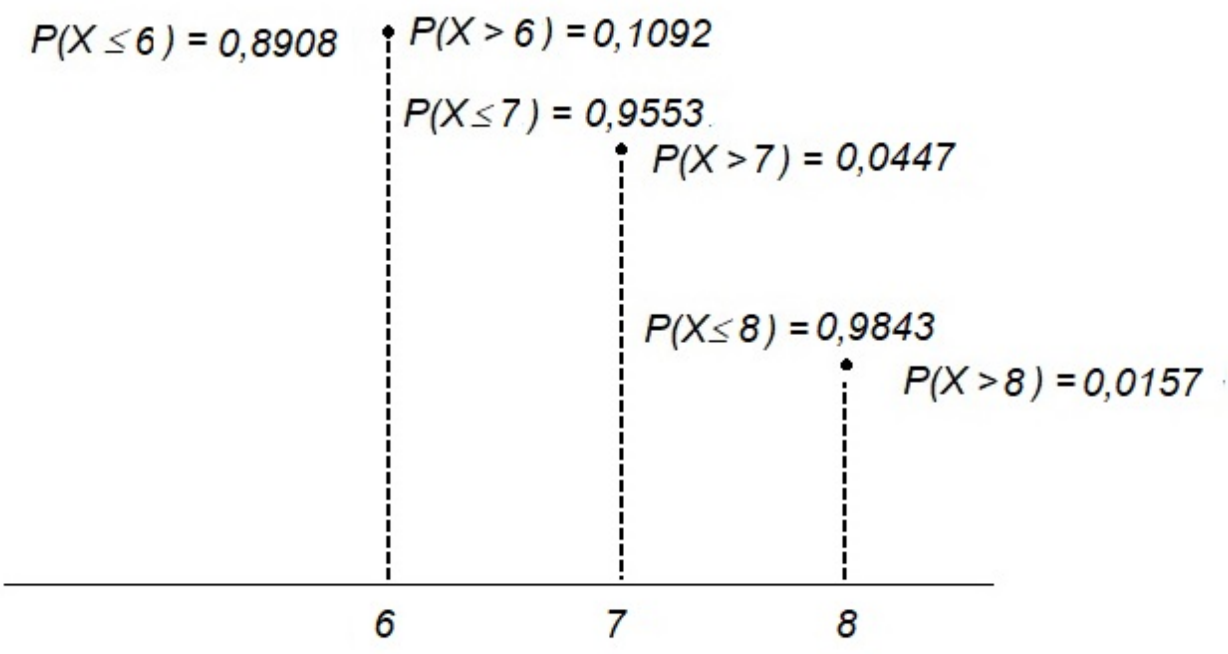
$$\mathbb{P}(To\ reject\ H_0 | H_0) \leq \alpha$$

We need to compute the lagrest $b_{1-\alpha}$ such that

$$\mathbb{P}(X > b_{1-\alpha} | H_0) < 0.05$$

By the definition of the quantiles it is the smallest number such that $\mathbb{P}(X \leq b_{1-\alpha}) \geq 1 - \alpha$.

```
> alpha <- 0.05
> qbinom(1 - alpha, 25, 1/6)
[1] 7
> pbinom(7, 25, 1/6)
[1] 0.9552681
> pbinom(6, 25, 1/6)
[1] 0.8907691
> pbinom(8, 25, 1/6)
[1] 0.9842926
> pbinom(7, 25, 1/6, lower.tail = FALSE)
[1] 0.04473193
> pbinom(6, 25, 1/6, lower.tail = FALSE)
[1] 0.1092309
> pbinom(8, 25, 1/6, lower.tail = FALSE)
[1] 0.01570738
> pValue <- pbinom(6, 25, 1/6, lower.tail = FALSE)
> pValue
[1] 0.1092309
```



By the definition of the critical area and as far as 7 is the smallest number such that $\mathbb{P}(X > 7) \leq \alpha$, the critical area is

$$W_\alpha = \{X > 7\}$$

If we have more than 7 sixes we can reject H_0 and we can say that the six is more favourable than the others. The die is biased in favour of six.

However in our case we have exactly 7 sixes, therefore, the sample is not in the critical area for H_0 and we cannot reject H_0. We cannot conclude that the die is not fair.

We can make the same by

```
> binom.test (7, 25, p = 1/6, alternative = "greater")
```

```
    Exact binomial test

data:  7 and 25
number of successes = 7, number of trials = 25, p-value =
0.1092
alternative hypothesis: true probability of success is
greater than 0.1666667
95 percent confidence interval:
 0.1394753 1.0000000
sample estimates:
```

```
probability of success
                  0.28
```

A two-tailed test would be the result of an alternative hypothesis saying "The die is biased".

Again let X be the number of six and p be the probability of the event "to have six points" on this die.

Then, $X \in Bi(25, p)$ on this die.
$H_0 : p = \frac{1}{6}$. The die is not biased.

$H_A : p \neq \frac{1}{6}$. The die is biased.

We chose $\alpha = 0.05$ and have to determine the critical area for H_0 in such a way that

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

We need to compute the largest $b_{\frac{\alpha}{2}}$ and the smallest $b_{1-\frac{\alpha}{2}}$ such that

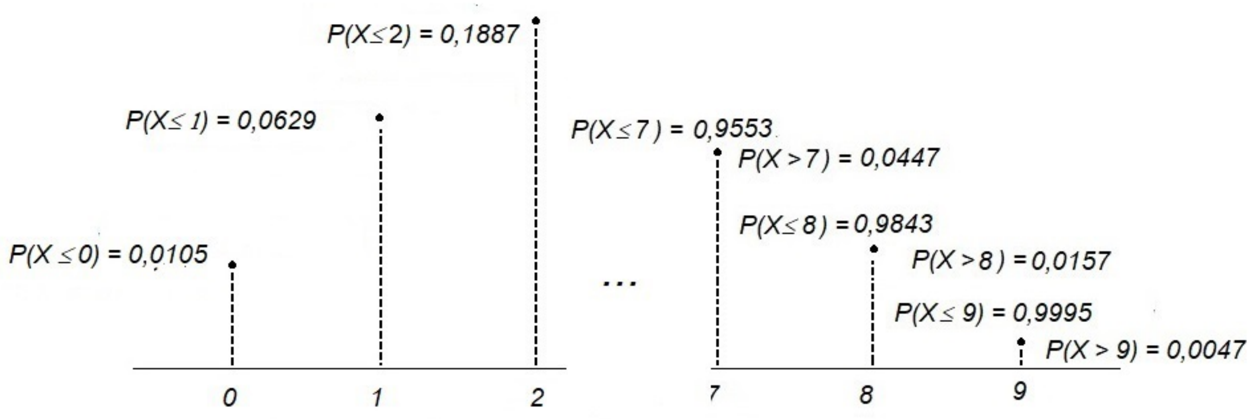
$$\mathbb{P}(X < b_{\frac{\alpha}{2}} \cup X > b_{1-\frac{\alpha}{2}} \,|\, H_0) \leq \alpha$$

By the definition of the quantiles $b_{\frac{\alpha}{2}}$ is the smallest number such that $\mathbb{P}(X \leq b_{\frac{\alpha}{2}}) \geq \frac{\alpha}{2}$.

