Hypothesis testing: critical values

Data Analysis for Psychology in R 1

Semester 2, Week 3

Dr Umberto Noè

Department of Psychology
The University of Edinburgh

Learning objectives

1. Understand the parallel between p-values and critical values
2. Be able to perform a one-sided or two-sided hypothesis test using the critical value method
3. Understand the link between t-scores and critical values

Part A

Introduction

Setting

- We cannot afford to collect data for the full population

- Data are only collected on **one** random sample of n individuals, where n = sample size

- After we have selected a sample at random, we know the measurements of the individuals in the sample.

- We are not interested in the individuals in the sample per se, but we collected data on them to **infer** from the sample data some property of the wider population the sample came from.

- You may want to:
    - Estimate a population parameter
    - Test whether a hypothesised parameter value is plausible

Estimation

If our goal is estimating a population mean, μ

- we use the average of the observations in the sample, $\bar{x}$, as the estimate
- the precision of our estimate is measured by the standard error, telling the average distance of a sample mean from the population mean
- a 95% (or 90% or 99%) **confidence interval** gives us a range of plausible values for the population mean. This is:

$$\left[\bar{x} - t^* \cdot \frac{s}{\sqrt{n}}, \ \ \bar{x} + t^* \cdot \frac{s}{\sqrt{n}}\right]$$

- for a 95% CI, the values $-t^*$ and $+t^*$ are found as:

```
qt(c(0.025, 0.975), df = n - 1)
```

Testing

If our goal is testing a hypothesis, for example:

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0$$

- Compute a **test statistic**, measuring some sort of "distance" between the sample data and the null hypothesis.

> **Definition: Test Statistic**
> A test statistic is any numerical quantity computed from the sample data with the purpose to make a test of some kind.

- For testing a population mean, we use the **t-statistic**:

$$t = \frac{\bar{x} - \mu_0}{SE} \qquad \text{where} \qquad SE = \frac{s}{\sqrt{n}}$$

- The t-statistic is the distance of the sample mean from the hypothesised parameter value, measured in units of the standard error.

- When you will perform a test on categorical variables you will see a different type of test statistic (the chi-squared statistic).

Part B

P-values and Critical Values

Sample mean

- Why is the sample mean a **good estimate** of the population mean?
- To study this, let's do a thought experiment that in practice we seldom can do:
    - **IF you could afford** to take not just one sample, but many samples from the population, each of size n.
    - You could compute the average of the data in each sample
    - You can plot a histogram of those averages
    - The centre of the histogram is the same value as the population mean
    - The spread of the histogram is the standard error

Thought experiment: Sampling distribution

- Suppose I gave you a population **with a mean of 20**:

  (21, 21, 18, 23, 21, 25, 16, 19, 17, 19, 21, 23, 19, 18, 19, 21, 20, 23, 19, 17)

- Take all possible samples of size $n = 4$.

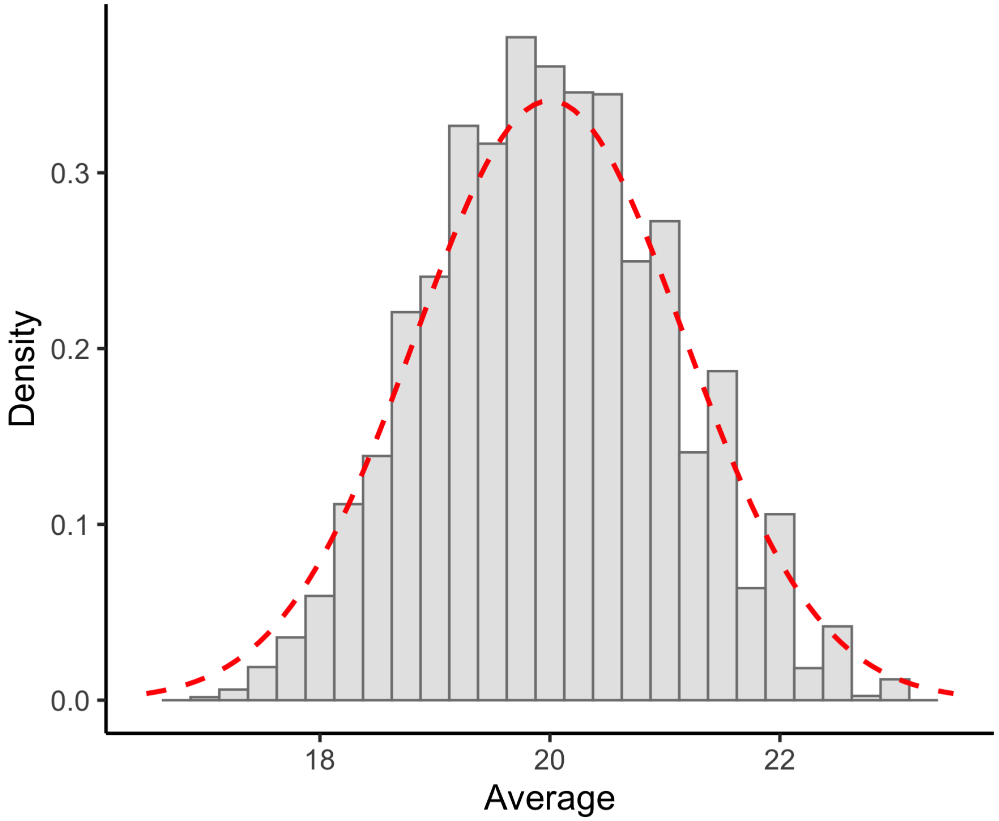
- **For each sample:**

  - Compute the average of the $n = 4$ numbers in the sample.

