DATA2002

Critical values, rejection regions and confidence intervals

Garth Tarr

THE UNIVERSITY OF SYDNEY

Random variables review

Critical values

Confidence intervals

Rejection regions

Random variables

Random variable basics

- A random variable can be thought of as a mathematical object which takes certain values with certain probabilities.
- We have *discrete* and *continuous* random variables, although we can always "approximate" a continuous one with a discrete one (taking values on a suitably fine grid).
- A simple discrete random variable X can be described as a *single random* draw from a "box" containing tickets, each with numbers written on them.
- In this case,
- $\mathrm{E}(X) = \mu$ (the average of the numbers in the box);
- $\mathrm{Var}(X) = \sigma^2$ (the *population variance* of the numbers in the box);
- $\mathrm{SD}(X) = \sigma$.

Random sample with replacement

- Next, consider taking a random sample of size n *with replacement*, denote the values $X_1, X_2, \ldots, X_n$.
- This means, one of *all possible samples of size* n is chosen in such a way that each is equally likely.
- If there are N tickets in the box, how many such samples are there?
- It turns out that these X_i's are *independent and identically distributed*. This means
  - each X_i has the same distribution as a single draw;
  - the X_i's are all mutually independent.
- Consider now taking the total $T = \sum_{i=1}^{n} X_i$.
- What is $\mathrm{E}(T)$?
- What is $\mathrm{Var}(T)$?

Expectation and variance of sums

The **expectation** of a sum is *always* the sum of the expectations. For example,

$$\mathrm{E}(T) = \mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = \underbrace{\mu + \cdots + \mu}_{n \text{ terms}} = n\mu\,.$$

Variance of sum of independent random variables

- The variance of a sum is *not always* the sum of the variances.
- However, it *is* if the X_i's are *independent*. So,

$$\mathrm{Var}(T) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = \underbrace{\sigma^2 + \cdots + \sigma^2}_{n \text{ terms}} = n\sigma^2\,.$$

Multiplying by a constant: for any random variable X and any constant c,

$$\mathrm{E}(cX) = c\,\mathrm{E}(X) \quad \text{and} \quad \mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)\,.$$

Sample mean

Consider the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}T$. What is $\mathrm{E}(\bar{X})$? What is $\mathrm{Var}(\bar{X})$?

- Thus since $\bar{X} = \frac{1}{n}T$,

$$\mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{1}{n}T\right) = \frac{1}{n}\mathrm{E}(T) = \frac{1}{n}n\mu = \mu\,.$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}T\right) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(T) = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n}\,.$$

Estimating μ

- In many applications, we model data $x_1, \ldots, x_n$ as values taken by such a sample $X_1, \ldots, X_n$ and we are interested in "estimating" or "learning" μ (which is an "unknown population mean").
- In this case the *estimator* is the sample mean $\bar{X}$ (regarded as a *random variable*).
- The *estimate* is $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, the observed value of the mean of the data (this is *conceptually different* to $\bar{X}$!!)
- An important theoretical quantity is the *standard error*, the *standard deviation of the estimator*.

$$\mathrm{SE} = \mathrm{SD}(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}\,.$$

this is (in general) *also an unknown parameter*.

Importance of the standard error

- The standard error (standard deviation of the estimator) is important to know, since it tells us the "likely size of the estimation error".
- An estimate on its own is not very useful, we need to also know how accurate or reliable the estimate is.
    - This is what the standard error provides.
- *Unfortunately* in most contexts the standard error *is also unknown*;
    - but we can usually (also) estimate the standard error!

Estimating the standard error

- The standard error (at least when estimating a population mean μ) involves the (usually unknown) population variance σ^2:

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}\,.$$

- Fortunately, we can usually estimate σ^2 using the *sample variance*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\,.$$

- The corresponding *estimated standard error* is

$$\widehat{\mathrm{SE}} = s/\sqrt{n}\,,$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the observed value of the sample variance.

Critical values and confidence intervals

More precise inference

- Usually, we want to know if a given value μ_0 is a "plausible value" for the unknown μ, based on observed data $x_1, \ldots, x_n$.
- Roughly speaking, we do this by
  1. computing the value of the *estimate* $\bar{x}$;
  2. computing the value of the *estimated standard error* $s/\sqrt{n}$;
  3. seeing if the *discrepancy* $\bar{x} - \mu_0$ is "large" compared to the standard error.
- The various procedures we look at:
  - t-tests (with corresponding p-values)
  - confidence intervals
  - rejection regions

  are all variations on this single idea.

What kind of discrepancies are of interest?

- We need to have it very clear in our minds which kind of discrepancies $\bar{x} - \mu_0$ we are interested in:
  - positive
  - negative
  - both
- Another way to think about it is, given a fixed μ_0 of interest and an observed sample mean $\bar{x}$, which of the following questions are we asking:
  1. Is $\bar{x}$ significantly *more* than μ_0? (*one-sided*)
  2. Is $\bar{x}$ significantly *less* than μ_0? (*one-sided*)
  3. Is $\bar{x}$ significantly *different* to μ_0? (*two-sided*)

Beer contents

Beer contents in a pack of six bottles (in millilitres) are:

374.8, 375.0, 375.3, 374.8, 374.4, 374.9

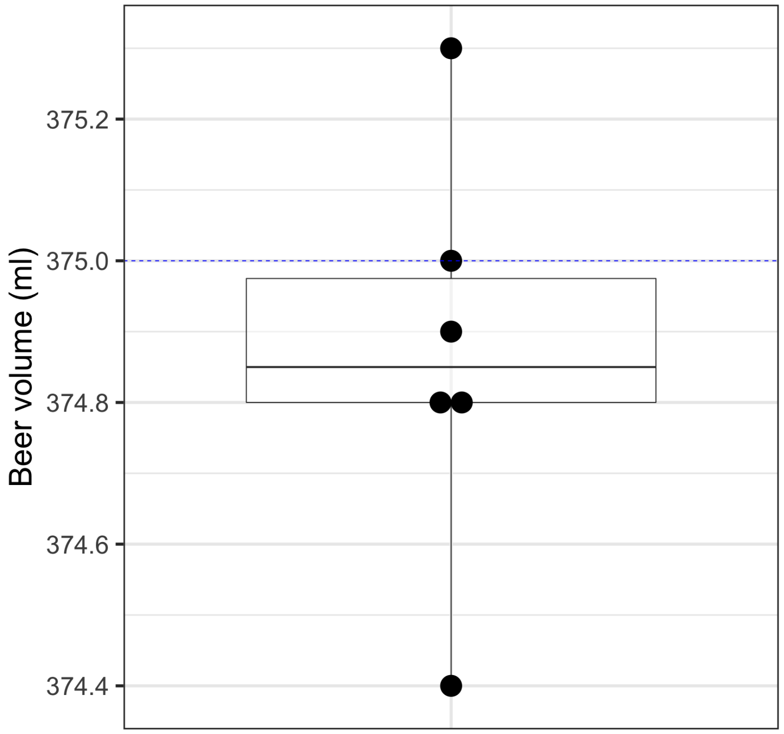
Does the mean beer content differ from the 375 mL claimed on the label?

```
x = c(374.8, 375.0, 375.3, 374.8, 374.4, 374.9)
mean(x)
```

```
## [1] 374.8667
```

```
sd(x)
```

```
## [1] 0.294392
```



Beer example

- In the beer example there are different possible points of view.
- For *consumers*, the results will only be "interesting" if $\bar{x}$ is significantly *less* than 375:
  - in this case the company is "ripping consumers off".
- However for the *beer producers*, both positive and negative discrepancies might be of interest:
  - if they are *underfilling*, consumers will be unhappy;
  - if they are *overfilling*, they are "wasting" some of their product.
- Thus both a *one-sided* and *two-sided* point of view are conceivable even for this example.

Two-sided discrepancies of interest

- When two-sided discrepancies are of interest we are basically asking: for a given μ_0, is the *absolute value* $|\bar{x} - \mu_0|$ large, compared to the standard error $s/\sqrt{n}$?

t-test approach: declare μ_0 not plausible if $|\bar{x} - \mu_0| > c\frac{s}{\sqrt{n}}$ for some "suitably chosen" constant c.

Confidence interval approach: the set of plausible values for the unknown μ is

$$\bar{x} \pm c\frac{s}{\sqrt{n}},$$

for some "suitably chosen" constant c.

- Note that if the *same* c is chosen in both approaches, the set of plausible values *is the same*:
  - μ_0 in the confidence interval $\Leftrightarrow |\bar{x} - \mu_0| \leq cs/\sqrt{n}$.

How to choose the constant c?

- The constant c can be chosen in a sensible way in each context.
- **Testing**: control the **false alarm rate**.
- **Confidence intervals**: control the **coverage probability**;
  - the coverage probability is commonly also called the confidence level and expressed as a percentage.

False alarm rate

- A "**false alarm**" is when we "reject incorrectly".
- Using our current language it is when we "reject a given value μ_0" when we shouldn't.
- That is, we declare μ_0 "not plausible" when it is in fact the true value!
- We pick choose small $0 \le \alpha \le 1$ for the desired "**false alarm rate**" e.g. 0.05, 0.01.
- Choose c such that (if possible)

$$P\left(|\bar{X} - \mu_0| > c\frac{S}{\sqrt{n}}\right) = \alpha\,;$$

- If this is not possible then just try to ensure that this probability does not *exceed* α!
- The **false alarm rate** is also called the **significance level**.

Normal population: use the t-distribution

- Under the special statistical model where the data are modelled as values taken by iid normal random variables, we know that *if the true population mean is indeed* $\boldsymbol{\mu_0}$, then the ratio

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

  and we can thus choose c such that

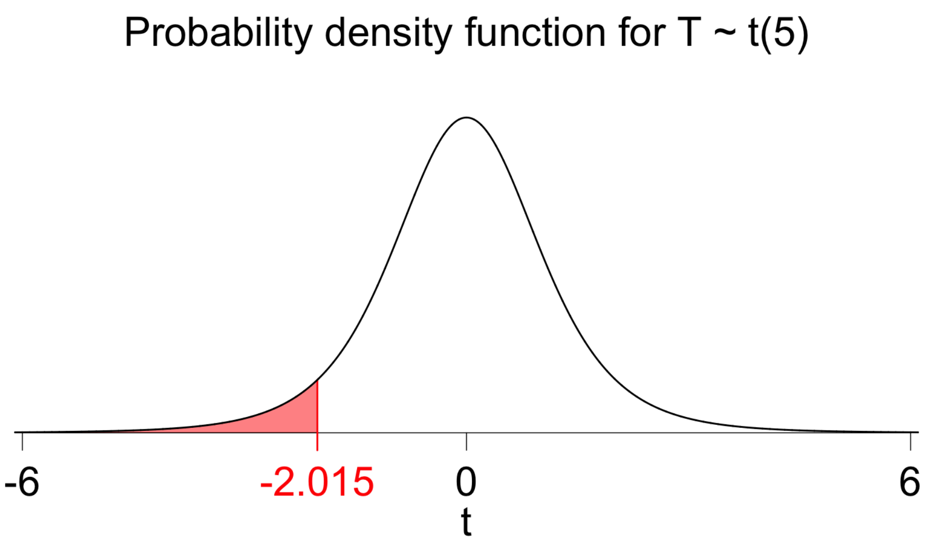
$$P\left(|\bar{X} - \mu_0| > c\frac{S}{\sqrt{n}}\right) = P\left(\frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} > c\right) = P(|t_{n-1}| > c) = \alpha\,. \qquad (*)$$

Finding quantiles in R

In R, we get quantiles using the `qDISTRIBUTION()` range of functions, e.g. `qt(p, n - 1)`, `qnorm(p)`, `qchisq(p, n - 1)` for t, normal and χ^2 distributions respectively.