Analogously $b_{1-\frac{\alpha}{2}}$ is the smallest number such that $\mathbb{P}(X \leq b_{1-\frac{\alpha}{2}}) \geq 1 - \frac{\alpha}{2}$.

```
> qbinom(alpha/2, 25, 1/6)
[1] 1
> qbinom(1 - alpha/2, 25, 1/6)
[1] 8
> pbinom(0, 25, 1/6)
[1] 0.0104826
> pbinom(1, 25, 1/6)
[1] 0.06289558
```

```
> pbinom(2, 25, 1/6)
[1] 0.1886867
> pbinom(8, 25, 1/6)
[1] 0.9842926
> pbinom(7, 25, 1/6)
[1] 0.9552681
> pbinom(9, 25, 1/6)
[1] 0.9952574
> pbinom(8, 25, 1/6, lower.tail = FALSE)
[1] 0.01570738
> pbinom(7, 25, 1/6, lower.tail = FALSE)
[1] 0.04473193
> pbinom(9, 25, 1/6, lower.tail = FALSE)
[1] 0.004742551
> pValue <- pbinom(1, 25, 1/6) + pbinom(6, 25, 1/6,
lower.tail = FALSE)
> pValue
[1] 0.1721265
```



By the definition of the critical area and as far as the largest number such that $\mathbb{P}(X < 1) \leq \dfrac{\alpha}{2}$ is 1, and the smallest number such that $\mathbb{P}(X > 8) \leq \dfrac{\alpha}{2}$ is 8, the critical area is not symmetric

$$W_\alpha = \{X < 1 \; or \; X > 8\}$$

Then if the six are 0 or > 8 we reject H_0. By condition the number six comes up 7 times. So, we cannot reject H_0.

We can make the same by

```
> binom.test (7, 25, p = 1/6)
```

```
    Exact binomial test

data:  7 and 25
number of successes = 7, number of trials = 25, p-value =
0.1721
alternative hypothesis: true probability of success is
not equal to 0.1666667
95 percent confidence interval:
 0.1207167 0.4938768
sample estimates:
probability of success
                  0.28
> prop.test(7, 25, p = 1/6)
Warning in prop.test(7, 25, p = 1/6): Chi-squared
approximation may be incorrect

    1-sample proportions test with continuity correction

data:  7 out of 25, null probability 1/6
X-squared = 1.568, df = 1, p-value = 0.2105
alternative hypothesis: true p is not equal to 0.1666667
95 percent confidence interval:
 0.1287239 0.4959901
sample estimates:
   p
0.28
```

The $p-value = 0.2105 > 0.05 = \alpha$, therefore the sample is not in the critical area for H_0 and we assume H_0, so the die isn't biased.

Note: For different α it is possible to obtain contradictions. A possible solution of the problem is to increase the sample and to perform analogous hypothesis testing again. Toss a die 30 times and test the same hypothesis.

Example 3:

You ask 100 people in a survey and 42 say "yes" to your question. Does this support the hypothesis that the true proportion is $50\,\%$?

In this case $X \in Bi(100, p)$.

To answer this, we set up a **two-sided hypothesis test**. The **null hypothesis** is

$$H_0 : p = 0.5$$

and the **alternative hypothesis** is

$$H_A : p \neq 0.5$$

We can use the `binom.test` function

```
> binom.test(42, 100, p = 0.5)

	Exact binomial test

data:  42 and 100
number of successes = 42, number of trials = 100, p-value
= 0.1332
alternative hypothesis: true probability of success is 
not equal to 0.5
95 percent confidence interval:
 0.3219855 0.5228808
sample estimates:
probability of success 
                  0.42
```

or `prop.test` function

```
> prop.test(42, 100, p = 0.5)

	1-sample proportions test with continuity correction

data:  42 out of 100, null probability 0.5
X-squared = 2.25, df = 1, p-value = 0.1336
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3233236 0.5228954
sample estimates:
   p 
0.42
```

The $p-value$ reports how likely is to see this data or worse assuming the null hypothesis. In particular $p-value$ is the probability of 42 or fewer or 58 or more answer “yes” when the chance a person will answer “yes” is fifty-fifty. The $p-value - 0.1336 > 0.05 = \alpha$ is not so small as to make an observation of 42 seem unreasonable in 100 samples assuming the H_0. So the sample is not in the critical part for the null hypothesis and we **don’t have reason to reject the null hypothesis**.

Let’s repeat the survey, but this time suppose we are asking 1000 people and 420 say yes. Does this support the hypothesis that the true proportion is $50\,\%$?

This time the hypothesis look the same.

$$H_0 : p = 0.5$$
$$H_A : p \neq 0.5$$

```
> binom.test(420, 1000, p = 0.5)

	Exact binomial test

data:  420 and 1000
number of successes = 420, number of trials = 1000, p-
value = 4.697e-07
alternative hypothesis: true probability of success is 
not equal to 0.5
95 percent confidence interval:
 0.3891836 0.4512888
sample estimates:
probability of success 
                  0.42 
> prop.test(420, 1000, p = 0.5)

	1-sample proportions test with continuity correction

data:  420 out of 1000, null probability 0.5
X-squared = 25.281, df = 1, p-value = 4.956e-07
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
```

```
 0.3892796 0.4513427
sample estimates:
   p
0.42
```

Now the $p-value = 4.956e-07 < 0.05 = \alpha$ and the **null hypothesis is not supported**. The sample is in the critical area for the null hypothesis and we reject H_0.

These to examples show that when the sample size is smaller the accuracy is smaller and, therefore, bigger differences between the tested and empirical value are not statistically significant. When the sample size increases the accuracy increases and smaller differences between the test and empirical value can be statistically significant.

Hypothesis testing for equality between the populational mean and a constant

Case 1. If $X \in N(\mu, \sigma^2)$, where $\mu = \mathbb{E}X$ and σ^2 is known.

$H_0 : \mu = \mu_0$, where $\mu_0 = const.$. The last means that the differences between the average of the sample and μ_0 is not statistically significant.

H_A could be formulated in one of the following ways.

✓ $H_A : \mu < \mu_0$,
In this case later on we speak about **left-sided critical area**.

✓ $H_A : \mu > \mu_0$,
In this case later on we speak about **right-sided critical area**.

✓$H_A : \mu \neq \mu_0$.

In this case later on we speak about **two-sided critical area**.

Although H_0 is for equality, when we analyze the result we explain it as the opposite of H_A.

