P-values and statistical tests
3. t-test

Marek Gierliński
Division of Computational Biology

Hand-outs available at http://is.gd/statlec

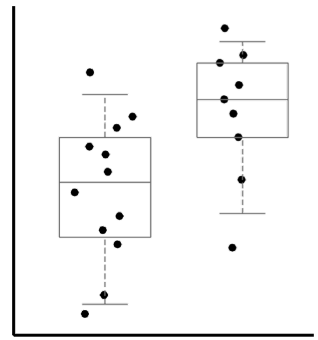
Statistical testing

Null hypothesis
H_0: no effect

All other assumptions

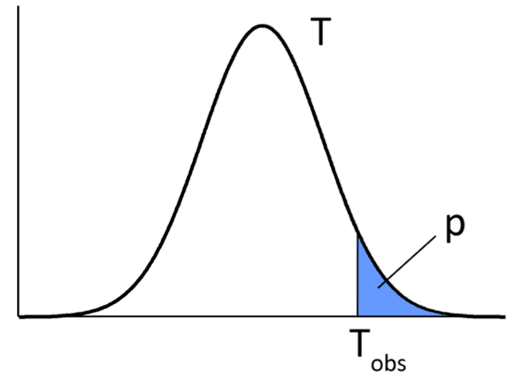
Significance level
$\alpha = 0.05$

Data



Test statistic T_{obs}

Statistical test against H_0



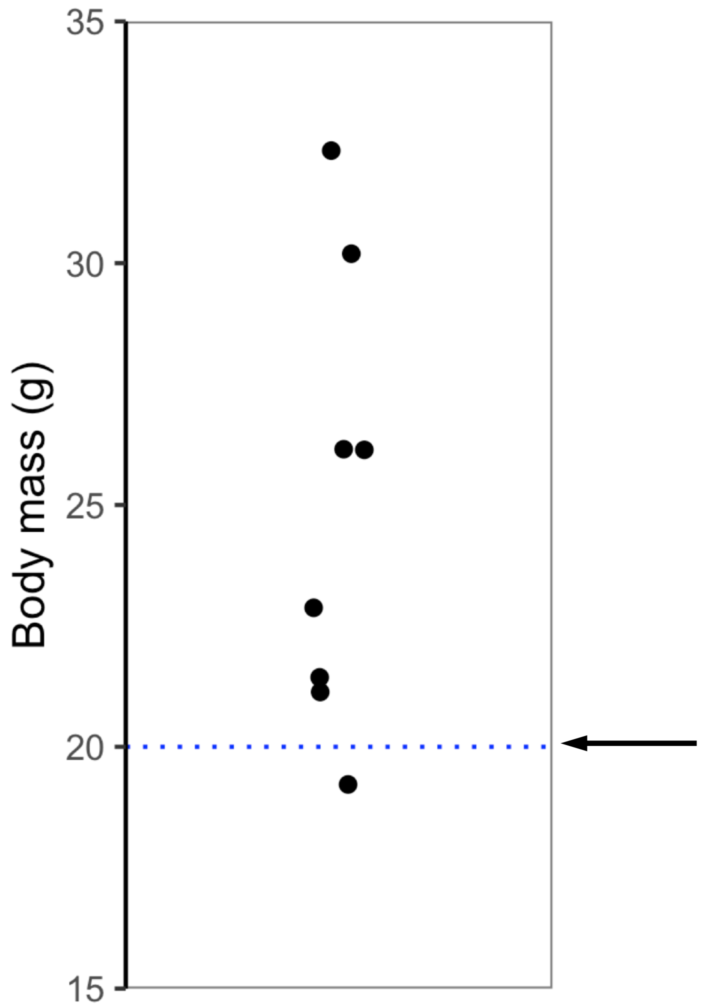
p-value: probability that the observed effect is random