- Plot all the averages $\bar{x}$'s using a histogram.

  - Centre?
  - Spread?

- The sample mean $\bar{x}$ fluctuates from sample to sample around the population mean $\mu = 20$, and the typical distance from the true value is given by the SE



Testing hypotheses

- Suppose we are testing

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- We can build a **test statistic** to assess how much the sample data are consistent with the null hypothesis we specified.

- This takes the form of a distance between the observed and hypothesised mean, measured in units of the SE

- The **test statistic** for testing a mean is the **t-statistic** or **t-score**. In our case, $\mu_0 = 20$ so:

$$t = \frac{\bar{x} - 20}{s/\sqrt{n}}$$

- We can compute the t-statistic for the observed sample. But is this an surprising value or not?

- To decide this we need to ask ourselves: what is the distribution of the t-statistic **when H_0 is true**? In other words, what is the **null distribution**?

Thought experiment: Null distribution

- Suppose I gave you a population **with a mean of 20**:

  (21, 21, 18, 23, 21, 25, 16, 19, 17, 19, 21, 23, 19, 18, 19, 21, 20, 23, 19, 17)

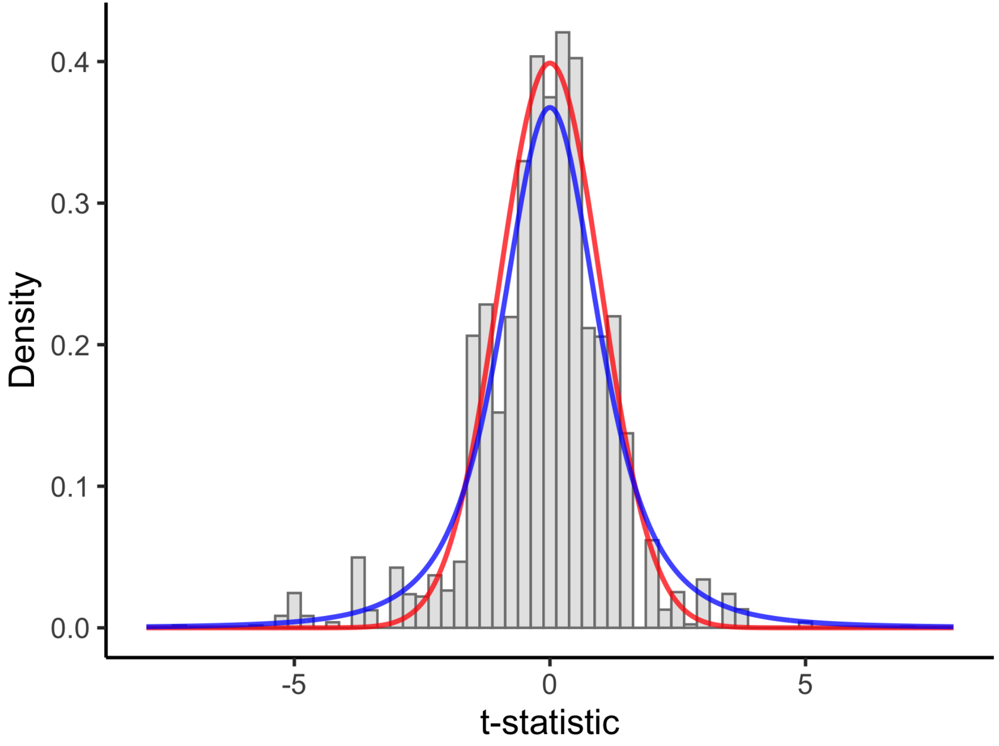
- Take all possible samples of size $n = 4$.

- **For each sample**:

  - Compute the average $\bar{x}$ of the $n = 4$ numbers in the sample.
  - Compute the SD s of the $n = 4$ numbers in the sample.
  - Compute the t-statistic $t = \frac{\bar{x}-20}{s/\sqrt{n}}$ for that sample.

- Histogram of all t-statistics shows a distribution with more variability than a standard normal:

$$t(n-1)$$



Red: Standard Normal
Blue: t(n-1)

Null distribution

- This thought-experiment shows us that the t-statistic, when the null hypothesis is true, follows a $t(n-1)$ distribution.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$$

- Why only when H_0 is true? Recall the previous example, in which the null hypothesis that $H_0 : \mu = 20$ was true.
- If that is the case, the sample means will fluctuate around 20. In turn, the distances of the sample means from 20, i.e. the t-scores, will fluctuate around 0.
- The null distribution shows us all the possible distances (t-statistics) between a sample mean and the hypothesised mean, when H_0 is true.
- If our observed sample gives us a t-statistic that is unlikely / surprising to obtain when H_0 is true, we start doubting the null hypothesis!