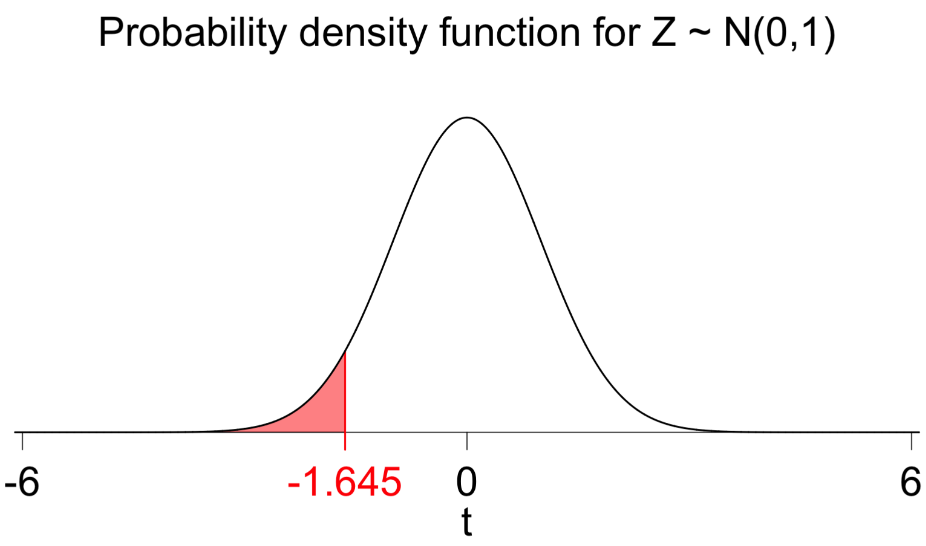
```r
qt(0.05, 5)
```

```
## [1] -2.015048
```




```r
qnorm(0.05)
```

```
## [1] -1.644854
```



Using `qt()`

- Note that if $P(|t_{n-1}| > c) = \alpha$ then

$$P(|t_{n-1}| \leq c) = P(-c \leq t_{n-1} \leq c) = 1 - \alpha$$

  and furthermore

$$P(t_{n-1} < -c) + P(t_{n-1} > c) = 2P(t_{n-1} > c) = \alpha$$

  so

$$P(t_{n-1} > c) = \frac{\alpha}{2} \quad \text{or equivalently} \quad P(t_{n-1} \leq c) = 1 - \frac{\alpha}{2}$$

- So for e.g.
  - $\alpha = 0.05$, we need c such that $P(t_{n-1} \leq c) = 1 - 0.025 = 0.975$ use `c = qt(0.975, df = n-1)`
  - $\alpha = 0.01$, we need c such that $P(t_{n-1} \leq c) = 1 - 0.005 = 0.995$ use `c = qt(0.975, df = n-1)`.

Beer example

- Recall we have observations

```
x = c(374.8, 375.0, 375.3, 374.8, 374.4, 374.9)
```

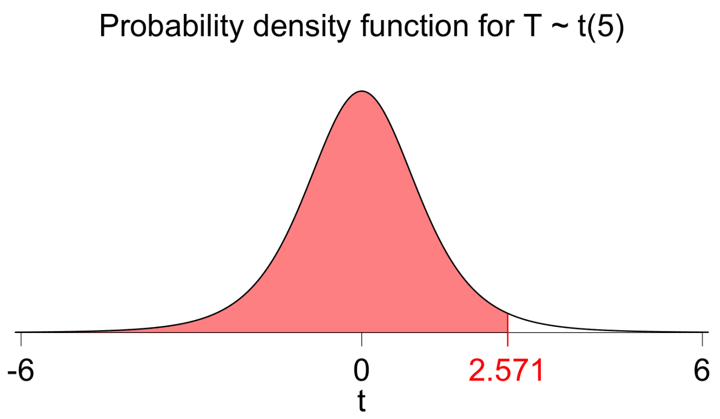
- Here the sample size $n = 6$ so if
  - $\alpha = 0.05$ we need c such that $P(t_5 \leq c) = 0.975$;
  - $\alpha = 0.01$ we need c such that $P(t_5 \leq c) = 0.995$.
- These are given by

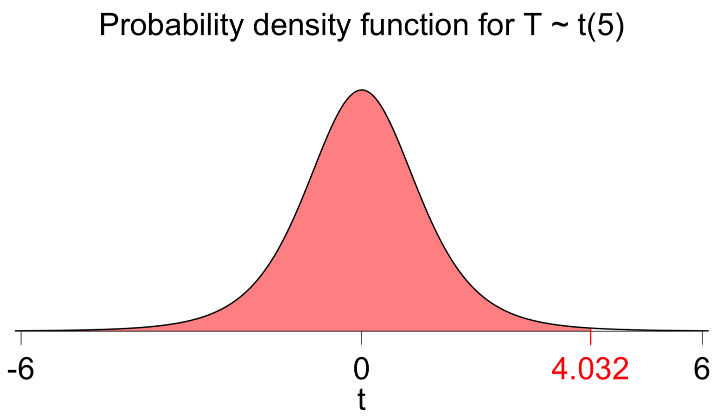
```
qt(0.975,5)
```

```
## [1] 2.570582
```

```
qt(0.995,5)
```

```
## [1] 4.032143
```





Beer example

- The sample mean is

```
xbar = mean(x)
xbar
```

```
## [1] 374.8667
```

- The standard error is

```
se = sd(x)/sqrt(6)
se
```

```
## [1] 0.120185
```

- The discrepancy from the "given value" 375 is

```
discrep=abs(xbar-375)
discrep
```

```
## [1] 0.1333333
```

- This is only slightly more than 1 (estimated) standard error.
- We need it to be at least 2.57 standard errors to "reject at the 0.05 **false alarm rate**":
- Therefore we cannot reject H_0, so 375 is a plausible value (in this two-sided sense).

