Statistics in R

Session-10

Contents

- Statistical tests
- Quiz

Statistical test

Contents

- Statistical test
- Quiz

Example

Let’s say we are interested in investigating monthly expenses of students living in Tallinn.

- **You created a google form to collect data anonymously.**

Age | Education | Gender | Expenses

Example

N = 100 responses

Age | Education | Gender | Expenses

	age	expense	gender	education
1	22	375.5307	M	Bachelor
2	21	305.1428	M	Bachelor
3	30	338.7994	M	Master
4	22	325.2335	F	Master
5	26	387.8250	F	Bachelor
6	29	348.7039	M	Master

Descriptive statistics

N = 100 responses

Age | Education | Gender | Expenses

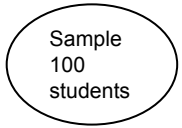
```
> summary(mdata)
      age            expense          gender             education
 Min.   :18.00   Min.   :281.2   Length:100         Length:100
 1st Qu.:21.00   1st Qu.:327.4   Class :character   Class :character
 Median :24.00   Median :347.7   Mode  :character   Mode  :character
 Mean   :24.06   Mean   :349.9
 3rd Qu.:27.00   3rd Qu.:373.5
 Max.   :30.00   Max.   :416.6
```

Descriptive statistics

N = 100 responses

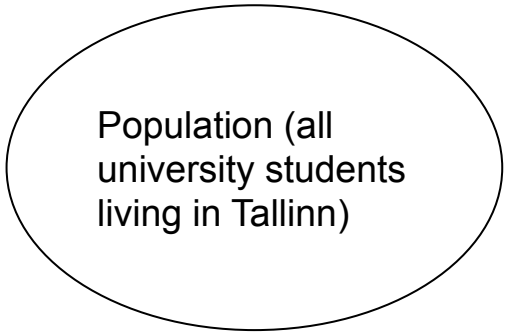
Age | Education | Gender | Expenses

It could be possible that on average a university students spends 350 euros per month as expenses.



What we know

Sample mean 350



What we want to find out

population mean ?

t-test

N = 100 responses

It could be possible that on average a university students spends 350 euros per month as expenses.

Assumptions:

- Normality
- Data independence (random sample)

Attribute of interest: Expenses (single attribute)
Type: Numeric

One sample t-test

t-test

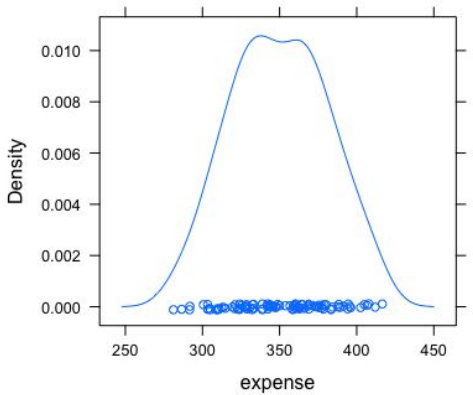
N = 100 responses

It could be possible that on average a university students spends 350 euros per month as expenses.

Assumptions:

- Normality
- Data independence (random sample)

```
densityplot(~expense,data=mdata)
```



Hypothesis testing

Null hypothesis: This is a hypothesis of no difference.

It always has an equal sign in the statement.

Alternative hypothesis: This is a contradictory statement.

It never has an equal sign in the statement.

H0: The population mean is equal to 350.

H1: The population mean is not equal to 350

https://opentextbc.ca/introstatopenstax/chapter/null-and-alternative-hypotheses/

Hypothesis testing

H0: The population mean is equal to 350.

