Biostatistics I: Hypothesis testing

Continuous data: Correlation tests

Eleni-Rosalina Andrinopoulou

Department of Biostatistics, Erasmus Medical Center

e.andrinopoulou@erasmusmc.nl

@erandrinopoulou

In this Section

- Pearson correlation test
- Spearman correlation test
- Examples

Pearson correlation test: Theory

Assumptions

- The variables must be continuous
- There is a linear relationship between the two variables
- The data has homoscedasticity
- The variables is approximately normally distributed
- The two variables represent paired observations
- The variables do not contain any outliers

Pearson correlation test: Theory

Scenario

Is the height of the students in my university linearly associated with their weight?

Connection with linear regression

$$scale(y_i) = \beta_0 + \beta_1 scale(x_i) + \epsilon_i$$
$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Pearson correlation test: Theory

Scenario

Is the height of the students in my university linearly associated with their weight?

Alternatively

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Pearson correlation test: Theory

Test statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

- Sample correlation: r
- Number of subjects: n

Sampling distribution

- t-distribution with $df = n - 2$
- Critical values and p-value

Pearson correlation test: Theory

Type I error

- Normally α = 0.05

Draw conclusions

- Compare test statistic (t) with the critical values or the p-value with α

Pearson correlation test: Application

Scenario

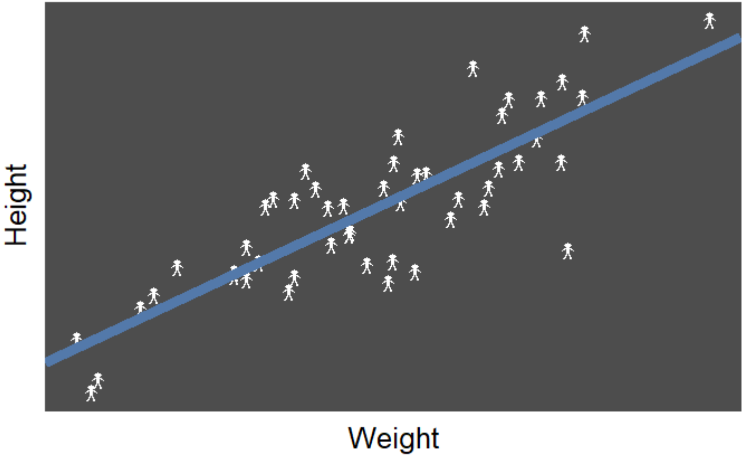
Is the height of the students in my university linearly associated with their weight?

Hypothesis

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Pearson correlation test: Application

Collect and visualize data



Pearson correlation test: Application

Hypothesis

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Test statistic

Let's assume that:

- Sample correlation
  $r = 0.83$
- Number of subjects
  $n = 50$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.83\sqrt{50-2}}{\sqrt{1-0.83^2}} = 10.31$$

Degrees of freedom

$df = 50 - 2 = 48$

Type I error

$\alpha = 0.05$

Pearson correlation test: Application

Critical values

Using R we get the critical values from the *t*-distribution:
critical value$_{\alpha/2}$ = critical value$_{0.05/2}$

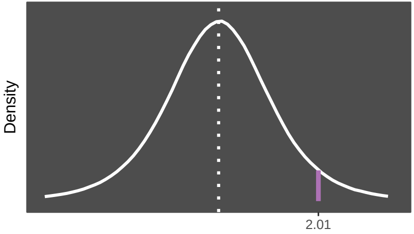
```
qt(p = 0.05/2, 48, lower.tail = FALSE)
```

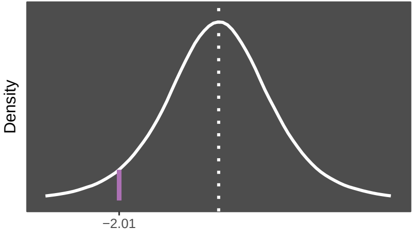
```
[1] 2.010635
```

-critical value$_{\alpha/2}$ = -critical value$_{0.05/2}$

```
qt(p = 0.05/2, 48, lower.tail = TRUE)
```

```
[1] -2.010635
```





Pearson correlation test: Application

Draw conclusions

We reject the H_0 if:

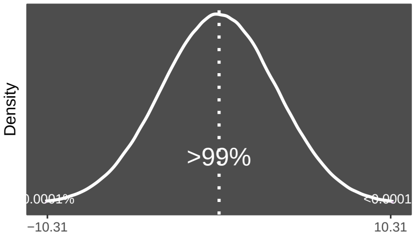
- t > critical value$_{\alpha/2}$ or t < - critical value$_{\alpha/2}$

We have $10.31 > 2.01 \Rightarrow$ we reject the H_0

Using R we obtain the p-value from the t-distribution:

```r
2 * pt(q = 10.31, df = 48,
       lower.tail = FALSE)
```

```
[1] 9.23435e-14
```



Spearman correlation test: Theory

Assumptions

- The variables must be continuous/ordinal
- There is a monotonic relationship between the two variables
- The two variables represent paired observations

Spearman correlation test: Theory

Scenario

Is the height of the students in my university monotonically associated with their weight?