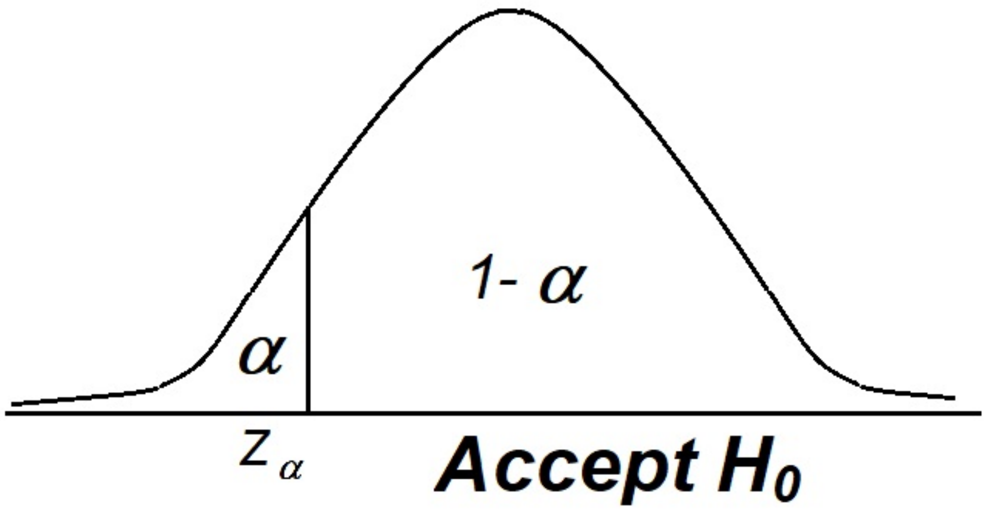
We chose **type I error**

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

The critical area for H_0 is

✓ In case of **left-sided**

$$W_\alpha = \left\{ \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha,\ N(0,1)} \right\}$$

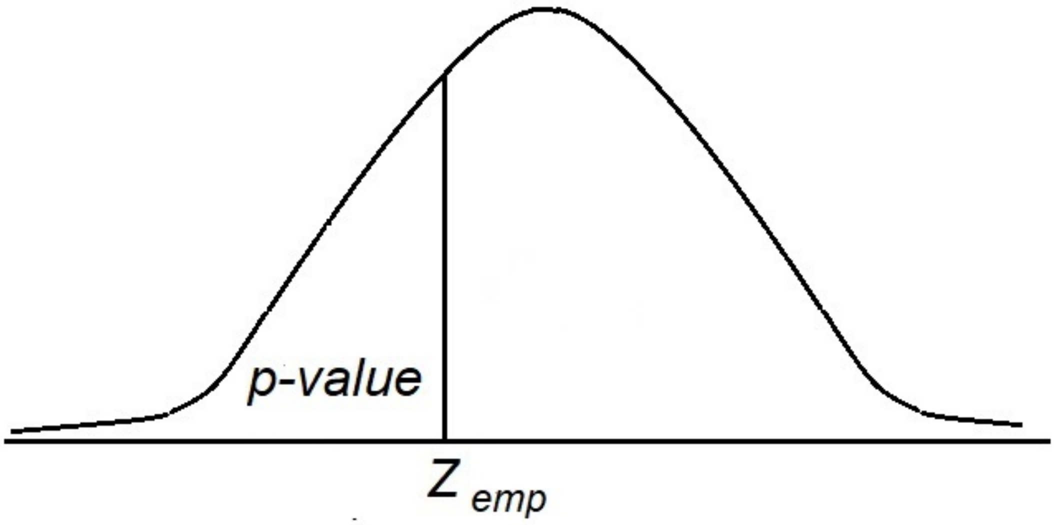


When we compute

$z_{emp} = \dfrac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ from the data we can compare z_{emp} with $z_{\alpha,N(0,1)}$.

If $z_{emp} > z_{\alpha,N(0,1)}$ 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 $z_{emp} \leq z_{\alpha,\ N(0,1)}$, then the sample is in the critical area of H_0. In this case we reject H_0 and assume H_A. The mean of the observed random variable X, i.e. $\mathbb{E}X$ is statistically significant less than μ_0.

$$p{-}value = \mathbb{P}(\eta \leq z_{emp})$$



Example 5:

Let us assume that a car gets $X \in N(\mu, 4)$ mpg. A manufacturer claims $\mu_0 = 25$ mpg. A consumer group asks 10 owners of this model to calculate their mpg and the mean value was 22 mpg. Is the manufacturer's claim supported? Check the hypothesis for significance level 0.05.

In this case we can have **one-sided hypothesis test**

$H_0 : \mu = 25$ - The manufacturer's claim cannot be rejected.
$H_A : \mu < 25$ - The manufacturer's claim can be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 22; sigma <- 2; n <- 10
> zemp <- (xbar - 25) / (sigma / sqrt(n)); zemp
[1] -4.743416
```

First way to solve this is to **compare the empirical value with the critical value** $z_{\alpha, N(0,1)}$

```
> alpha <- 0.05
> zcritical <- qnorm(alpha); zcritical
[1] -1.644854
```

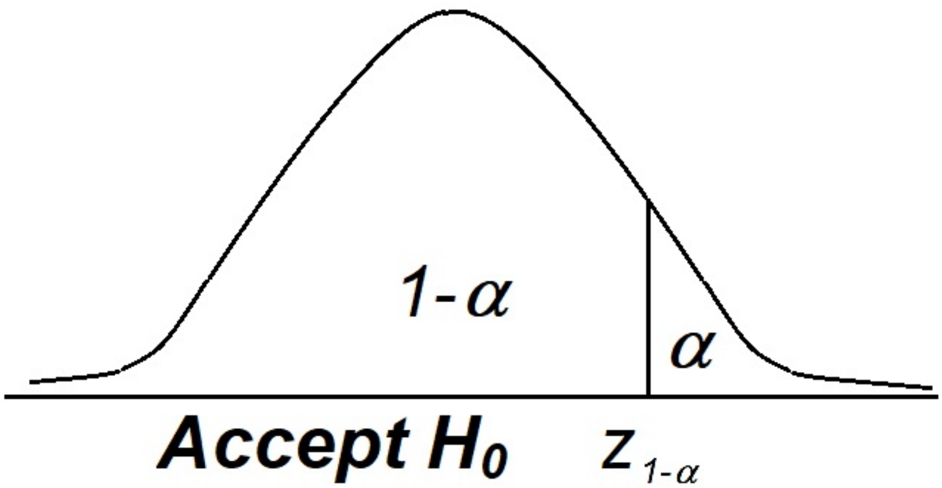
The empirical value -4.743416 is less than $z_{0.05,N(0,1)} = -1.644854$. Therefore the sample is **in the critical area for the null hypothesis** and we **reject** H_0. The manufacturer's claim is suspicious.

Second way to solve this is to compute the $p\text{−}value$ and to compare it with α.