$p < \alpha$
Reject H_0
Effect is real

$p \geq \alpha$
Insufficient evidence

One-sample t-test

One-sample t-test



Null hypothesis: the sample came from a population with mean $\mu = 20$ g

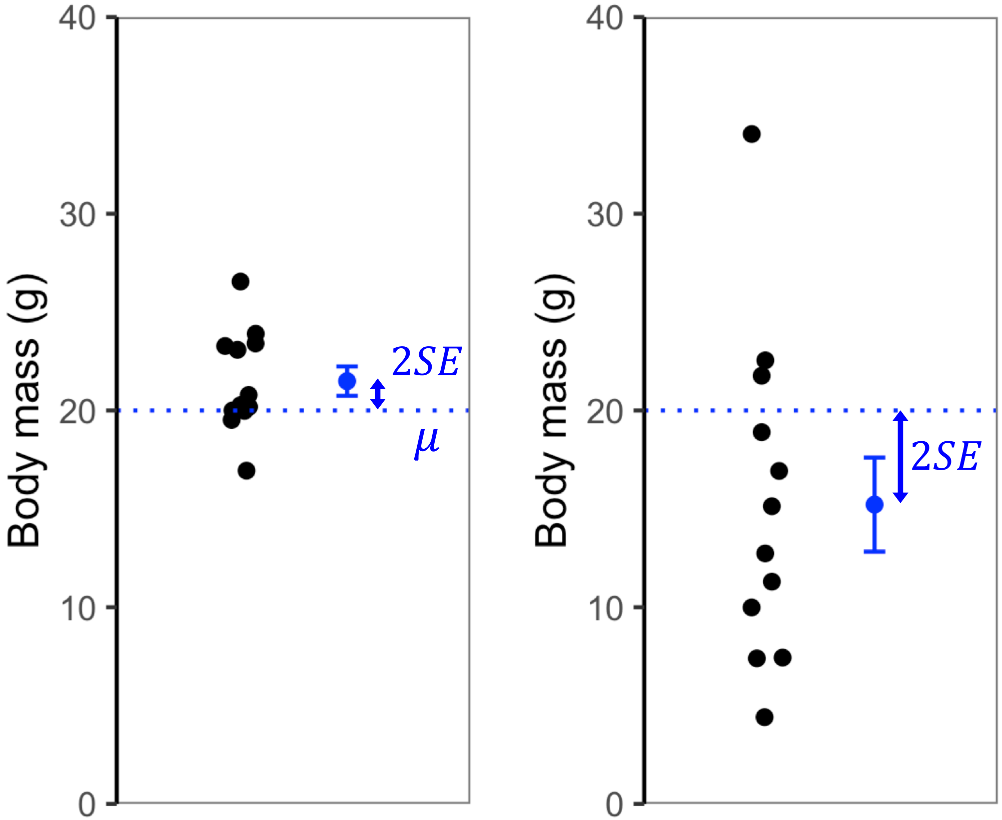
t-statistic

- Sample $x_1, x_2, \ldots, x_n$

  M - mean

  SD - standard deviation

  $SE = SD/\sqrt{n}$ - standard error



- From these we can find

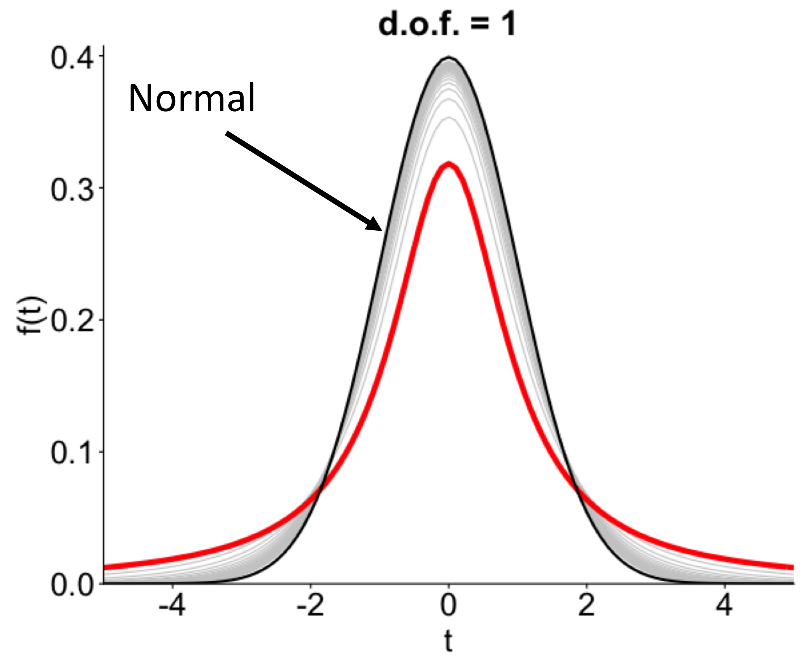
$$t = \frac{M - \mu}{SE}$$

- more generic form:

$$t = \frac{\text{deviation}}{\text{standard error}}$$

Note: Student's *t*-distribution

- *t*-statistic is distributed with *t*-distribution
- Standardized
- One parameter: degrees of freedom, ν
- For large ν approaches normal distribution



William Gosset

- Brewer and statistician
- Developed Student's *t*-distribution

- Worked for Guinness, who prohibited employees from publishing any papers
- Published as "Student"

- Worked with Fisher and developed the *t*-statistic in its current form
- Always worked with experimental data
- Progenitor bioinformatician?

William Sealy Gosset (1876-1937)

William Gosset

- Brewer and statistician
- Developed Student's *t*-distribution

- Worked for Guinness, who prohibited employees from publishing any papers
- Published as "Student"

- Worked with Fisher and developed the *t*-statistic in its current form
- Always worked with experimental data
- Progenitor bioinformatician?

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

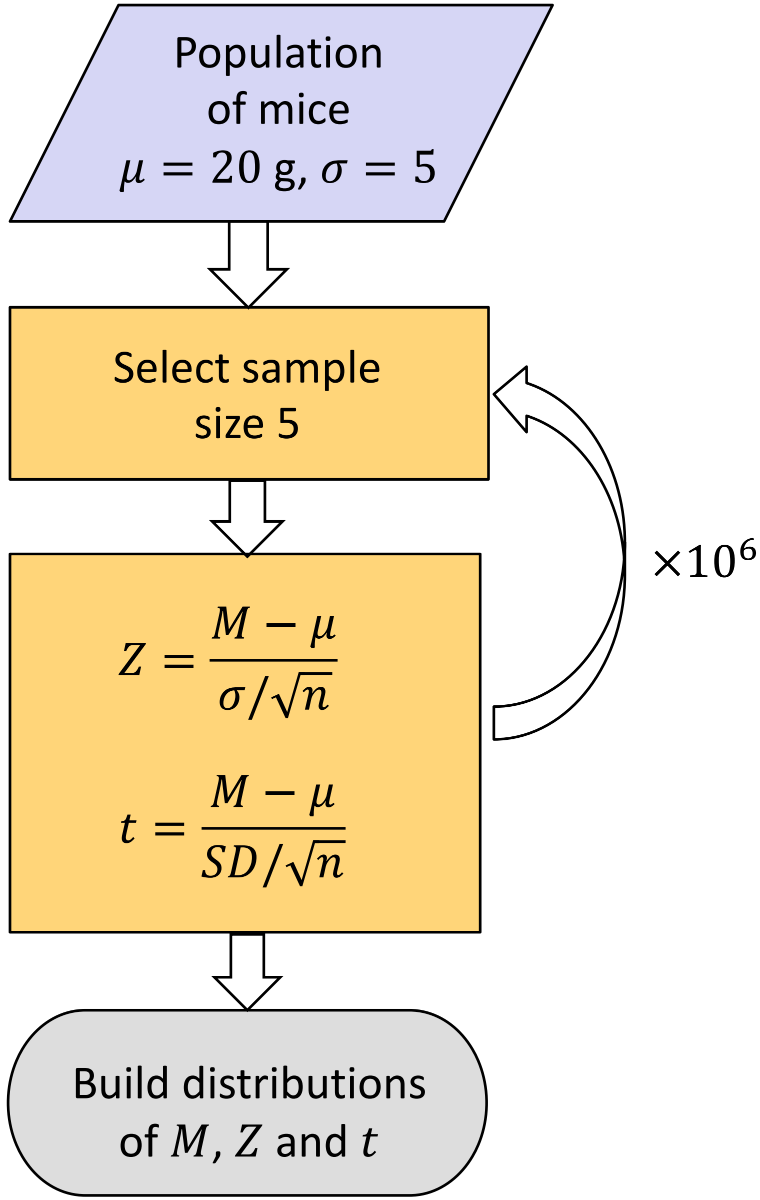
Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

