Tests of hypotheses

Confidence interval: Form an interval (on the basis of data) of plausible values for a population parameter.

Test of hypothesis: Answer a yes or no question regarding a population parameter.

Examples:

- $\longrightarrow$ Do the two strains have the same average response?
- $\longrightarrow$ Is the concentration of substance X in the water supply above the safe limit?
- $\longrightarrow$ Does the treatment have an effect?

Example

We have a quantitative assay for the concentration of antibodies against a certain virus in blood from a mouse.

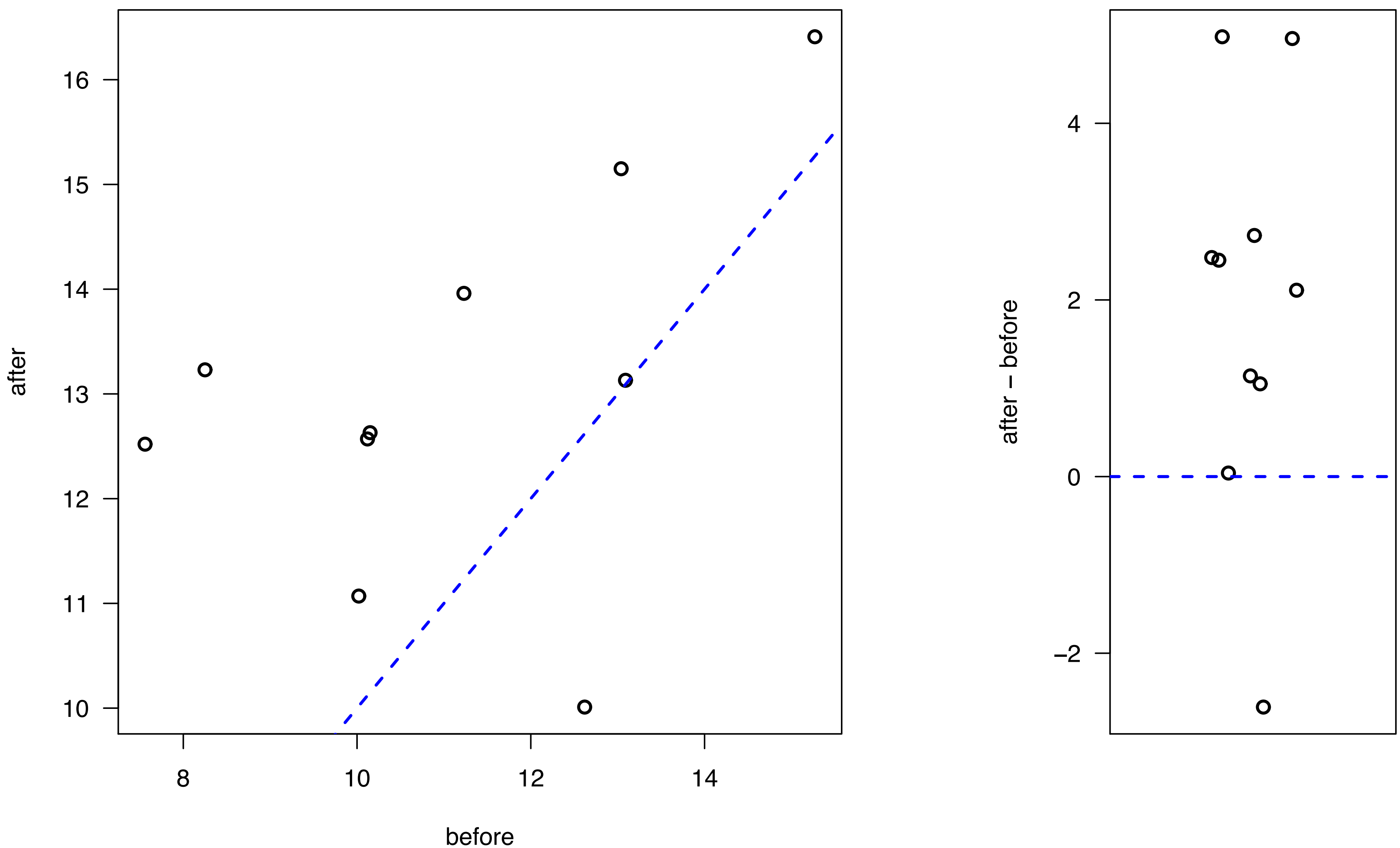
We apply our assay to a set of ten mice before and after the injection of a vaccine. (This is called a “paired” experiment.)