```
> pnorm(zemp, 0, 1)
[1] 1.050718e-06
```

The $p\text{−}value < 0.05 = \alpha$. The sample is in the critical area for the null hypothesis and we reject H_0. The manufacturer's claim is suspicious.

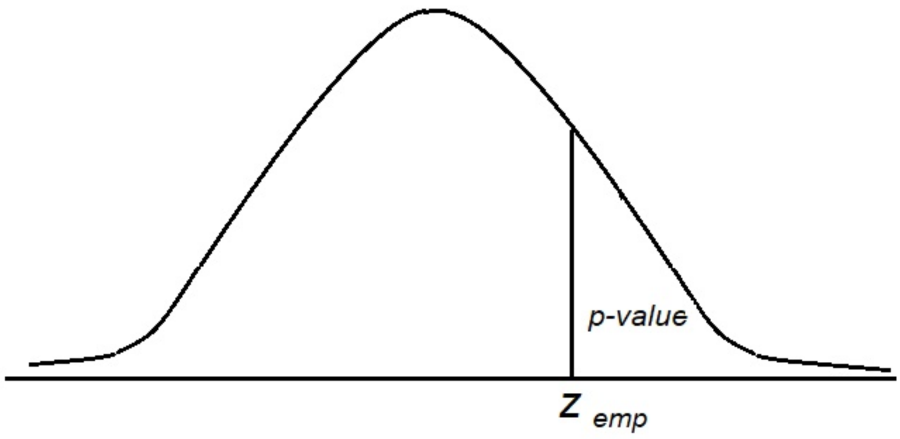
✓ In case of right-sided critical area.



$$W_\alpha = \left\{ \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_{1-\alpha,N(0,1)} \right\}$$

When we compute $z_{emp} = \dfrac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ from the data we can compare z_{emp} with $z_{1-\alpha,N(0,1)}$. If $z_{emp} < z_{1-\alpha,N(0,1)}$, 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 $z_{emp} \geq z_{1-\alpha,N(0,1)}$, then the sample is in the critical area of H_0. In this case we reject H_0 and assume H_A. The mean of the observed random variable $\mathbb{E}X$ is statistically significantly bigger than μ_0.

$$p-value = \mathbb{P}(\eta \geq z_{emp}), \eta \in N(0,1).$$



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 and assume H_A. The mean of the observed random variable $\mathbb{E}X$ is statistically significantly bigger than μ_0.

Example 6:

Let us assume that the standard weight of a mushroom is at most 40 grams and it is $X \in N(\mu,9)$. A customer claims that the mushrooms of a producer are not standard. The producer chooses 15 mushrooms at random and weighs them. The average of the mushrooms was 41 grams. Can the customer's claim be rejected?

In this case we can have **right-sided hypothesis test**

$H_0 : \mu = 40$ - The customer's claim can be rejected.

$H_A : \mu > 40$ - The customer's claim can not be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 41; sigma <- 9; n <- 15
```

```
> zemp <- (xbar - 40) / (sigma / sqrt(n)); zemp
[1] 0.4303315
```

First way to solve this is to compare the empirical value with the critical value $z_{1-\alpha,N(0,1)}$.

```
> alpha <- 0.05
> zcritical <- qnorm(1 - alpha); zcritical
[1] 1.644854
```

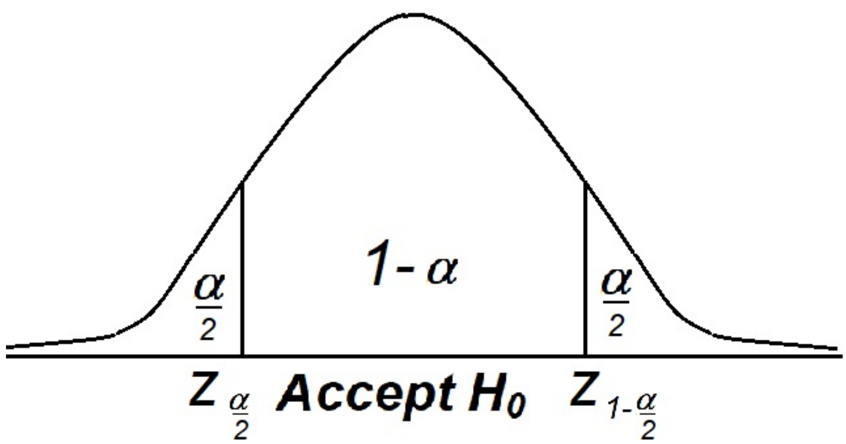
The empirical value 0.4303315 is less than $z_{1-0.05,N(0,1)} = z_{0.95,N(0,1)} = 1.644854$. Therefore, the sample is not in the critical area for the null hypothesis and we have no reason to reject H_0. The customer's claim can be rejected.

Second way to solve this is to compute the $p-value$ and to compare it with α.

```
> pnorm(zemp, 0, 1, lower.tail = FALSE)
[1] 0.3334773
```

The $p-value = 0.33 > 0.05 = \alpha$. The sample is not in the critical area for the null hypothesis and we have no reason to reject H_0. The 41 is not statistically significantly bigger than 40. The customer's claim can be rejected.

✓ In case of two-sided critical area.

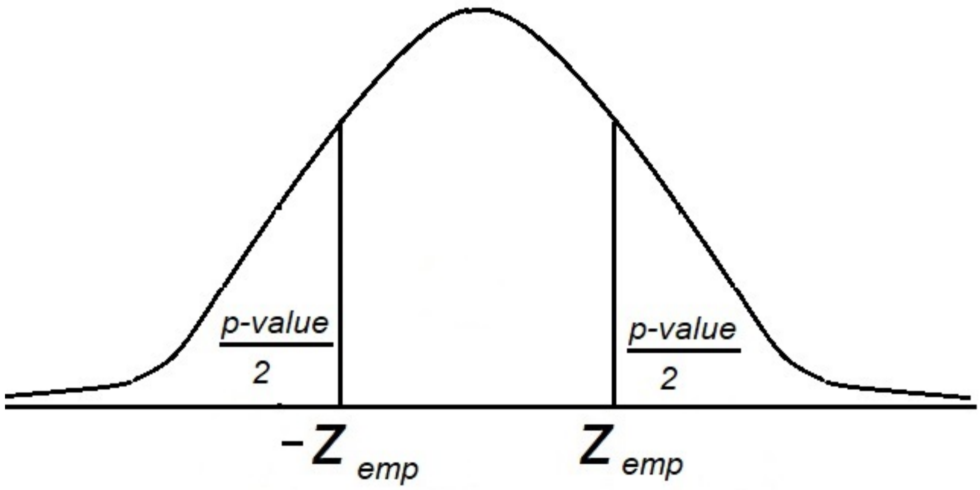


$$W_\alpha = \left\{ \frac{|\overline{X}_n - \mu_0|}{\frac{\sigma}{\sqrt{n}}} \geq z_{1-\frac{\alpha}{2}, N(0,1)} \right\}$$

When we compute $z_{emp} = \frac{|\overline{X}_n - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$ from the data we can compare z_{emp} with $z_{1-\frac{\alpha}{2},N(0,1)}$ and $-z_{1-\frac{\alpha}{2},N(0,1)}$.