Coverage probability

- For a **confidence interval**, the **coverage probability** is simply the probability that the "true" value of the unknown parameter lies inside (is "covered by") the **confidence interval**.
- This is a *long run property* and should be interpreted in the context of *repeated experiments*.
- We choose a (small) *non-coverage probability* α, say 0.05 or 0.01;
    - then the **coverage probability** is $1-\alpha$.
- Thus, under some statistical model we choose c so that the **coverage probability** *under the model* satisfies (with μ the *true population mean*):

$$P\left(\bar{X} - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + c\frac{S}{\sqrt{n}}\right) = P\left(|\bar{X} - \mu| \leq c\frac{S}{\sqrt{n}}\right) = 1 - \alpha\,.$$

Equivalent to false alarm rate condition for t-test

- The **coverage probability** condition on the previous slide is an equivalent statement to the **false alarm rate** condition for the t-test (for the same α).
- Thus if the desired **coverage probability** is
  - 0.95 (i.e. non-coverage probability $\alpha = 0.05$) then we need c such that

$$P(t_{n-1} \leq c) = 1 - 0.025 = 0.975\,;$$

  - 0.99 (i.e. non-coverage probability $\alpha = 0.01$) then we need c such that

$$P(t_{n-1} \leq c) = 1 - 0.005 = 0.995\,.$$

Beer example

- For a 95% **confidence interval** for μ we thus choose c via

```
c_95 = qt(0.975,5)
c_95
```

```
## [1] 2.570582
```

giving

```
xbar + c(-1,1) * c_95 * se
```

```
## [1] 374.5577 375.1756
```

- Note that this includes the "special value" 375 and so is consistent with our 0.05 **false-alarm rate** test earlier.

- For a 99% **confidence interval** for μ we thus choose c via

```
c_99 = qt(0.995,5)
c_99
```

```
## [1] 4.032143
```

giving

```
xbar + c(-1,1)*c_99*se
```

```
## [1] 374.3821 375.3513
```

- As we'd expect, this CI is wider, and also includes 375.

Using `t.test()`

- Compare our "manual" computations above with the output of the R function `t.test()`:

- First the default:

```
t.test(x, mu = 375)
```

```
##
##      One Sample t-test
##
## data:  x
## t = -1.1094, df = 5, p-value = 0.3177
## alternative hypothesis: true mean is not equal to 375
## 95 percent confidence interval:
##  374.5577 375.1756
## sample estimates:
## mean of x
##  374.8667
```

- Setting `conf.level=0.99`:

```
t.test(x, mu = 375, conf.level = 0.99)
```

```
##
##      One Sample t-test
##
## data:  x
## t = -1.1094, df = 5, p-value = 0.3177
## alternative hypothesis: true mean is not equal to
## 99 percent confidence interval:
##  374.3821 375.3513
## sample estimates:
## mean of x
##  374.8667
```

- Note the default in R is *two-sided*.

One-sided discrepancies of interest

- The "two-sided" approach just outlined would be of interest to the beer producers, but not necessarily the beer consumers.
- Let us consider the point of view of the consumers now.
- t-test approach: declare
  - μ_0 not plausible if $\bar{x} - \mu_0 < -c\frac{s}{\sqrt{n}} \Leftrightarrow \bar{x} < \mu_0 - c\frac{s}{\sqrt{n}}$ for some "suitably chosen" constant c.
- **Confidence interval** approach: set of plausible values for the unknown μ are those "not too much bigger than $\bar{x}$", i.e.

$$\Big(-\infty, \bar{x} + c\frac{s}{\sqrt{n}}\Big]$$

  for a "suitably chosen" constant c.
  - the upper endpoint is sometimes called an "upper confidence limit"
  - it can be interpreted as "the largest value consistent with the data".

Same set of plausible values

- Again, note that for the same c these two approaches give the *same set of plausible values* for μ:
  - μ_0 is in the (one-sided) **confidence interval** $\Leftrightarrow \bar{x} \geq \mu_0 - c\frac{s}{\sqrt{n}}$.

Controlling the (one-sided) false alarm rate

- We use a similar approach to the two-sided case, but with a crucial difference!
- We again know that under the iid normal model with population mean μ, $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$.
- We thus choose c so that if μ_0 is the true value,

$$P\left(\bar{X} < \mu_0 - c\frac{S}{\sqrt{n}}\right) = P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -c\right) = P(t_{n-1} < -c) = \alpha\,.$$

- By symmetry we must also have

$$P(t_{n-1} > c) = \alpha \quad \text{or} \quad P(t_{n-1} \leq c) = 1 - \alpha\,.$$

- Thus for **false alarm rate**
  - 0.05 we need c such that $P(t_{n-1} \leq c) = 1 - 0.05 = 0.95$;
  - 0.01 we need c such that $P(t_{n-1} \leq c) = 1 - 0.01 = 0.99$.

Beer example

For the $\alpha = 0.05$ **false alarm rate**, since $n = 6$ we need

```
c_05 = qt(.95, 5)
c_05
```

```
## [1] 2.015048
```

Note this is *smaller* than the two-sided version.

- We have already seen that the discrepancy is only slightly more than 1 standard error:

```
c(xbar - 375, se)
```

```
## [1] -0.1333333  0.1201850
```

so in this one-sided sense, 375 is a plausible value.

For the $\alpha = 0.01$ **false alarm rate**, since $n = 6$ we need

```
c_01 = qt(.99, 5)
c_01
```

```
## [1] 3.36493
```

- Note that this is also smaller than the two-sided version.
- This makes the one-sided tests "more sensitive" than the two-sided versions.

One-sided confidence intervals

- Again we fix the **coverage probability** $1 - \alpha$:

$$P\left(\mu_0 \leq \bar{X} + c\frac{S}{\sqrt{n}}\right) = P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq -c\right) = P(t_{n-1} \geq -c) = P(t_{n-1} \leq +c) = 1 - \alpha\,.$$

  which is again the same as the corresponding **false alarm rate** condition.