Connection with linear regression

The slope becomes the correlation if we use the rank of the two variables of interest

What is rank?
Ranks are integers indicating the rank of some values. E.g. the rank of 3, 10, 16, 6, 2 is 2, 4, 5, 3, 1:

```
rank(c(3, 10, 16, 6, 2))
```

```
[1] 2 4 5 3 1
```

Spearman correlation test: Theory

Connection with linear regression

$rank(y_i) = \beta_0 + \beta_1 rank(x_i) + \epsilon_i$
$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

Alternatively

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Spearman correlation test: Theory

Test statistic

$t = \frac{r_R\sqrt{n-2}}{\sqrt{1-r_R^2}}$

- Sample correlation based on ranked data: r_R
- Number of subjects: n

Sampling distribution

- t-distribution with $df = n - 2$
- Critical values and p-value

Spearman correlation test: Theory

Type I error

- Normally $\alpha = 0.05$

Draw conclusions

- Compare test statistic (t) with the critical values$_{\alpha/2}$ or the p-value with α

Spearman correlation test: Application

Scenario

Is the height of the students in my university monotonically associated with their weight?

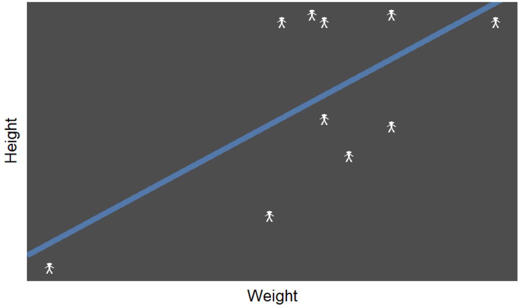
Hypothesis

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Spearman correlation test: Application

Collect and visualize data

x	y	rank_x	rank_y
1.72	49	2.0	2.0
1.85	57	7.0	3.0
1.81	62	5.5	5.0
1.81	75	5.5	7.0
1.92	76	8.5	9.5
1.36	42	1.0	1.0
1.79	76	4.0	9.5
1.92	61	8.5	4.0
1.74	75	3.0	7.0
2.09	75	10.0	7.0



Spearman correlation test: Application

Hypothesis

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Test statistic

Let's assume that:

- Sample correlation
  $r_R = 0.41$
- Number of subjects
  $n = 10$

$$t = \frac{r_R\sqrt{n-2}}{\sqrt{1-r_R^2}} = \frac{0.41\sqrt{10-2}}{\sqrt{1-0.41^2}} = 1.27$$

Degrees of freedom

$df = 10 - 2 = 8$

Type I error

$\alpha = 0.05$

Spearman correlation test: Application

Critical values

Using R we get the critical values from *t*-distribution:

critical value$_{\alpha/2}$ = critical value$_{0.05/2}$

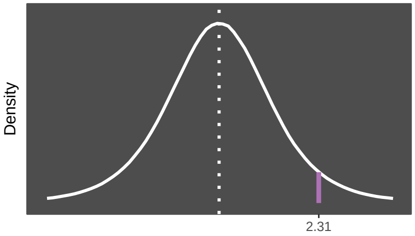
```
qt(p = 0.05/2, df = 8, lower.tail = FALSE)
```

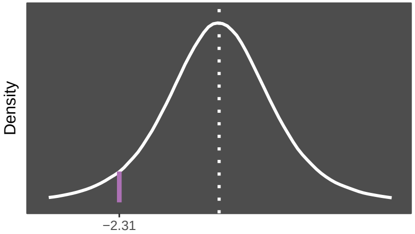
```
[1] 2.306004
```

-critical value$_{\alpha/2}$ = -critical value$_{0.05/2}$

```
qt(p = 0.05/2, df = 8, lower.tail = TRUE)
```

```
[1] -2.306004
```





Spearman correlation test: Application

Draw conclusions

We reject the H_0 if:

- t > critical value$_{\alpha/2}$ or t < - critical value$_{\alpha/2}$

We have 1.27 < 2.31 $\Rightarrow$ we do *not* reject the H_0

Using R we obtain the p-value from the t-distribution:

```r
2 * pt(q = 1.27, df = 8, lower.tail = FALSE)
```

```
[1] 0.2397765
```