If $-z_{1-\frac{\alpha}{2},N(0,1)} < z_{emp} < z_{1-\frac{\alpha}{2},N(0,1)}$, 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 $z_{emp} \leq -z_{1-\frac{\alpha}{2},N(0,1)}$ or $z_{1-\frac{\alpha}{2},N(0,1)} \leq z_{emp}$, 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 $\mathbb{E}X$ is statistically significantly different from μ_0.

$$p{-}value = \mathbb{E}\left(|\eta| \geq z_{emp} \right) = \mathbb{P}\left(\eta \leq -z_{emp} \cup \eta \geq z_{emp} \right), \eta \in N(0,1)$$



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 and assume H_A. The mean of the observed random variable $\mathbb{E}X$ is statistically significantly different from μ_0.

Example 7:

Let us assume that the standard weight of a mushroom is 40 grams and it is $X \in N(\mu, 9)$. A customer claims that the mushrooms of a producer are not standard. The producer chooses 15 mushrooms at random and weighs them. The average of the mushrooms was 41 grams. Can the customer's claim be rejected?

In this case we can have **two-sided hypothesis test**

$H_0 : \mu = 40$ - The customer's claim can be rejected.

$H_A : \mu \neq 40$ - The customer's claim can not be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 41; sigma <- 3; n <- 15
> zemp <- abs(xbar - 40) / (sigma / sqrt(n)); zemp
[1] 1.290994
```

First way to solve this is to compare the empirical value with the critical values $-z_{1-\frac{\alpha}{2}, N(0,1)}$ and $z_{1-\frac{\alpha}{2}, N(0,1)}$.

```
> zcritical1<- -qnorm(1 - alpha/2); zcritical1
[1] -1.959964
> zcritical2<- qnorm(1 - alpha/2); zcritical2
[1] 1.959964
```

The empirical value is 1.290994 in the interval between $-z_{1-\frac{0.05}{2} < N(0,1)} = -z_{0.975, N(0,1)} = -1.959963$ and $z_{0.975, N(0,1)} = 1.959964$. Therefore, the sample is not in the critical area for the null hypothesis and we have no reason to reject H_0. The customer's claim can be rejected.

Second way to solve this is to compute the $p-value$

and to compare it with α.

```
> 2*pnorm(zemp, 0, 1, lower.tail = FALSE)
[1] 0.1967056
```

The $p-value = 0.1967 > 0.05 = \alpha$. The sample is in not the critical area for the null hypothesis and we have no reason to reject H_0. The difference between 41 and 40 is not statistically significant. The customer's claim can be rejected.

Case 2. If the observed random variable $X \in N(\mu, \sigma^2)$ and σ^2 is **unknown**

In this case we estimate σ^2 by its unbiased estimate s^2 from the sample and replace $N(0,1)$ with $t(n-1)$.

As far as when the degrees of freedom of t distribution are at least 30 it almost coincides with $N(0,1)$ distribution the results from the next approach is different from the previous one only when the sample size $n \leq 30$.

The tested hypothesis are the analogous as in the previous case.

The critical areas for H_0 are as follows:

✓ If the critical area is **left-sided**

$$W_\alpha = \left\{ \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \leq t_{\alpha,t(n-1)} \right\},$$

where $t_{\alpha,t(n-1)}$ is the α quantile of $t(n-1)$ distribution.

When we compute $t_{emp} = \dfrac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}}$ from the data we can compare t_{emp} with $t_{\alpha,t(n-1)}$. If $t_{emp} > t_{\alpha,t(n-1)}$ 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 $t_{emp} \leq t_{\alpha,t(n-1)}$, then the sample is in the critical area of H_0. In this case we reject H_0 and assume H_A.

The mean of the observed random variable $\mathbb{E}X$ is statistically significantly less than μ_0.

$$p\text{−}value = \mathbb{P}(\eta \leq t_{emp}),\ \eta \in t(n-1)$$

If $p\text{−}value > \alpha$, 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 $p\text{−}value < \alpha$, 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 $\mathbb{E}X$ is statistically significantly less than μ_0.

Example 8:

Let us assume that a car gets $X \in N(\mu, \sigma^2)$. A manufacturer claims $\mu_0 = 25$ mpg. A consumer group asks 10 owners of this model to calculate their mpg. The mean value in the sample was 22 mpg. and the standard deviation in the sample was 3 mpg. Is the manufacturer's claim supported?

In this case we can have **one-sided hypothesis test**

$H_0 : \mu = 25$ - The manufacturer's claim can not be rejected.
$H_A : \mu < 25$ - The manufacturer's claim can be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 22; s <- 3; n <- 10
> temp <- (xbar - 25) / (s / sqrt(n)); temp
[1] -3.162278
> alpha <- 0.05
> tcritical <- qt(alpha, n - 1); tcritical
[1] -1.833113
> pt(temp, n-1)
[1] 0.005753993
```

The $p\text{−}value = 0.005753993 < 0.05 = \alpha$. The sample is in the critical area for the null hypothesis and we reject H_0. The manufacturer's claim is suspicious.

✓ If the critical area is **right-sided**

$$W_\alpha = \left\{ \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha,t(n-1)} \right\}$$

When we compute $t_{emp} = \dfrac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}}$ from the data we can compare t_{emp} with $t_{1-\alpha,t(n-1)}$. If $t_{emp} < t_{1-\alpha,t(n-1)}$, 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 $t_{rmp} \geq t_{1-\alpha,t(n-1)}$, then the sample is in the critical area of H_0. In this case we reject H_0 and assume H_A. The mean of the observed random variable $\mathbb{E}X$ is statistically significantly bigger than μ_0.

$$W_\alpha = \left\{ \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha,t(n-1)} \right\}.$$

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 $\mathbb{E}X$ is statistically significantly bigger than μ_0.

Example 9:

Let us assume that the standard weight of a mushroom is at most 40 grams and it is $X \in N(\mu, \sigma^2)$. A customer claims that the mushrooms of a producer are not standard. The producer chooses 15 mushrooms at random and weighs them. The average of the mushrooms is 41 grams. The standard deviation of the sample is $s = 3$ grams. Can the customer's claim be rejected?