- Thus for non-coverage probability
  - 0.05 we need c such that $P(t_{n-1} \leq c) = 1 - 0.05 = 0.95$;
  - 0.01 we need c such that $P(t_{n-1} \leq c) = 1 - 0.01 = 0.99$.

Beer example

- We can use `c_05` and `c_01` already obtained.
- The 95% "upper confidence limit" is thus

```
xbar + c_05*se
```

```
## [1] 375.1088
```

which gives the one-sided **confidence interval**

```
c(-Inf, xbar + c_05*se)
```

```
## [1]      -Inf 375.1088
```

- For 99%,

```
c(-Inf, xbar + c_01*se)
```

```
## [1]      -Inf 375.2711
```

- These both include 375!

Using `t.test()`

- We need to explicitly ask for a one-sided analysis:

```
t.test(x, mu = 375, alternative = "less")
```

```
## 
##      One Sample t-test
## 
## data:  x
## t = -1.1094, df = 5, p-value = 0.1589
## alternative hypothesis: true mean is less than 375
## 95 percent confidence interval:
##      -Inf 375.1088
## sample estimates:
## mean of x 
##  374.8667
```

```r
t.test(x, mu = 375, alternative = "less", conf.level = 0.99)
```

```
##
##     One Sample t-test
##
## data:  x
## t = -1.1094, df = 5, p-value = 0.1589
## alternative hypothesis: true mean is less than 375
## 99 percent confidence interval:
##      -Inf 375.2711
## sample estimates:
## mean of x
##  374.8667
```

Observed significance level: the p-value

- Finally, to tie all of this together we relate it all to the p-value.
- The *observed signficance level* (or *p-value*) is the value of α for which the observed data is "right on the edge".
- More precisely that is
  - the smallest **false alarm rate** for which we would "reject" a given value μ_0;
  - the *non-coverage probability* (i.e. $1-$ confidence level) for which μ_0 is on the boundary of the **confidence interval**.

Beer example: two-sided

```
t.test(x, mu = 375, conf.level = 1 - 0.3177)
```

```
##
##      One Sample t-test
##
## data:  x
## t = -1.1094, df = 5, p-value = 0.3177
## alternative hypothesis: true mean is not equal to 375
## 68.23 percent confidence interval:
##  374.7333 375.0000
## sample estimates:
## mean of x
##  374.8667
```

Beer example: one-sided

```r
t.test(x, mu = 375, alternative = "less", conf.level = 1-0.1589)
```

```
## 
## 	One Sample t-test
## 
## data:  x
## t = -1.1094, df = 5, p-value = 0.1589
## alternative hypothesis: true mean is less than 375
## 84.11 percent confidence interval:
##  -Inf  375
## sample estimates:
## mean of x 
##  374.8667
```

Rejection regions

Decision rules

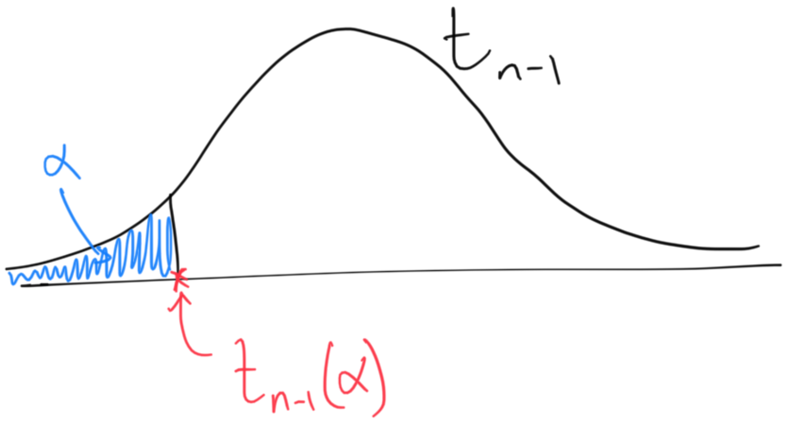
- To test a hypothesis, we previously defined a **decision rule** to reject H_0. That is when the p-value is less than certain fixed preassigned levels, say p-value $\leq \alpha$ where $\alpha = 0.05, 0.10$, etc.
- In other words, we reject or do not reject H_0 according to whether the p-value is less than α or greater than α.
- The α is called the significance level of the test, which is the boundary between rejecting and not rejecting H_0.

Notation

Let $t_{n-1}(\alpha)$ be the **critical value** (or quantile) given by

$$P(t_{n-1} \leq t_{n-1}(\alpha)) = \alpha,$$

or if we are using the standard normal distribution $Z \sim N(0, 1)$ then $z(\alpha)$ is defined by $P(Z \leq z(\alpha)) = \alpha$.



Critical value decision rule

The critical value depends on the level of significance, α, and the distribution of T under H_0, t_{n-1}.

Decision rule

For a test of $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu > \mu_0$, the **decision rule** at level α is:

- reject H_0 if $t_0 \geq t_{n-1}(1-\alpha)$ or equivalently reject H_0 if $t_0 \geq |t_{n-1}(\alpha)|$