Example

- Suppose you have collected data on one sample, with sample size 4. The sample data are:

$$(32, 36, 26, 28)$$

- We wish to test whether this sample comes from a population with a mean different from 20:

$$H_0 : \mu = 20 \qquad \text{vs} \qquad H_1 : \mu \neq 20$$

```
data_sample <- tibble(x = c(32, 36, 26, 28))
data_sample
```

```
## # A tibble: 4 × 1
##       x
##   <dbl>
## 1    32
## 2    36
## 3    26
## 4    28
```

```
xbar <- mean(data_sample$x)
xbar
```

```
## [1] 30.5
```

```
n <- nrow(data_sample)
s <- sd(data_sample$x)
se <- s / sqrt(n)

mu0 <- 20
tobs <- (xbar - mu0) / se
tobs
```

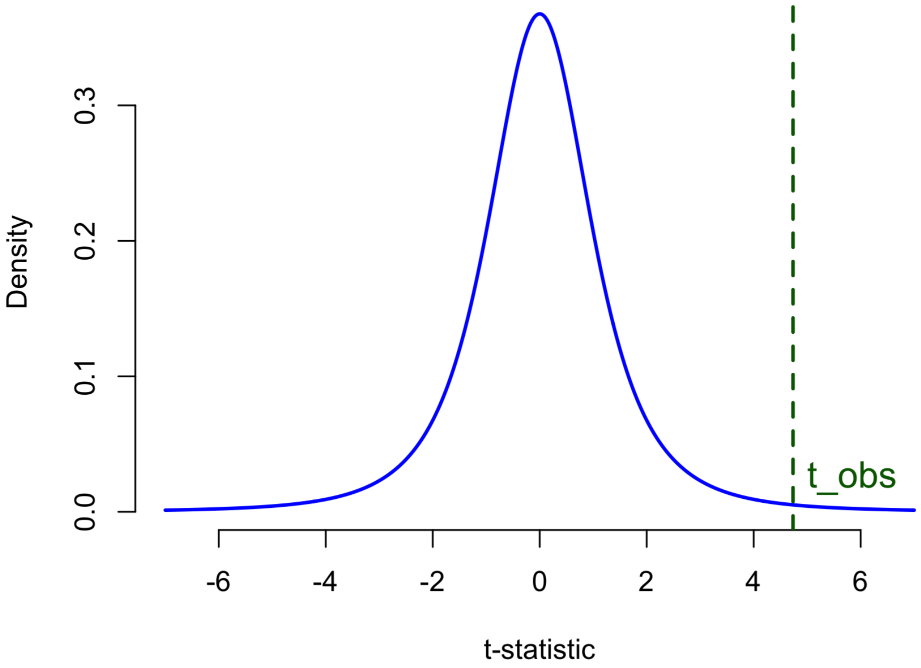
```
## [1] 4.735
```

P-value

- Last week we learned to assess significance by computing the p-value.
- We choose a significance level, $\alpha = 0.05$ say.
- As H_1 is two-sided, we compute the p-value as:

```
# Twice the area to the right of observed t
pvalue <- 2 * pt(abs(tobs), df = n-1,
                 lower.tail = FALSE)
pvalue
```

```
## [1] 0.01785
```



P-value

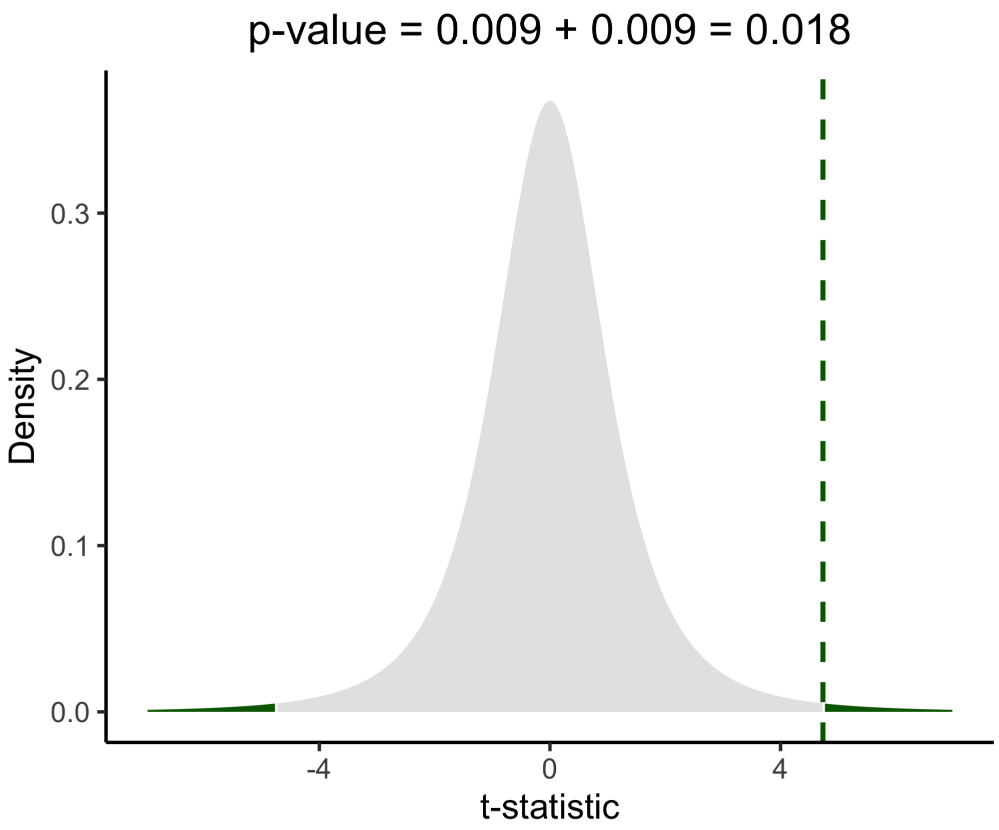
- As H_1 is two-sided, we compute the p-value as:

```
# Twice the area to the right of observed t
pvalue <- 2 * pt(abs(tobs), df = n-1,
                 lower.tail = FALSE)
pvalue
```

```
## [1] 0.01785
```