Let X_i denote the differences between the measurements (“after” minus “before”) for mouse i.

We imagine that the X_i are independent and identically distributed Normal(μ, σ).

$\longrightarrow$ Does the vaccine have an effect? In other words: Is $\mu \neq 0$?

The data



Hypothesis testing

We consider two hypotheses:

Null hypothesis, H_0: $\mu = 0$ Alternative hypothesis, H_a: $\mu \neq 0$

Type I error:	Reject H_0 when it is true	(false positive)
Type II error:	Fail to reject H_0 when it is false	(false negative)

We set things up so that a Type I error is a worse error (and so that we are seeking to prove the alternative hypothesis). We want to control the rate (the significance level, α) of such errors.

⟶ Test statistic: $T = (\overline{X} - 0)/(S/\sqrt{10})$

⟶ We reject H_0 if $|T| > t^\star$, where $t^\star$ is chosen so that

$\Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \Pr(|T| > t^\star \mid \mu = 0) = \alpha$.

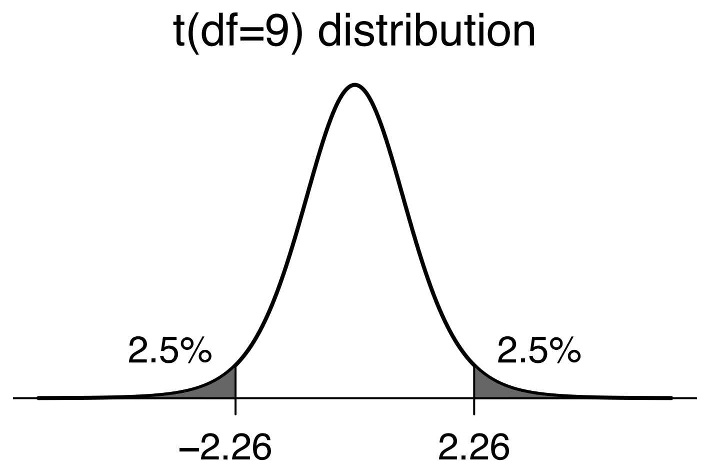
(generally α = 5%)

Example (continued)

Under H_0 (i.e., when $\mu = 0$),

$T = (\overline{X} - 0)/(S/\sqrt{10}) \sim t(df = 9)$

We reject H_0 if $|T| > 2.26$.



As a result, if H_0 is true, there's a 5% chance that you'll reject it!

For the observed data:

$\bar{x} = 1.93$, $s = 2.24$, $n = 10$ $T = (1.93 - 0) / (2.24/\sqrt{10}) = 2.72$

$\longrightarrow$ Thus we reject H_0.

The goal

⟶ We seek to prove the alternative hypothesis.

⟶ We are happy if we reject H_0.

⟶ In the case that we reject H_0, we might say:

Either H_0 is false, or a rare event occurred.

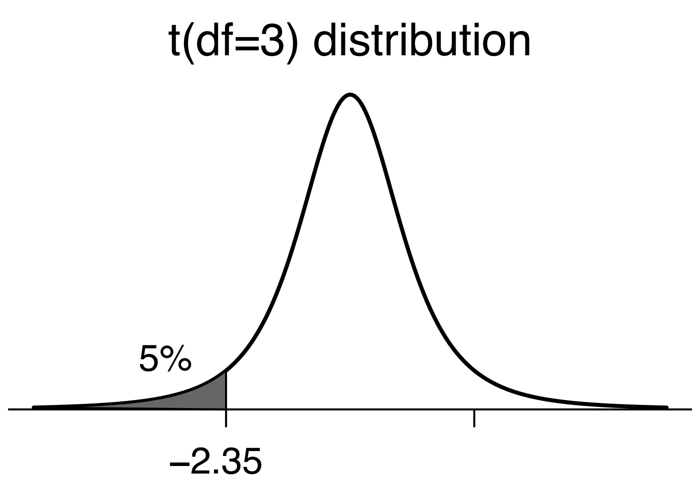
Another example

Question: is the concentration of substance X in the water supply above the safe level?

$X_1, X_2, \ldots, X_4 \sim$ iid Normal(μ, σ).

$\longrightarrow$ We want to test H$_0$: $\mu \geq 6$ (unsafe) versus H$_a$: $\mu < 6$ (safe).