In this case we can have **right-sided hypothesis test**

$H_0 : \mu = 40$ - The customer's claim can be rejected.
$H_A : \mu > 40$ - The customer's claim can not be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 41; s <- 3; n <- 15
> temp <- (xbar - 40) / (s / sqrt(n)); temp
[1] 1.290994
> alpha <- 0.05
> tcritical <- qt(1 - alpha, n - 1); tcritical
[1] 1.76131
> pt(temp, n - 1, lower.tail = FALSE)
[1] 0.1088085
```

The $p-value = 0.1088 > 0.05 = \alpha$. The sample is not in the critical area for the null hypothesis and we have no reason to reject H_0.
The 41 is not statistically significantly bigger than 40. The customer's claim can be rejected.

✓ If the critical area is **two-sided**

$$W_\alpha = \left\{ \frac{|\overline{X}_n - \mu_0|}{\frac{s}{\sqrt{n}}} \geq t_{1-\frac{\alpha}{2},t(n-1)} \right\}$$

where $t_{1-\frac{\alpha}{2},t(n-1)}$ is the $1 - \dfrac{\alpha}{2}$ quantile of $t(n-1)$ distribution.

When we compute $t_{emp} = \dfrac{|\overline{X}_n - \mu_0|}{\frac{s}{\sqrt{n}}}$ from the data we can compare t_{emp} with $t_{1-\frac{\alpha}{2},t(n-1)}$ and $-t_{1-\frac{\alpha}{2},t(n-1)}$.

If $-t_{1-\frac{\alpha}{2}<t_{emp}<t_{1-\frac{\alpha}{2},t(n-1)}}$, 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 $t_{emp} \leq -t_{1-\frac{\alpha}{2},t(n-1)}$ or $t_{1-\frac{\alpha}{2},t(n-1)} \leq t_{emp}$, 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 $\mathbb{E}X$ is statistically significantly different from μ_0.

$$p-value = \mathbb{P}(|\eta| \geq t_{emp}) = \mathbb{P}(\eta \leq -t_{emp} \cup \eta \geq t_{emp}),\ \eta \in t(n-1)$$

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 $\mathbb{E}X$ is statistically significantly different from μ_0.

Example 10:

Let us assume that the standard weight of a mushroom is 40 grams and it is $X \in N(\mu, \sigma^2)$. A customer claims that the mushrooms of a producer are not standard. The producer chooses 15 mushrooms at random and weighs them. The average of the mushrooms is 41 grams. The standard deviation in the sample is $s = 3$ grams. Can the customer's claim be rejected?

In this case we can have **two-sided hypothesis test**

$H_0 : \mu = 40$ - The customer's claim can be rejected.
$H_A : \mu \neq 40$ - The customer's claim can not be rejected.

The data are summarized and we have the average. So,

```
> xbar <- 41; s <- 3; n <- 15
> temp <- abs(xbar - 40) / (3 / sqrt(n)); temp
[1] 1.290994
> alpha <- 0.05
> tcritical <- qt(1 - alpha/2, n - 1); tcritical
[1] 2.144787
> 2 * pt(temp, n-1, lower.tail = FALSE)
[1] 0.217617
```

The $p-value = 0.2176 > 0.05 = \alpha$. The sample is not in the critical area for the null hypothesis and we have no reason to reject H_0. The difference between 41 and 40 is not statistically significant. The customer's claim can be rejected.

If the data are not summarized in advance we can use $t.test$ function in R.

Example 11:

We already use the `puerto` data set. Let's assume that the income is normally distributed and to see if we can say that the mean of Puerto Rican immigrants to Miami's income is equal to 277.

In this case we have **two-sided hypothesis test**

$$H_0 : \mu = 277$$
$$H_A : \mu \neq 277$$

```
> 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
> t.test(puerto, mu = 277, conf.level = 0.95, alternative
= "two.sided")

    One Sample t-test
```

```
data:  puerto
t = 0.046571, df = 49, p-value = 0.963
alternative hypothesis: true mean is not equal to 277
95 percent confidence interval:
 255.9244 299.0756
sample estimates:
mean of x 
    277.5
```

The $p-value = 0.963 > 0.05 = \alpha$. The sample isn't in the critical area for the null hypothesis, so we have no reason to reject H_0.

Case 3. When the **sample size is large** and the **variance** of the observed random variable X is **finite** the previous approaches can be used without any information if X is normal or not.

Example 12:

We already use the `puerto` data set. Let's assume that the income has a **finite variance** and to see if we can say that the mean of Puerto Rican immigrants to Miami's income is equal to 277.