- This is the probability of observing a t-statistic having at least the same distance from 0 as the observed t-statistic, when H_0 is true.

- An observed mean of 30 is as distant from 20 as 10 is. So both would be equally "different" from the hypothesised value, 20.



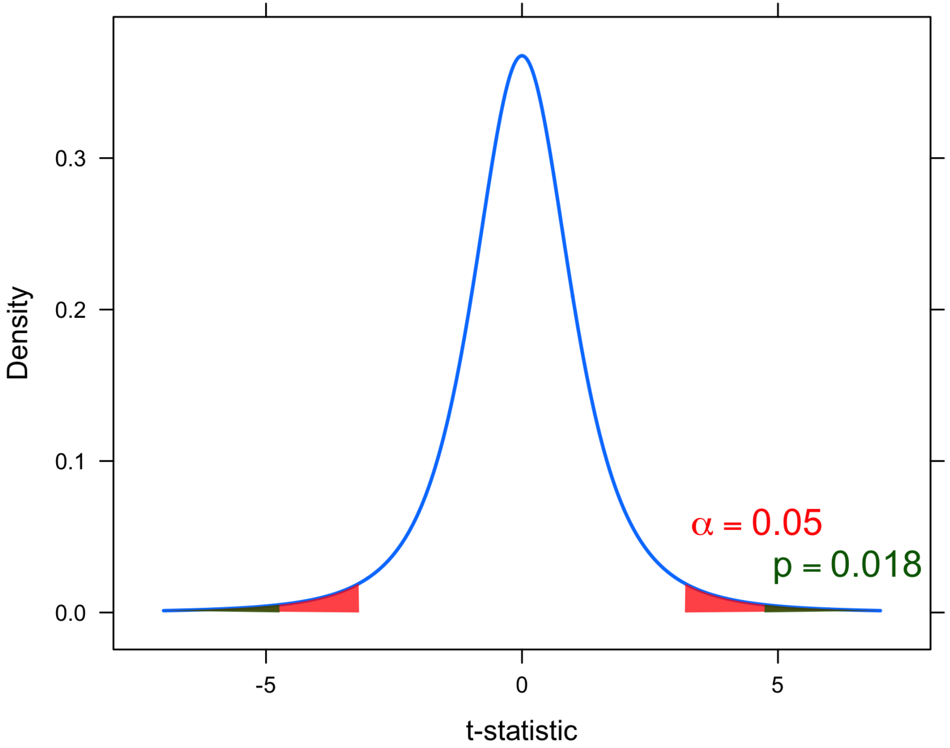
Making a decision

- To make a decision on whether or not to reject H_0 we need to compare the computed p-value with the chosen significance level of 5%.

- The p-value is 0.018, which is less than the chosen significance level, so we reject the null hypothesis.

- In doing so, we compared the *green* area, corresponding to the p-value, against the *red* area, corresponding to the $\alpha = 0.05$ significance level.

- Recalle that the $\alpha = 0.05$ probability is equally divided among the two tails in this case, because the alternative hypothesis is two-sided.



Equivalent approach!

- Rather than comparing the area of α (0.05, in red) to the area of the p-value (0.018, in green), we can compare the corresponding t-statistics along the x-axis.

- The p-value is computed using the observed t-statistic, 4.74.