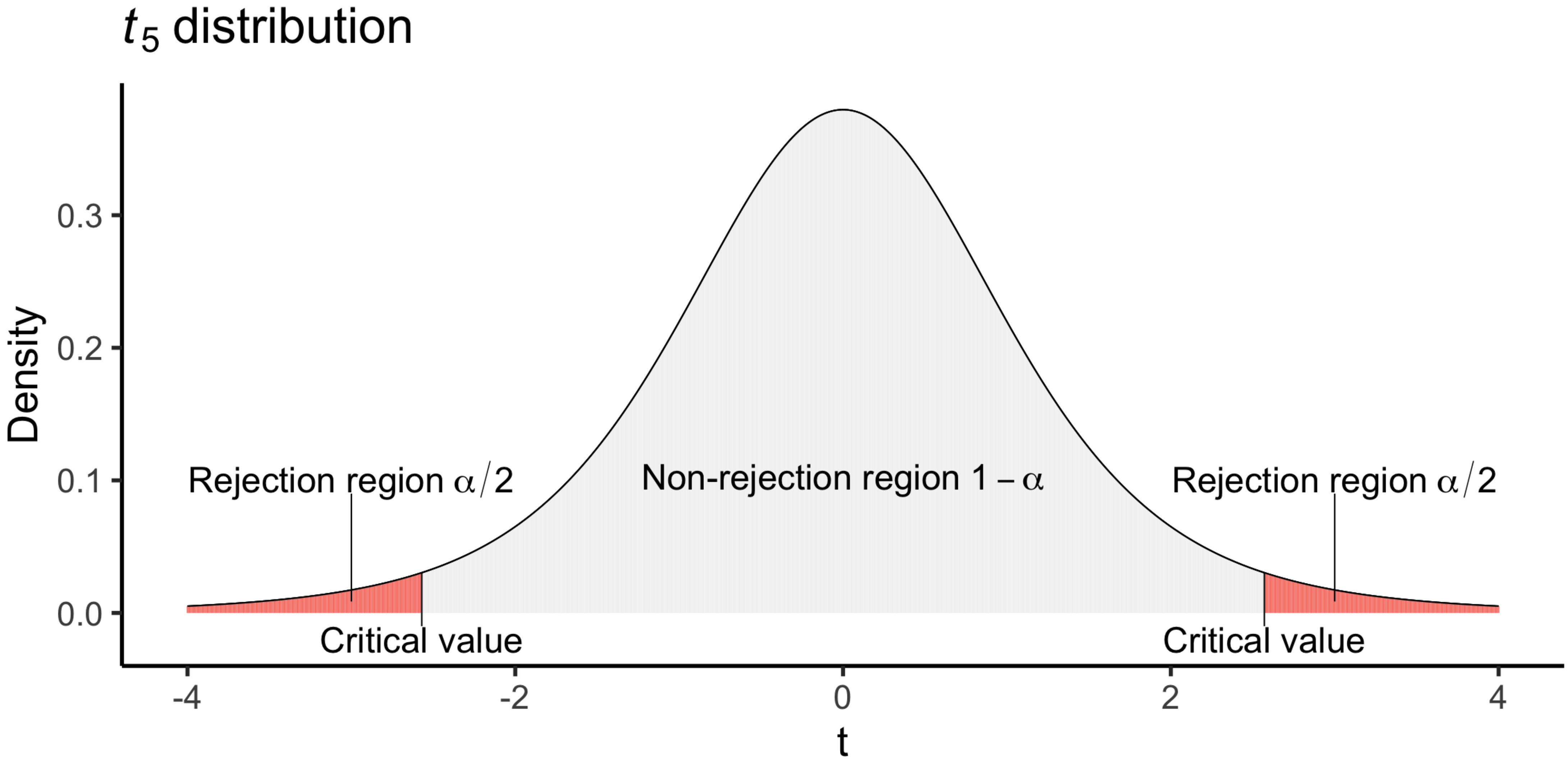
For a test of $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu < \mu_0$, the **decision rule** at level α is:

- reject H_0 if $t_0 \leq t_{n-1}(\alpha)$

For a test of $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu \neq \mu_0$, the **decision rule** at level α is:

- reject H_0 if $|t_0| \geq |t_{n-1}(\alpha/2)|$
- do not reject H_0 if $|t_0| < |t_{n-1}(\alpha/2)|$

Rejection region for two-sided test, $H_1\colon\ \mu \neq \mu_0$



Rejection region for test statistics

- **Hypothesis:** $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu > \mu_0,\ \mu < \mu_0,\ \mu \neq \mu_0$
- **Assumptions:** X_i are iid $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is **unknown**.
- **Test statistic:** $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$
- **Observed test statistic:** $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
- **Rejection region:**
- $H_1\colon\ \mu \lessgtr \mu_0$: $t_0 \leq t_{n-1}(\alpha)$ or $t_0 \geq |t_{n-1}(\alpha)|$
- $H_1\colon\ \mu \neq \mu_0$: $|t_0| \geq |t_{n-1}(\alpha/2)|$
- **Decision:** We reject H_0 if t_0 is in the rejection region.

- **Hypothesis:** $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu > \mu_0,\ \mu \neq \mu_0$
- **Assumptions:** X_i are iid $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is **known**.
- **Test statistic:** $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- **Observed test statistic:** $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
- **Rejection region:**
- $H_1\colon\ \mu \lessgtr \mu_0$: $z_0 \leq z(\alpha)$ or $z_0 \geq |z(\alpha)|$
- $H_1\colon\ \mu \neq \mu_0$: $|z_0| \geq |z(\alpha/2)|$
- **Decision:** We reject H_0 if z_0 is in the rejection region.

Beer contents

We have $n = 6$, $\bar{x} = 374.87$, $s = 0.29$, $t_0 = -1.11$. Hypothesis test using critical value.

- **Hypothesis:** $H_0\colon\ \mu = 375$ vs $H_1\colon\ \mu < 375$
- **Assumptions:** X_i are *iid* rv and follow $N(\mu, \sigma^2)$.
- **Test statistic:** $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$. Under H_0, $T \sim t_{n-1}$.
- **Observed test statistic:** $t_0 = \dfrac{374.87 - 375}{0.29/\sqrt{6}} = -1.11$

- **Critical value:** $t_5(0.05) = -2.015$. I.e. reject if t_0 is less than -2.015
- **Decision:** the observed test statistic, $t_0 = -1.11$ is greater than -2.015, so do not reject H_0.