```
> length(puerto)
[1] 50
```

The solution is the same as in the Example 11.

Hypothesis testing for the median

In cases when we have no information if the observed random variable X has finite or infinite variance the Wilcoxon (Mann-Whitney) test can be useful.

Example 13:

Study of cell-phone usage for a user gives the following lengths for the calls.

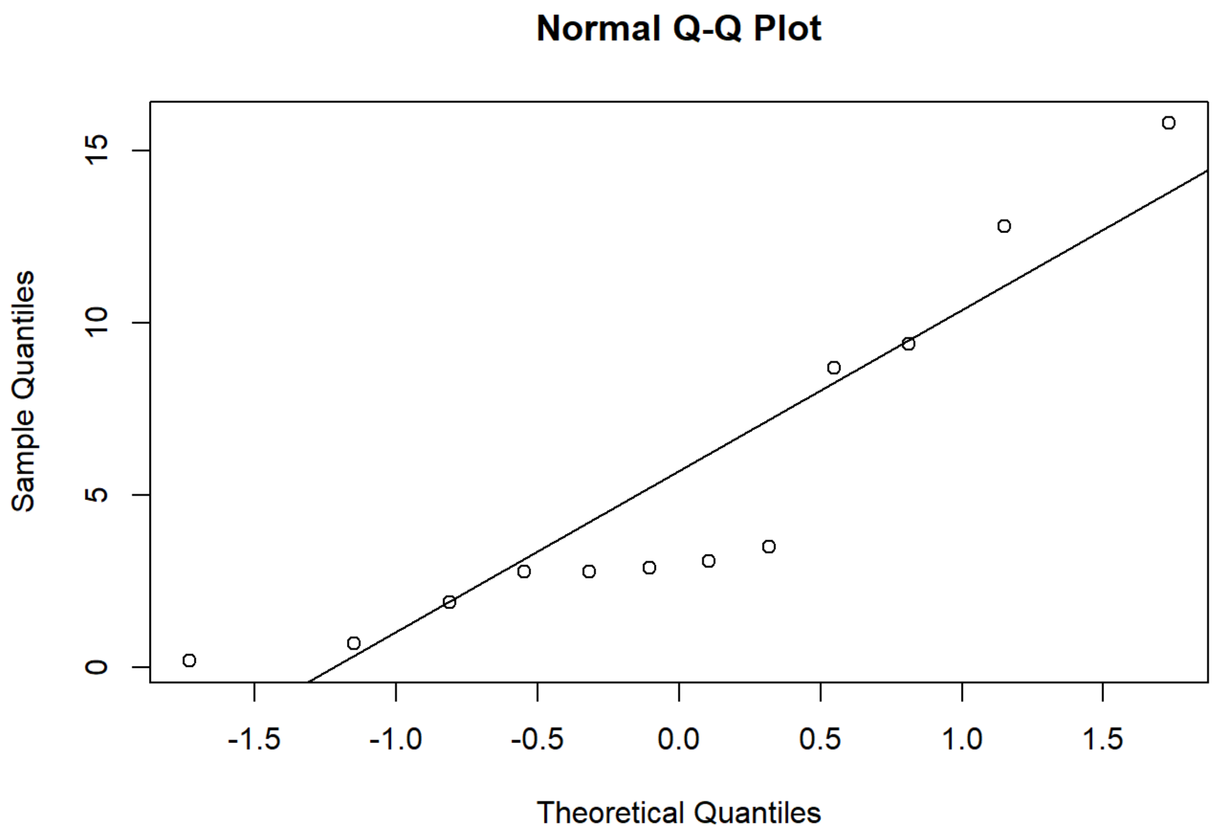
```
> x <- c(12.8, 3.5, 2.9, 9.4, 8.7, 0.7, 0.2, 2.8, 1.9, 
2.8, 3.1, 15.8)
```

What is the appropriate test for center?

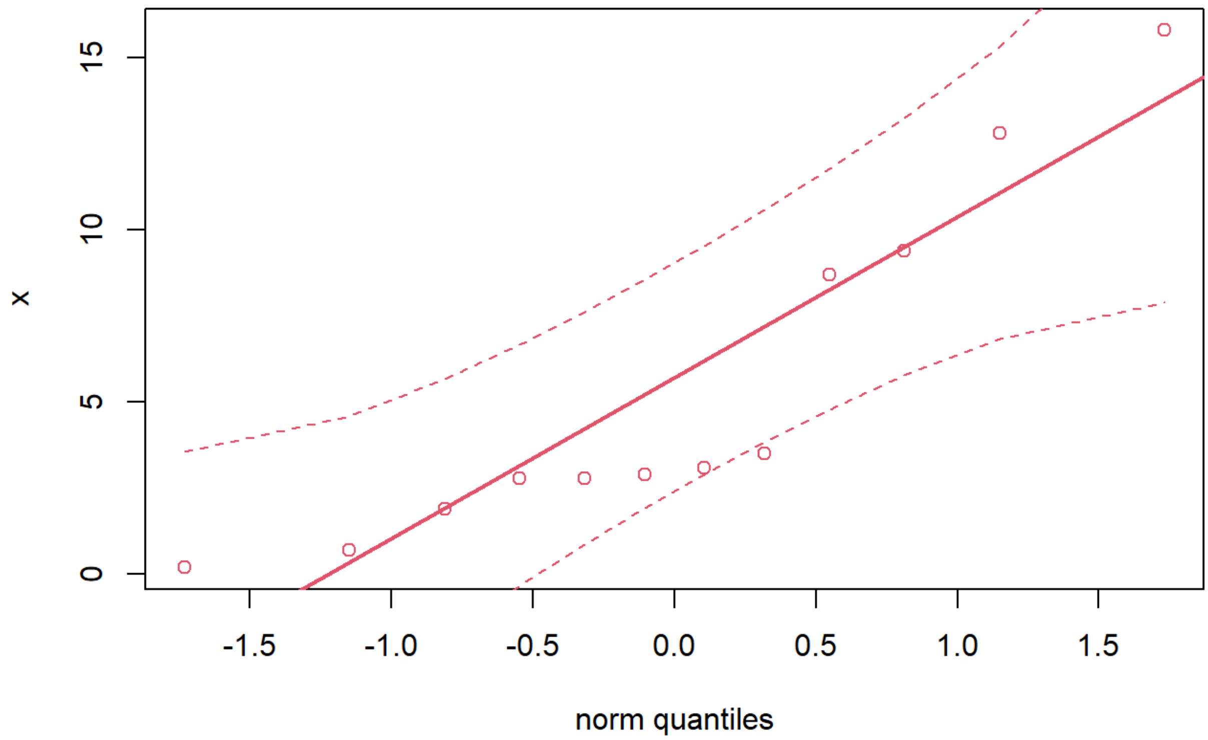
Let's first check if the distribution of the observed random variable is normal.

H_0 : X is normally distributed
H_A : X isn't normally distributed

```
> qqnorm(x)
> qqline(x)
```




```
> 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(x)
```




```
> shapiro.test(x)

	Shapiro-Wilk normality test

data:  x
W = 0.83988, p-value = 0.0276
```

As we see from the graphics and from the Shapiro test $p-value = 0.0276 < 0.05 = \alpha$ we reject H_0, so the data is **not normally distributed**. We are going to make a one-side hypothesis test for the median.

$$H_0 : Me = 5$$
$$H_A : Me > 5$$

We can use the `wilcox.test` function

```
> wilcox.test(x, mu = 5, alternative = "greater")
Warning in wilcox.test.default(x, mu = 5, alternative =
"greater"): cannot
compute exact p-value with ties
```