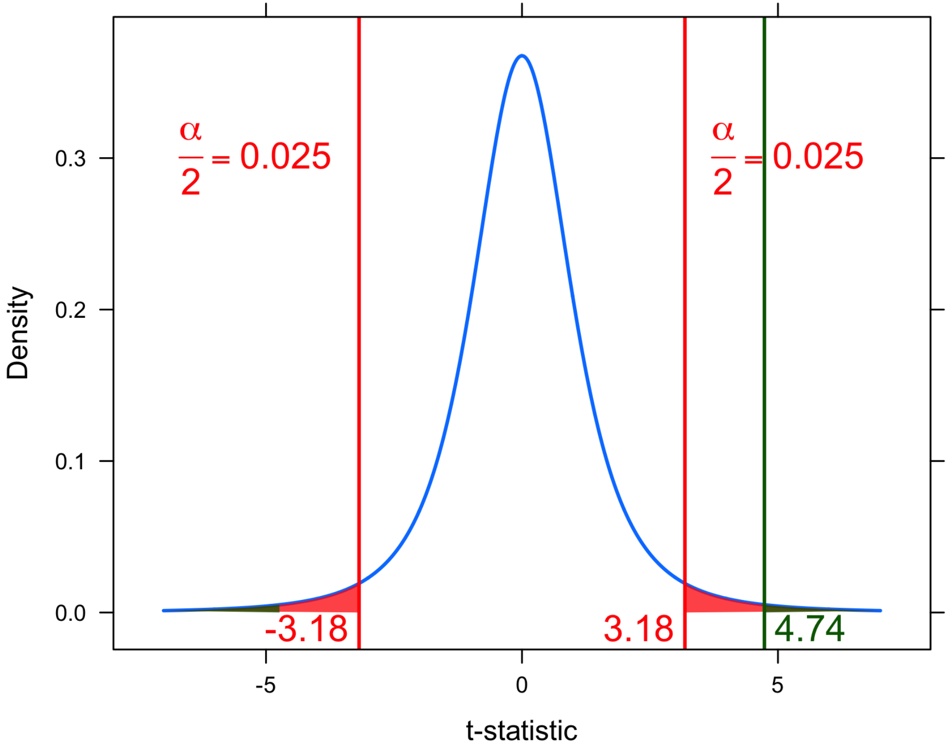
- The t values that cut an area of 0.025 to the left and 0.025 to the right are called the **critical values** for $\alpha = 0.05$ and denoted $-t^*$ and $+t^*$:

```
qt(c(0.025, 0.975), df = n-1)
```

```
## [1] -3.182  3.182
```

- We reject H_0 when either $t \leq -t^*$ or $t \geq +t^*$.

- In this case, $t = 4.74$ is larger than the upper critical value, $t^* = 3.18$.



Example 2

- Suppose now that the collected sample, with sample size 4, was:

$$(18, 21, 19, 23)$$

- We wish to test whether this sample comes from a population with a mean different from 20:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

```
data_sample2 <- tibble(x = c(18, 21, 19, 23))
data_sample2
```

```
## # A tibble: 4 × 1
##       x
##   <dbl>
## 1    18
## 2    21
## 3    19
## 4    23
```

Example 2

```
xbar <- mean(data_sample2$x)
xbar
```

```
## [1] 20.25
```

```
n <- nrow(data_sample2)
s <- sd(data_sample2$x)
se <- s / sqrt(n)

mu0 <- 20
tobs <- (xbar - mu0) / se
tobs
```

```
## [1] 0.2255
```

Example 2

- Compute the critical values for a $t(n-1)$ distribution with $\alpha = 0.05$.

```
tstar <- qt(c(0.025, 0.975), df = n-1)
tstar
```

```
## [1] -3.182  3.182
```

- Is the observed t-statistic $t =$ 0.23 smaller than or equal to the lower critical value? *No!*

```
tobs <= tstar[1]
```

```
## [1] FALSE
```

- Is the observed t-statistic $t =$ 0.23 greater than or equal to the upper critical value? *No!*

```
tobs >= tstar[2]
```

```
## [1] FALSE
```

Example 2

- As our observed t-statistic lies in between the two critical values, rather than beyond, it lies in the middle 95% of the null distribution.
- If you were to compute the p-value for t, it would be larger than the area arising from the critical values $\pm t^*$ (the significance level α).
- We do not have sufficient evidence to reject H_0 at the 5% significance level.

Part C

Body temperature example

Body temperature example

> Has the average body temperature for healthy humans changed from the long-thought 37 °C?

- We are testing:

$$H_0 : \mu = 37 \qquad \text{vs} \qquad H_1 : \mu \neq 37$$

- Read the data:

```
library(tidyverse)
tempsample <- read_csv('https://uoepsy.github.io/data/BodyTemp.csv')
head(tempsample)
```

```
## # A tibble: 6 × 2
##   BodyTemp Pulse
##      <dbl> <dbl>
## 1     36.4    69
## 2     37.4    77
## 3     37.2    75
## 4     37.1    84
## 5     36.7    71
## 6     37.2    76
```

Body temperature example

```r
xbar <- mean(tempsample$BodyTemp)
xbar
```

```
## [1] 36.81
```

- The observed t-statistic: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

```r
n <- nrow(tempsample)
n
```

```
## [1] 50
```

```r
s <- sd(tempsample$BodyTemp)
SE <- s / sqrt(n)

mu0 <- 37
tobs <- (xbar - mu0) / SE
tobs
```

```
## [1] -3.141
```

Body temperature example

- The observed t-statistic is $t =$ -3.14.
- Compute the critical values of a t(49) distribution with $\alpha = 0.05$:

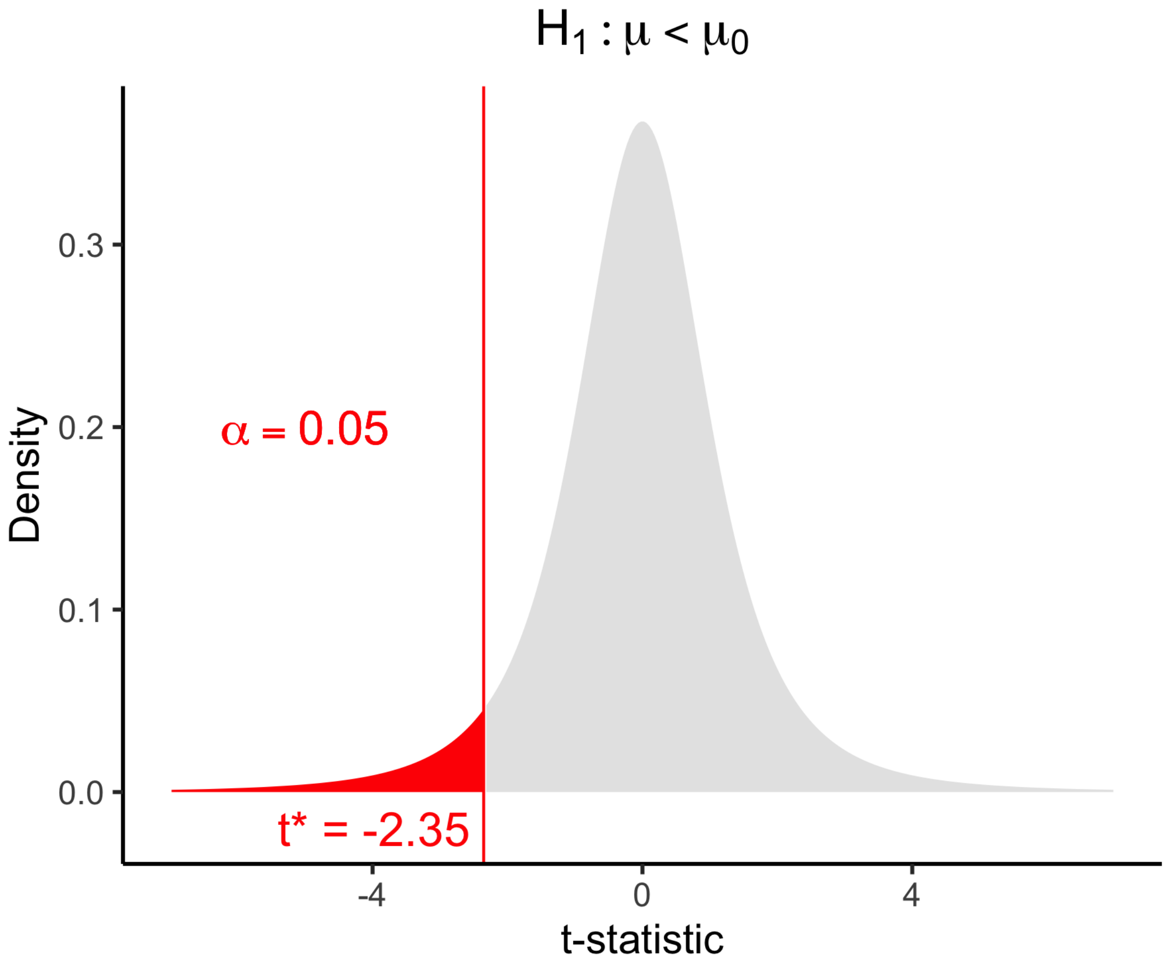
```
qt(c(0.025, 0.975), df = n - 1)
```

```
## [1] -2.01  2.01
```

- The observed t-statistic lies beyond the critical values, and as such falls in the 5% probability tails of the null distribution.
- If you were to compute the p-value, it would be smaller than 0.05.
- We reject the null hypothesis as the observed t-statistic is unlikely to be obtained if the null hypothesis were true.
- In terms of reporting, when the observed t is beyond the critical values, $p < \alpha$. Otherwise, $p > \alpha$.