Rejection region on the data scale

Smoking

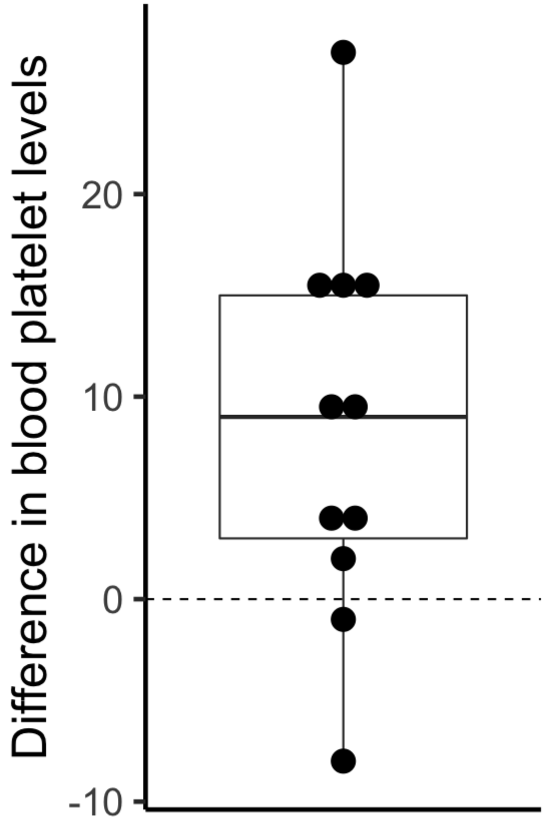
Blood samples from 11 individuals before and after they smoked a cigarette are used to measure aggregation of blood platelets.

```
before = c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)
after =  c(27, 29, 37, 36, 46, 82, 57, 80, 61, 59, 43)
df = data.frame(before, after, difference = after-before)
```

This is a match-pair sample. We reduce the data to one sample by considering the aggregation difference.

Let X_i and Y_i be the blood platelet aggregation levels for the i^{th} person before and after smoking, respectively. Define the change in person i's platelet aggregation levels as $D_i = Y_i - X_i$ and the population mean change in platelet aggregation levels as μ_d.

> Is blod platelet aggregation affected by smoking?



The paired sample t-test on whether the aggregation is affected by smoking.

- **Hypothesis:** $H_0\colon\ \mu_d = 0$ vs $H_1\colon\ \mu_d \neq 0$.
- **Assumptions:** $D_i \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is unknown. The symmetric boxplot shows that the normal assumption is at least approximately satisfied.
- **Test statistic:** $T = \frac{\bar{D}-\mu_d}{S_d/\sqrt{n}}$. Under H_0, $T \sim t_{10}$.
- **Observed test statistic:** $t_0 = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{8.45}{9.65/\sqrt{11}} = 2.9$
- **Rejection region:** Large value of $|t_0|$ argue against H_0 in favour of H_1. Specifically, the critical value is, $|t_{n-1}(\alpha/2)| = |t_{10}(0.025)| = 2.228$
- **Decision:** Since $|t_0| = 2.9 > |t_{10}(0.025)| = 2.2$, there is strong evidence against H_0. Hence we reject H_0 and conclude that the aggregation is affected by smoking at the $\alpha = 0.05$ level of significance.

```r
n = length(df$difference)
dbar = mean(df$difference)
s_d = sd(df$difference)
t0 = dbar/(s_d/sqrt(n))
c(n, dbar, s_d, t0) %>% round(2)
```

```
## [1] 11.00  8.45  9.65  2.91
```

```r
alpha = 0.05
qt(1-alpha/2, n - 1)
```

```
## [1] 2.228139
```

Rejection region for sample mean

The rejection regions for the test using test statistic

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq t_{n-1}(\alpha)$$

on the standardized scale can be transformed to the measurement scale.

We can do this because...

$$\begin{aligned} \alpha &= P\left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq t_{n-1}(\alpha)\right) \\ &= P\left(\bar{x} - \mu_0 \geq t_{n-1}(\alpha) s/\sqrt{n}\right) \\ &= P\left(\bar{x} \geq t_{n-1}(\alpha) s/\sqrt{n} + \mu_0\right) \end{aligned}$$

Which means we can define a rejection region on the measurement scale

$$\{\bar{x} : \bar{x} \geq k_0 = \mu_0 + t_{n-1}(\alpha) s/\sqrt{n}\} \quad \text{for} \quad H_1\colon\ \mu > \mu_0.$$

