Biostatistics I: Hypothesis testing

Continuous data: One-sample tests

Eleni-Rosalina Andrinopoulou

Department of Biostatistics, Erasmus Medical Center

e.andrinopoulou@erasmusmc.nl

@erandrinopoulou

Erasmus MC
University Medical Center Rotterdam

In this Section

- One-sample t-test
- One-sample Wilcoxon signed rank test
- Examples

One-sample t-test: Theory

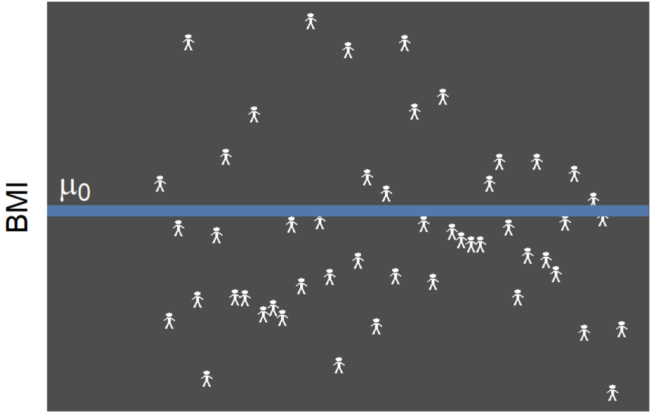
Assumptions

- The dependent variable must be continuous
- The observations are independent
- The dependent variable is approximately normally distributed
- The dependent variable does not contain any outliers

One-sample t-test: Theory

Scenario

Is the mean BMI of the students in my university different from the mean BMI of all students?



One-sample t-test: Theory

Scenario

Is the mean BMI of the students in my university different from the mean BMI of all students?

Connection with linear regression

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $x_i = 0$

$H_0 : \beta_0 = 0$
$H_1 : \beta_0 \neq 0$

One-sample t-test: Theory

Scenario

Is the mean BMI of the students in my university different from the mean BMI of all students?

Alternatively

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

More general

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

One-sample t-test: Theory

If **one-tailed**

Is the mean BMI of the students in my university larger than the mean BMI of all students?

$H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

or

Is the mean BMI of the students in my university smaller than the mean BMI of all students?

$H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$

One-sample t-test: Theory

Test statistic

$t = \frac{\bar{x} - \mu_0}{sd(x)/\sqrt{n}}$

- Sample mean: $\bar{x}$ (sample of students in my university)
- Population mean: μ_0 (all students)
- Standard deviation of the sample: $sd(x)$
- Number of subjects: n

Sampling distribution

- t-distribution with $df = n - 1$
- Critical values and p-value

One-sample t-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

If **one-tailed**: Compare test statistic with the critical value$_{\alpha}$

One-sample t-test: Application

Scenario

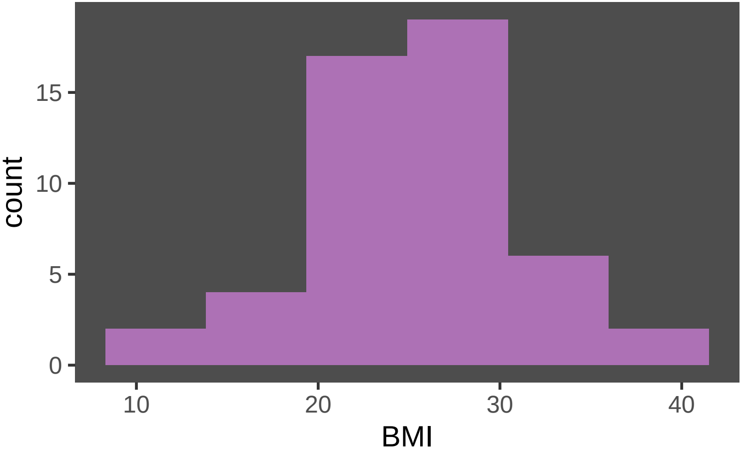
Is the mean BMI of the students in my university different from the mean BMI of all students?