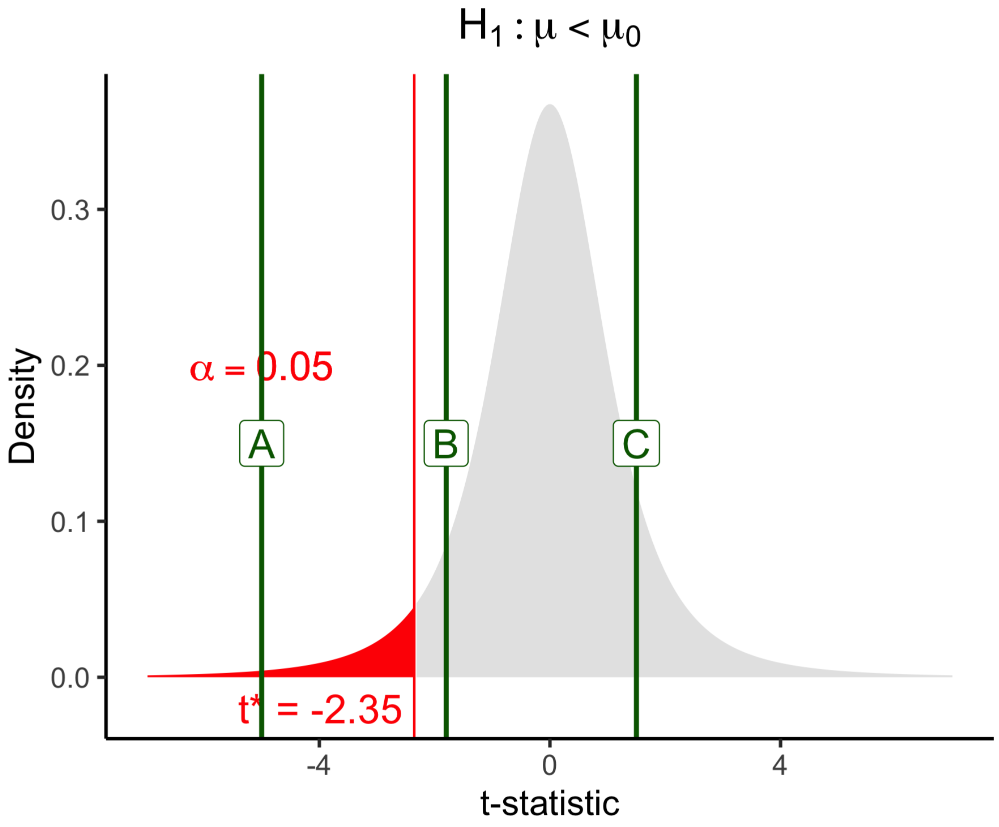
> At the 5% significance level, we performed a two-sided hypothesis test against the null hypothesis that the mean body temperature for all healthy humans is equal to 37 °C.
> As the observed t-statistics lies beyond the critical values, the sample results provide strong evidence against the null hypothesis and in favour of the alternative one that the average body temperature differs from 37 °C;
> $t(49) = 3.14, p < .05$, two-sided.