We have $n = 11$, $\bar{d} = 8.45$, $s_d = 9.65$, $t_0 = 2.91$

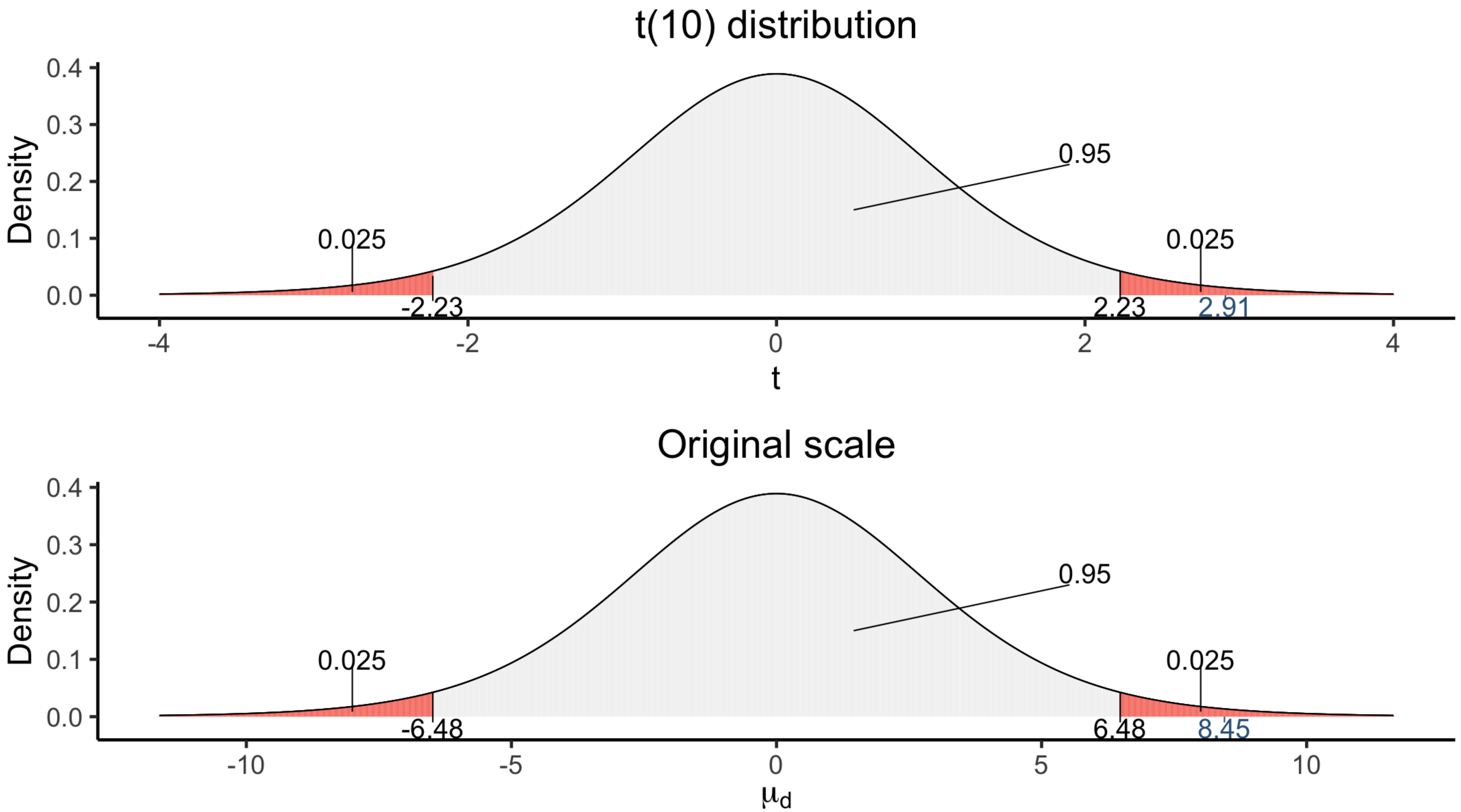
- **Observed test statistic:** $t_0 = \dfrac{\bar{d}}{s_d/\sqrt{n}} = \dfrac{8.45}{9.65/\sqrt{11}} = 2.91$
- **Rejection region:** $\left|\dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}\right| > t_{10}(0.025) = 2.228$, rearranging,

$$\begin{aligned} \bar{d} &< \mu_d - t_{n-1}(0.025)\, s_d/\sqrt{n} \\ \bar{d} &< 0 - 2.228 \times 9.65/\sqrt{11} \\ \bar{d} &< -6.48 \end{aligned}$$

and

$$\begin{aligned} \bar{d} &> \mu_d + t_{n-1}(0.025)\, s_d/\sqrt{n} \\ \bar{d} &> 0 + 2.228 \times 9.65/\sqrt{11} \\ \bar{d} &> 6.48 \end{aligned}$$

- **Decision:** If $\bar{d} < -6.48$ or $\bar{d} > 6.48$ then reject H_0. In this case, $\bar{d} = 8.45 > 6.48$ so we reject H_0.



t(10) distribution
Density
0.4
0.3
0.2
0.1
0.0
0.025
0.95
0.025
-2.23
2.23
2.91
-4
-2
0
2
4
t
Original scale
Density
0.4
0.3
0.2
0.1
0.0
0.025
0.95
0.025
-6.48
6.48
8.45
-10
-5
0
5
10
μ_d

```r
before = c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)
after =  c(27, 29, 37, 36, 46, 82, 57, 80, 61, 59, 43)
df = data.frame(before, after, difference = after-before)
(s_d = sd(df$difference))
```

```
## [1] 9.647421
```

```r
n=nrow(df); mu0=0
(crit_val=qt(0.975,n-1))
```

```
## [1] 2.228139
```

```r
rrlower=mu0-crit_val*s_d/sqrt(n)
rrupper=mu0+crit_val*s_d/sqrt(n)
c(rrlower,rrupper) %>% round(2)
```

```
## [1] -6.48  6.48
```

Beer contents

We have $n = 6$, $\bar{x} = 374.87$, $s = 0.29$, $t_0 = -1.11$. Hypothesis test using rejection region with $\alpha = 0.05$.

- **Hypothesis:** $H_0\colon\ \mu = 375$ vs $H_1\colon\ \mu < 375$
- **Assumptions:** X_i are *iid* rv and follow $N(\mu, \sigma^2)$.
- **Test statistic:** $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$. Under H_0, $T \sim t_{n-1}$.

- **Rejection region (on the data scale):**

$$\begin{aligned} \frac{\bar{X} - \mu}{s/\sqrt{n}} &< t_{n-1}(0.05) \\ \bar{X} &< \mu + t_{n-1}(0.05)\, s/\sqrt{n} \\ \bar{X} &< 375 - 2.015 \times 0.29/\sqrt{6} \\ \bar{X} &< 374.74 \end{aligned}$$

  I.e. reject if $\bar{x}$ is less than 374.74.

- **Decision:** the observed sample mean, $\bar{x} = 374.9$ is greater than 374.74, so do not reject H_0.

Confidence intervals

To link decision rules with **confidence intervals**:

- if the population parameter is inside the **confidence interval** then it is within the range of plausible values
- **do not reject** H_0 at the α level if significance if the value of the population parameter under the null hypothesis is **inside** the $100(1-\alpha)\%$ **confidence interval**

References

For further details see Larsen and Marx (2012), sections 6.1, 6.2 and 6.4.

Larsen, R. J. and M. L. Marx (2012). *An Introduction to Mathematical Statistics and its Applications*. 5th ed. Boston, MA: Prentice Hall. ISBN: 978-0-321-69394-5.