Hypothesis

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

One-sample t-test: Application

Collect and visualize data



One-sample t-test: Application

Hypothesis

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

Test statistic

Let's assume that:

- Sample mean $\bar{x} = 24$
- Mean of all students $\mu_0 = 20$
- Standard deviation of the sample $sd(x) = 6$
- Number of subjects $n = 50$

$$t = \frac{\bar{x} - \mu_0}{sd(x)/\sqrt{n}} = \frac{24 - 20}{6/\sqrt{50}} = 4.7$$

Degrees of freedom

$df = 50 - 1 = 49$

Type I error

$\alpha = 0.05$

One-sample t-test: Application

Critical values

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

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

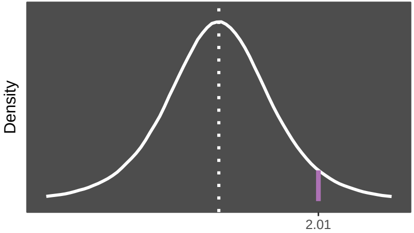
```
qt(p = 0.05/2, 49, lower.tail = FALSE)
```

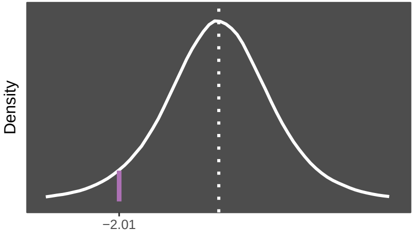
```
[1] 2.009575
```

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

```
qt(p = 0.05/2, 49, lower.tail = TRUE)
```

```
[1] -2.009575
```





One-sample t-test: Application

Critical values

If **one-tailed**

critical value$_{\alpha}$:

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

or

-critical value$_{\alpha}$:

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

One-sample t-test: Application

Draw conclusions

We reject the H_0 if:

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

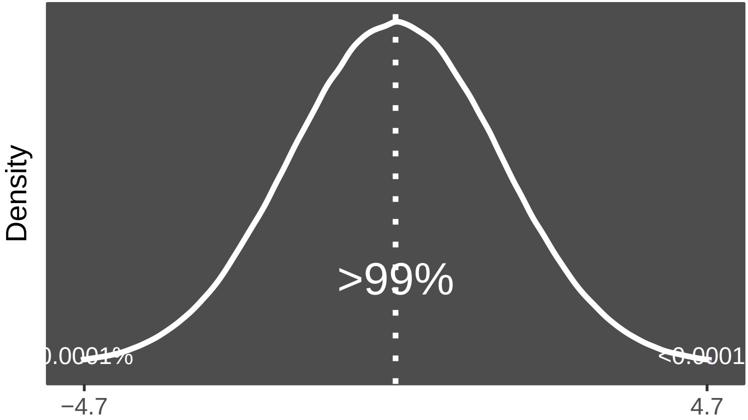
We have 4.7 > 2.01 $\Rightarrow$ we reject the H_0

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

```r
2 * pt(q = 4.7, df = 49, lower.tail = FALSE)
```

```
[1] 2.146314e-05
```

One-sample t-test: Application



One-sample t-test: Application

Draw conclusions

If **one-tailed**

We reject the H_0 if: t > critical value$_\alpha$
Using `R` we obtain the p-value from the *t*-distribution:

```
pt(q = t, df, lower.tail = FALSE)
```

or

We reject the H_0 if: t < -critical value$_\alpha$
Using `R` we obtain the p-value from the *t*-distribution:

```
pt(q = t, df, lower.tail = TRUE)
```

One-sample Wilcoxon signed rank test: Theory

Assumptions

- Population distribution is symmetric
- The observations are independent of one another

One-sample Wilcoxon signed rank test: Theory

▶ Scenario

Is the median score value of the students in my university different from the median score value of all students?

▶ Connection with linear regression

What is signed rank??
Ranks are integers indicating the rank of some values.

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

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

For signed ranks we obtain the rank according to the absolute value and we add the sign.

```
sign(c(3, -10, 16, 6, 2)) *
  rank(abs(c(3, -10, 16, 6, 2)))
```

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

One-sample Wilcoxon signed rank test: Theory

- **Connection with linear regression**

$signed\_rank(y_i) = \beta_0 + \beta_1 x_i + \epsilon_i$, where $x_i = 0$

$H_0 : \beta_0 = 0$
$H_1 : \beta_0 \neq 0$

- **Alternatively**

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

- **More general**

$H_0 : m = m_0$
$H_1 : m \neq m_0$

One-sample Wilcoxon signed rank test: Theory

If **one-tailed**

Is the median score value of the students in my university larger than the median score value of all students?

$H_0 : m = m_0$

$H_1 : m > m_0$

or

Is the median score value of the students in my university smaller than the median score value of all students?