```
	Wilcoxon signed rank test with continuity correction

data:  x
V = 39, p-value = 0.5156
alternative hypothesis: true location is greater than 5
```

The $p-value = 0.5156 > 0.05 = \alpha$. The sample isn't in the critical area for the null hypothesis and we don't have reason to reject H_0.

Example 14: Let us come back to Example 11.

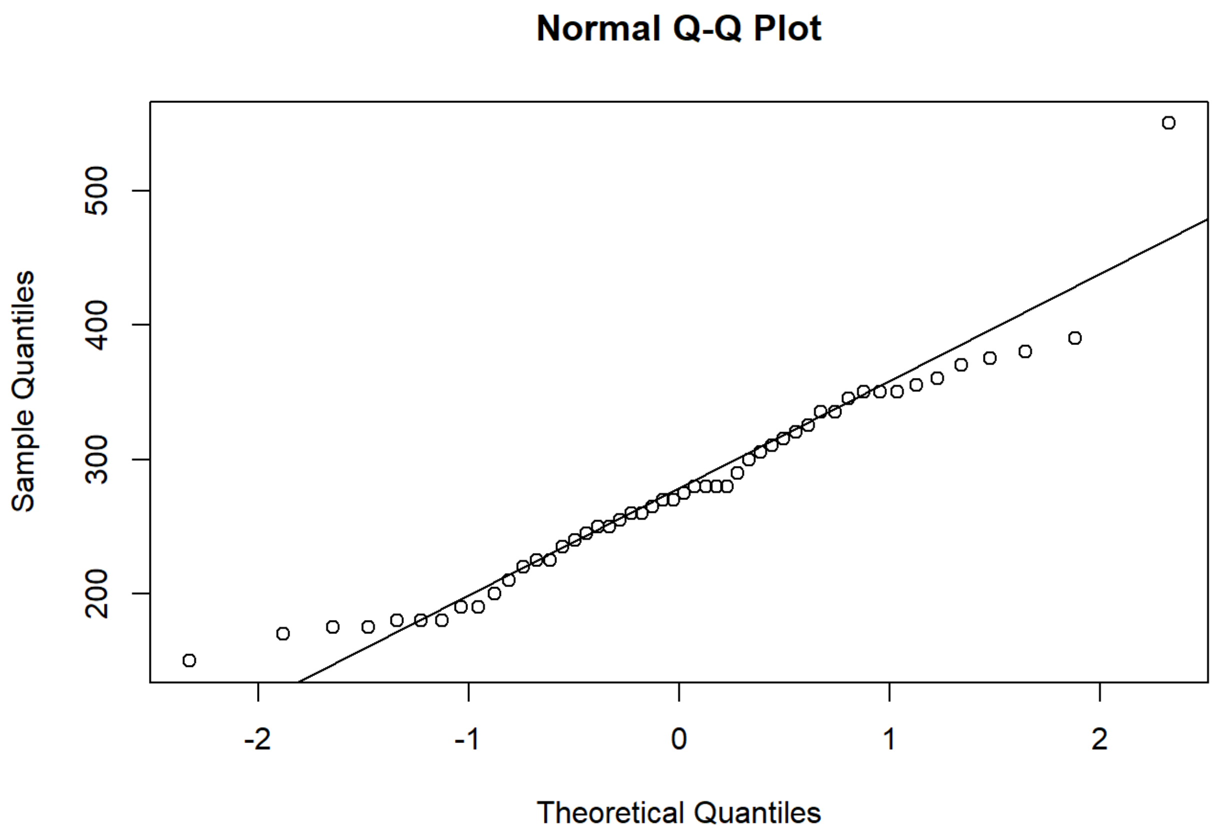
We already use the `puerto` data set. Let's us see if we can say that the median of Puerto Rican immigrants to Miami's income is equal to 273.

What is the appropriate test for center?

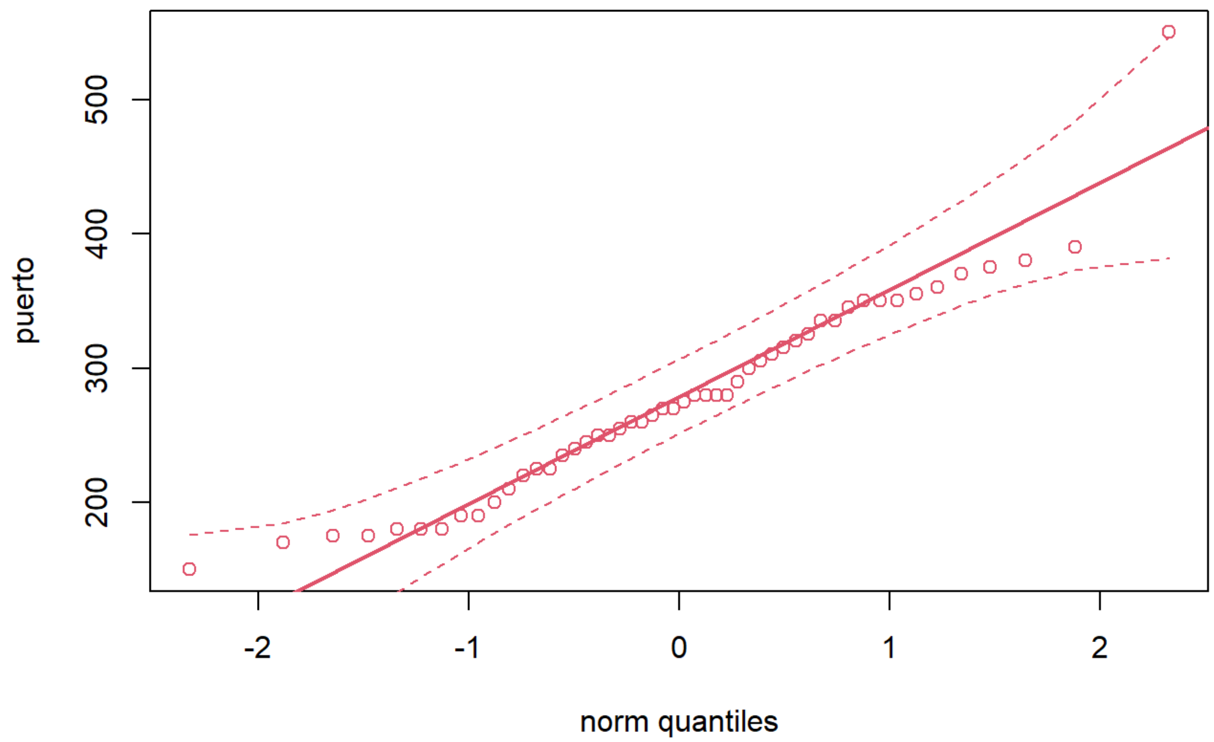
Let's first check if the distribution of the observed random variable is normal.

H_0 : X is normally distributed
H_A : X isn't normally distributed

```
> qqnorm(puerto)
> qqline(puerto)
```




```
> qqplot.das(puerto)
```



```
> shapiro.test(puerto)

	Shapiro-Wilk normality test

data:  puerto
W = 0.94538, p-value = 0.02212
```

As we see from the graphics and from the Shapiro test $p-value = 0.0221 < 0.05 = \alpha$ we reject H_0, so the data is not normally distributed. The distribution looks skewed with a possibly heavy tail. We are going to make a one-side hypothesis test for the median.

$$H_0 : Me = 273$$
$$H_A : Me \neq 273$$

We can use the `wilcox.test` function

```
> wilcox.test(puerto, mu = 273, alternative =
"two.sided")

	Wilcoxon signed rank test with continuity correction

data:  puerto
V = 642, p-value = 0.9692
alternative hypothesis: true location is not equal to 273
```

The $p-value = 0.9692 > 0.05 = \alpha$. The sample isn't in the critical area for the null hypothesis, so we have no reason to reject H_0.

Rank tests

Similar hypothesis test for the median.

```
> x <- c(12.8, 3.5, 2.9, 9.4, 8.7, 0.7, 0.2, 2.8, 1.9,
2.8, 3.1, 15.8)
> simple.median.test(x, median = 5)
[1] 0.3876953
```

The $p-value = 0.3876953 > 0.05 = \alpha$. The sample isn't in the critical area for the null hypothesis, so we have no reason to reject H_0.

```
> simple.median.test(x, median = 10)
[1] 0.03857422
```

The $p-value = 0.0385742 < 0.05 = \alpha$. The sample is in the critical area for the null hypothesis, so we reject H_0.