Test statistic: $T = \dfrac{\overline{X} - 6}{S/\sqrt{4}}$

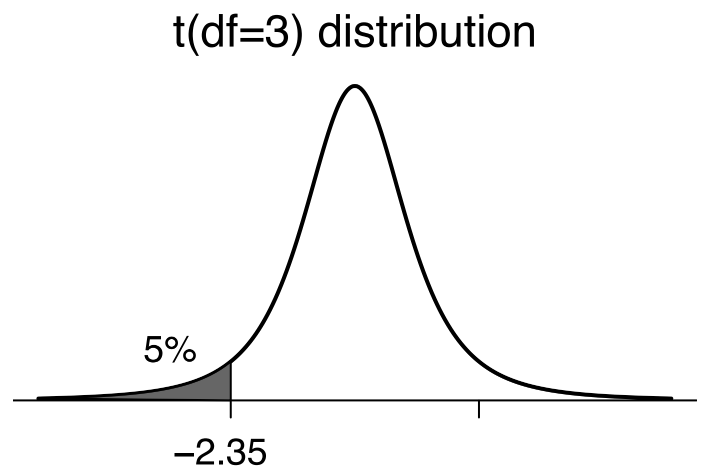


If we wish to have the significance level α = 5%, the rejection region is $T < t^{\star} = -2.35$.

One-tailed vs two-tailed tests

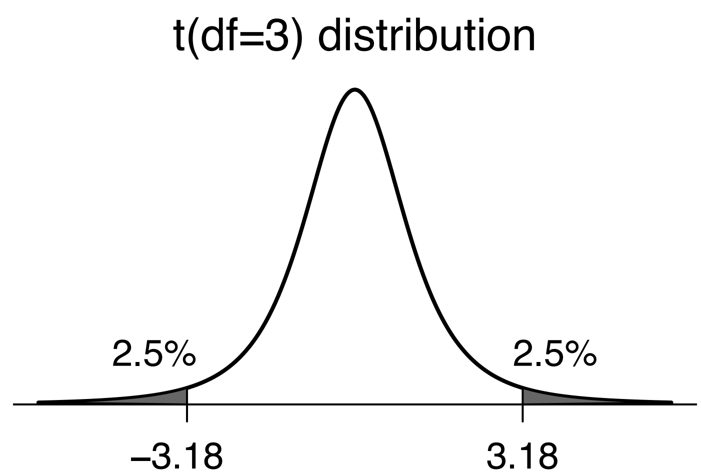
If you are trying to prove that a treatment improves things, you want a one-tailed (or one-sided) test.

You'll reject H_0 only if $T < t^\star$.



If you are just looking for a difference, use a two-tailed (or two-sided) test.

You'll reject H_0 if $T < t^\star$ or $T > t^\star$.



P-values

P-value: $\longrightarrow$ the smallest significance level (α) for which you would fail to reject H_0 with the observed data.

$\longrightarrow$ the probability, if H_0 was true, of receiving data as extreme as what was observed.

$X_1, \ldots, X_{10} \sim$ iid Normal(μ, σ), H_0: $\mu = 0$; H_a: $\mu \neq 0$.

$\bar{x} = 1.93$; s = 2.24

$T_{obs} = \frac{1.93-0}{2.24/\sqrt{10}} = 2.72$

P-value = $\Pr(|T| > T_{obs}) = 2.4\%$.

```
2*pt(-2.72,9)
```

Another example

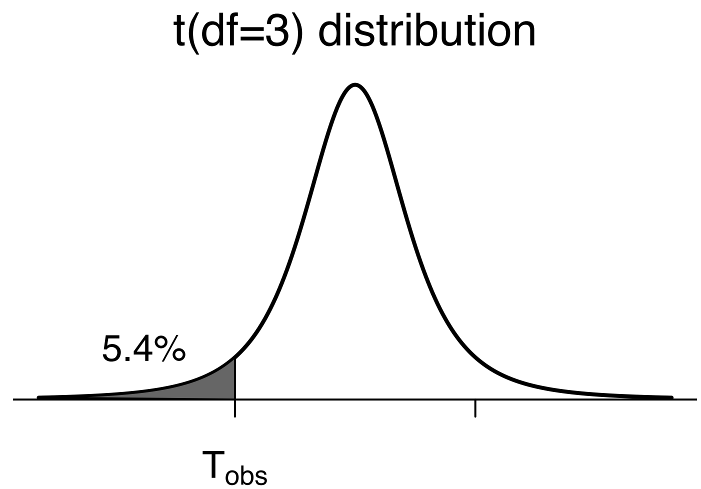
$X_1, \ldots, X_4 \sim$ Normal(μ, σ) H$_0$: $\mu \geq 6$; H$_a$: $\mu < 6$.