$H_0 : m = m_0$

$H_1 : m < m_0$

One-sample Wilcoxon signed rank test: Theory

Test statistic

- Calculate the ranks of the absolute difference
  - If ties $\Rightarrow$ assign the average of the tied ranks
  - If a pair of scores are equal $\Rightarrow$ they are dropped from the analysis and the sample size is reduced
- Obtain the sum of those ranks where the difference was positive or negative $W_+ = \sum R_d^+$ or $W_- = \sum R_d^-$, where R_d are the ranks of the differences
- The test statistic (W) is the minimum of W_+ and W_-

If **one-tailed**: use either W_+ or W_- for the test statistic (W) depending on the direction of the alternative hypothesis

One-sample Wilcoxon signed rank test: Theory

Sampling distribution

- For large sample size, we use the normal approximation, that is, W is normally distributed

$$\mu_W = \frac{n(n+1)}{4} \text{ and}$$
$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$
$$z = \frac{|max(W_+, W_-) - \mu_W| - 1/2}{\sigma_W}$$

When there are ties, the mean stays the same but the variance is reduced by a quantity

- For small sample size, we can use the exact distribution (more details in the application)

Get critical values and p-value

One-sample Wilcoxon signed rank test: Theory

Type I error

- Normally $\alpha = 0.05$

Draw conclusions

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

If **one-tailed**: Compare test statistic with the critical value$_{\alpha}$

One-sample Wilcoxon signed rank test: Application

Scenario

Is the median score value of the students in my university different from the median score value of all students?

Hypothesis

$H_0 : m = m_0$
$H_1 : m \neq m_0$

One-sample Wilcoxon signed rank test: Application

Collect and visualize data

x	m_0	Difference	|Difference|	rank
9.75508	10	-0.244920	0.244920	1
11.10491	10	1.104913	1.104913	3
10.69730	10	0.697299	0.697299	2

Hypothesis

$H_0 : m = m_0$

$H_1 : m \neq m_0$

Test statistic

$W_- = 1$ and $W_+ = 5$

Type I error

$\alpha = 0.05$

One-sample Wilcoxon signed rank test: Application

Exact distribution

If $n = 3$ with no ties, we have $0, \ldots, 6$ possible values for W. Each of the three data points would be assigned a rank of either $1, 2,$ or 3. Depending on whether the data point fell above or below m_0, each of the three possible ranks $1, 2,$ or 3 would remain either a positive signed rank or become a negative.

			W	Probability
1	2	3	6	0.125
-1	2	3	5	0.125
1	-2	3	4	0.125
1	2	-3	3	0.250
-1	-2	3	3	0.250
-1	2	-3	2	0.125
1	-2	-3	1	0.125
-1	-2	-3	0	0.125

Summarize the probs:

W	Probability
0	0.125
1	0.125
2	0.125
3	0.250
4	0.125
5	0.125
6	0.125

One-sample Wilcoxon signed rank test: Application

Critical values

Using R we get the critical values from the exact distribution:

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

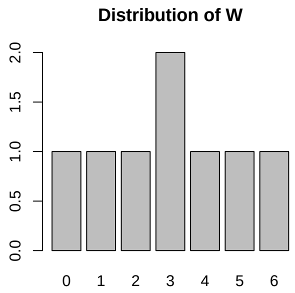
```
qsignrank(p = 0.05/2, n = 3, lower.tail = TRUE)
```

```
[1] 0
```

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

```
qsignrank(p = 0.05/2, n = 3, lower.tail = FALSE)
```

```
[1] 6
```



One-sample Wilcoxon signed rank test: Application

Critical values

If **one-tailed**

low critical value$_\alpha$:

```
qsignrank(p = 0.05, n = n, lower.tail = TRUE)
```

or

high critical value$_\alpha$:

```
qsignrank(p = 0.05, n = n, lower.tail = FALSE)
```

One-sample Wilcoxon signed rank test: Application

Draw conclusions

We reject the H_0 if:

- $W >$ high critical value$_{\alpha/2}$ or $W <$ low critical value$_{\alpha/2}$

We have $5 < 6$ and $1 > 0 \Rightarrow$ we do *not* reject the H_0

One-sample Wilcoxon signed rank test: Application

Draw conclusions

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

$p - value = 2 * Pr(W <= 1):$

```
2 * psignrank(q = 1, n = 3, lower.tail = TRUE)
```

```
[1] 0.5
```

or

$p - value = 2 * Pr(W >= 5) = 2 * (1 - Pr(W < 5)):$

```
2 * (1 - psignrank(q = 5 - 1, n = 3, lower.tail = TRUE))
```

```
[1] 0.5
```

```
2 * psignrank(q = 5 - 1, n = 3, lower.tail = FALSE)
```

```
[1] 0.5
```

One-sample Wilcoxon signed rank test: Application

Draw conclusions

If **one-tailed**

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

`psignrank(q =` W`, n = n, lower.tail = TRUE)`

or

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

`1 - psignrank(q =` W `- 1, n = n, lower.tail = TRUE)`

or

`psignrank(q =` W `- 1, n = n, lower.tail = FALSE)`