H1: The population mean is not equal to 350

Level of significance

You can select 5%, 1%, .1%

$\alpha = 5\ \%$ or .05

Perform the test

```
> t.test(mdata$expense,mu=350)

        One Sample t-test

data:  mdata$expense
t = -0.023605, df = 99, p-value = 0.9812
alternative hypothesis: true mean is not equal to 350
95 percent confidence interval:
 343.7131 356.1391
sample estimates:
mean of x
 349.9261
```

Perform the test

p-value =< α : Alternative hypothesis

p-value > α : Null hypothesis

p-value = .9812

Draw the conclusion

The null hypothesis can not be rejected. Thus, the population mean may be 350 euros (or the average monthly expenditure of university students living in Tallinn may be 350 euros)

Example

N = 100 responses

Age | Education | Gender | Expenses

	age	expense	gender	education
1	22	375.5307	M	Bachelor
2	21	305.1428	M	Bachelor
3	30	338.7994	M	Master
4	22	325.2335	F	Master
5	26	387.8250	F	Bachelor
6	29	348.7039	M	Master

Is there any difference between male and female students in terms of expenditure per month?

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

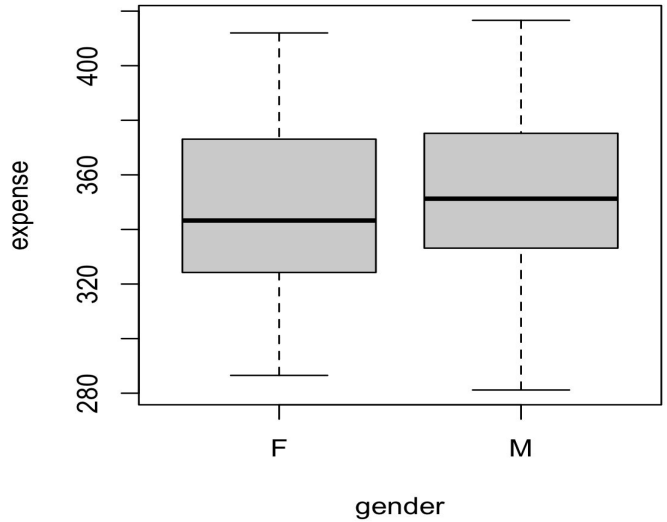
- Descriptive statistics
- Check assumptions
- State Null and alternative hypotheses
- Set level of significance
- Perform the test and get the p-value
- Conclude on the basis of p-value

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- **Descriptive statistics**
- Check assumptions
- State Null and alternative hypotheses
- Set level of significance
- Perform the test and get the p-value
- Conclude on the basis of p-value



Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- Descriptive statistics
- **Check assumptions**
- State Null and alternative hypotheses
- Set level of significance
- Perform the test and get the p-value
- Conclude on the basis of p-value

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- Descriptive statistics
- Check assumptions
- **State Null and alternative hypotheses**
- Set level of significance
- Perform the test and get the p-value
- Conclude on the basis of p-value

H0: The expenses of male students are equal to the expenses of female students.
H1: The expenses of male students are not equal to the expenses of female students.

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- Descriptive statistics
- Check assumptions
- State Null and alternative hypotheses
- **Set level of significance**
- Perform the test and get the p-value
- Conclude on the basis of p-value

α = 5%

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- Descriptive statistics
- Check assumptions
- State Null and alternative hypotheses
- Set level of significance
- **Perform the test and get the p-value**
- Conclude on the basis of p-value

```
> t.test(mdata$expense~mdata$gender,var=T)

        Two Sample t-test

data:  mdata$expense by mdata$gender
t = -0.72375, df = 98, p-value = 0.4709
alternative hypothesis: true difference in
95 percent confidence interval:
 -17.055631   7.939682
sample estimates:
mean in group F mean in group M
       347.8294        352.3874
```

Example

Is there any difference between male and female students in terms of expenditure per month?

N = 100 responses

- Descriptive statistics
- Check assumptions
- State Null and alternative hypotheses
- Set level of significance
- Perform the test and get the p-value
- **Conclude on the basis of p-value**

p-value = 0.4709

α = 5% or .05

p-value =< α : Alternative hypothesis

p-value > α : Null hypothesis

Thank you

http://bit.ly/content-r