$\bar{x}$ = 5.51; s = 0.43

$T_{obs} = \frac{5.51 - 6}{0.43/\sqrt{4}}$ = –2.28

P-value = Pr(T $<$ T_{obs} | μ = 6) = 5.4%.

```
pt(-2.28, 3)
```



$\longrightarrow$ The P-value quantifies how likely it is to get data as extreme as the data observed, assuming the null hypothesis was true.

Recall: We want to prove the alternative hypothesis (i.e., reject H$_0$, receive a small P-value)

Hypothesis tests and confidence intervals

→ The 95% confidence interval for μ is the set of values, μ_0, such that the null hypothesis $H_0 : \mu = \mu_0$ would not be rejected by a two-sided test with α = 5%.

The 95% CI for μ is the set of plausible values of μ. If a value of μ is plausible, then as a null hypothesis, it would not be rejected.

For example:

9.98 9.87 10.05 10.08 9.99 9.90 assumed to be iid Normal(μ,σ)

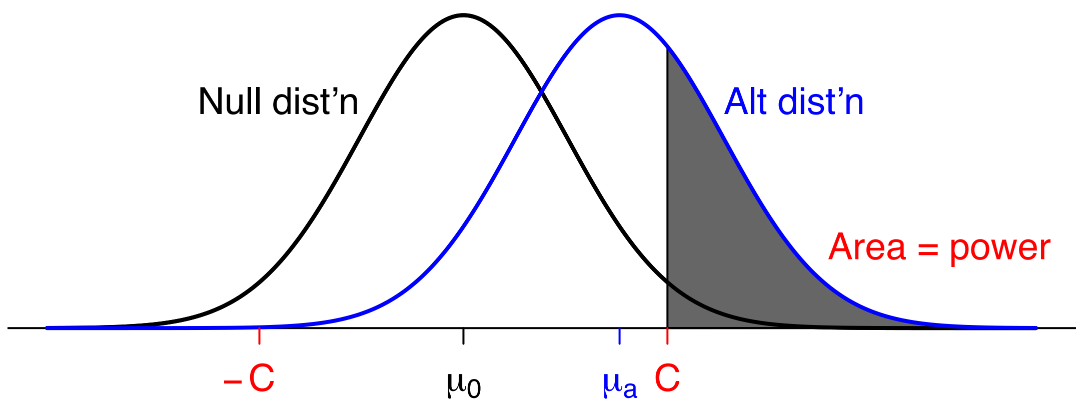
$\bar{x}$ = 9.98; s = 0.082; n = 6; `qt(0.975,5)` = 2.57

The 95% CI for μ is

$$9.98 \pm 2.57 \times 0.082 / \sqrt{6} \quad = \quad 9.98 \pm 0.086 \quad = \quad (9.89,10.06)$$

Power

The power of a test = Pr(reject H_0 | H_0 is false).



The power depends on:
- The null hypothesis and test statistic
- The sample size
- The true value of μ
- The true value of σ

Why "fail to reject"?

If the data are insufficient to reject H_0, we say,

The data are insufficient to reject H_0.

We shouldn't say, *We have proven H_0.*

→ We may only have low power to detect anything but extreme differences.

→ We control the rate of type I errors ("false positives") at 5% (or whatever), but we may have little or no control over the rate of type II errors.

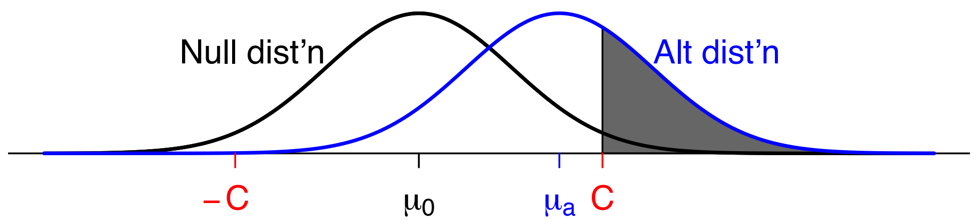
The effect of sample size

Let $X_1, \ldots, X_n$ be iid Normal(μ, σ).

We wish to test $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$.

Imagine $\mu = \mu_a$.

n = 4



n = 16