$H_1 : \mu < \mu_0$, example with $t(3)$

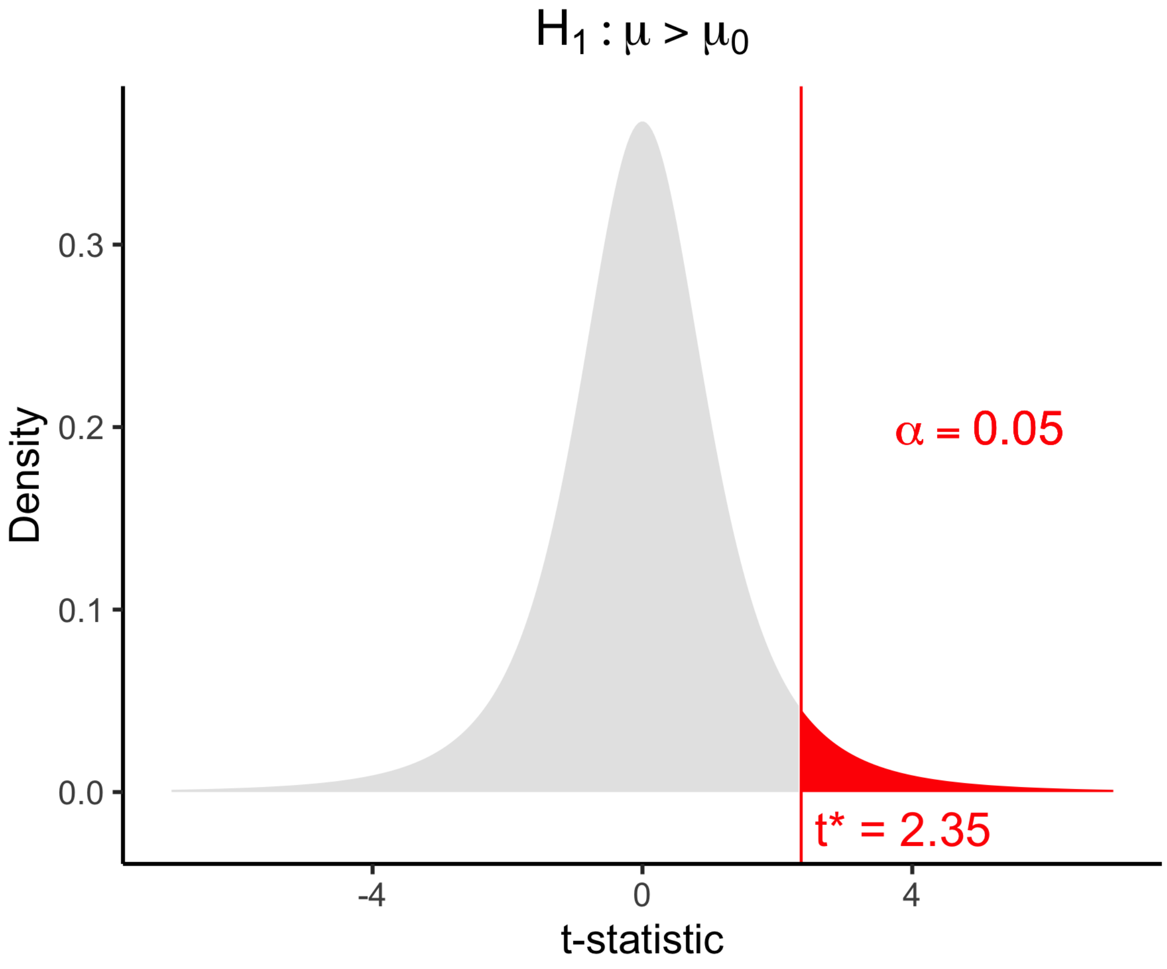


$H_1 : \mu < \mu_0$, example with $t(3)$

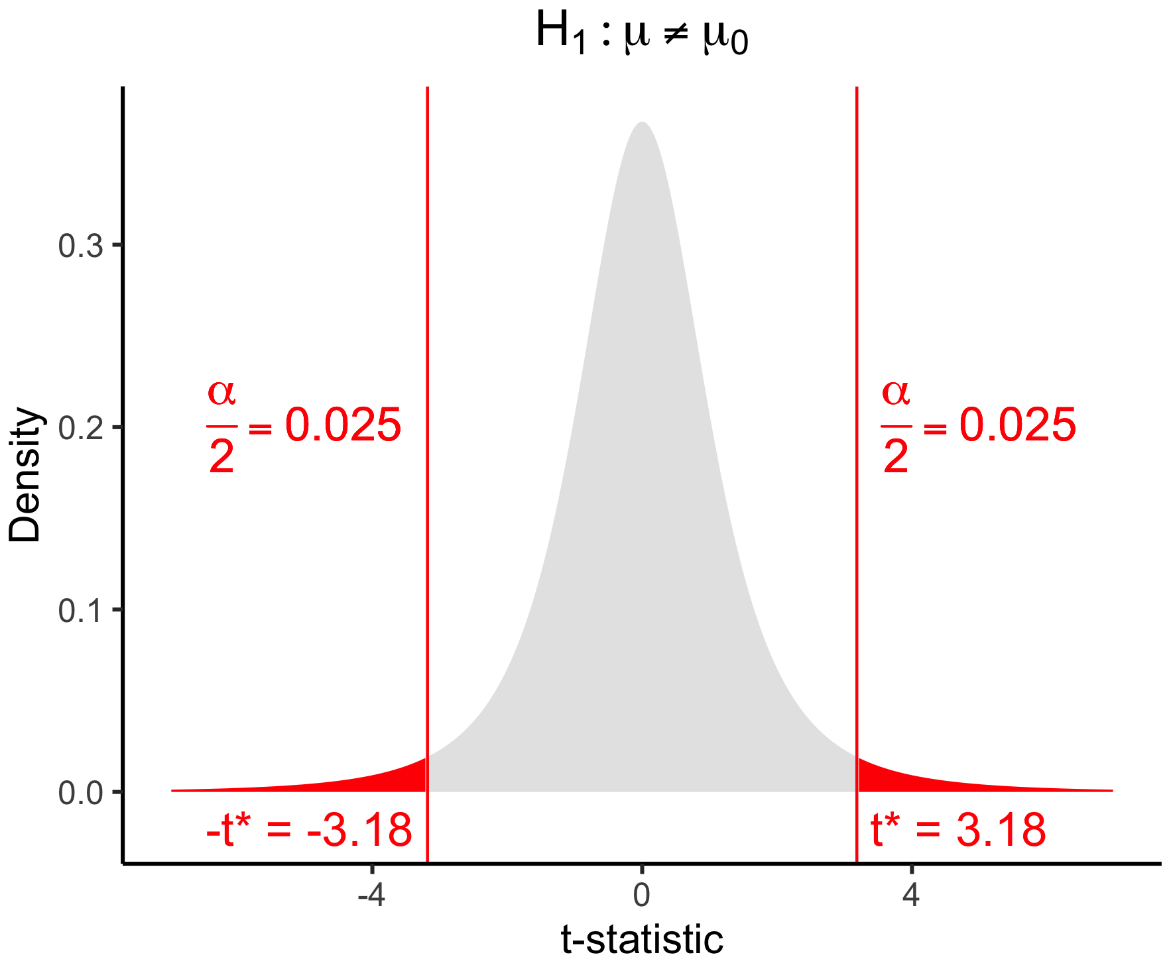
- t = A will lead to a p-value < 0.05
- t = B will lead to a p-value > 0.05
- t = C will lead to a p-value > 0.05



$H_1 : \mu > \mu_0$, example with $t(3)$



$H_1 : \mu \neq \mu_0$, example with $t(3)$


$H_1 : \mu \neq \mu_0$
$\frac{\alpha}{2} = 0.025$
$\frac{\alpha}{2} = 0.025$
-t* = -3.18
t* = 3.18
Density
t-statistic