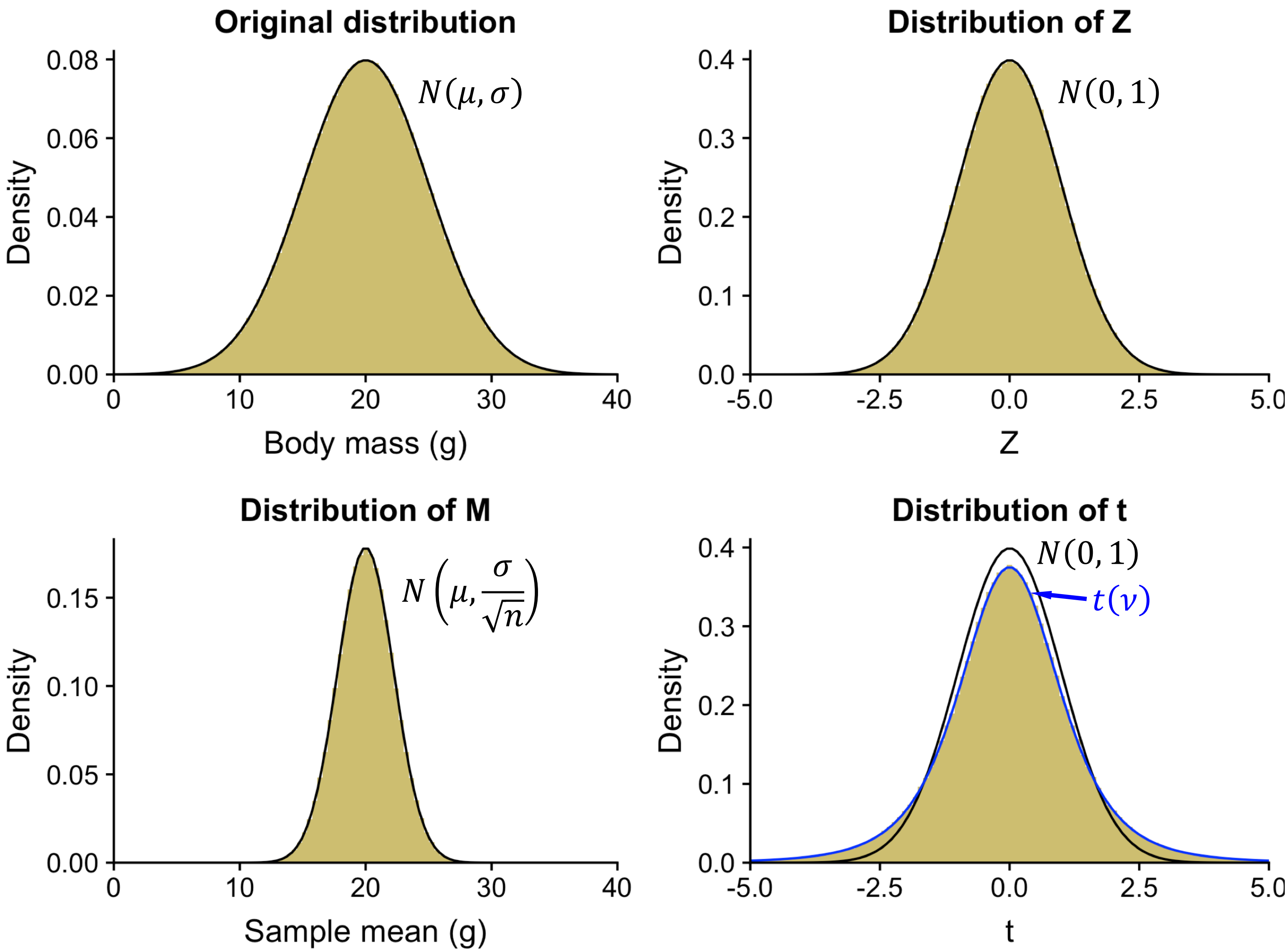
Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.

Null distribution for the deviation of the mean



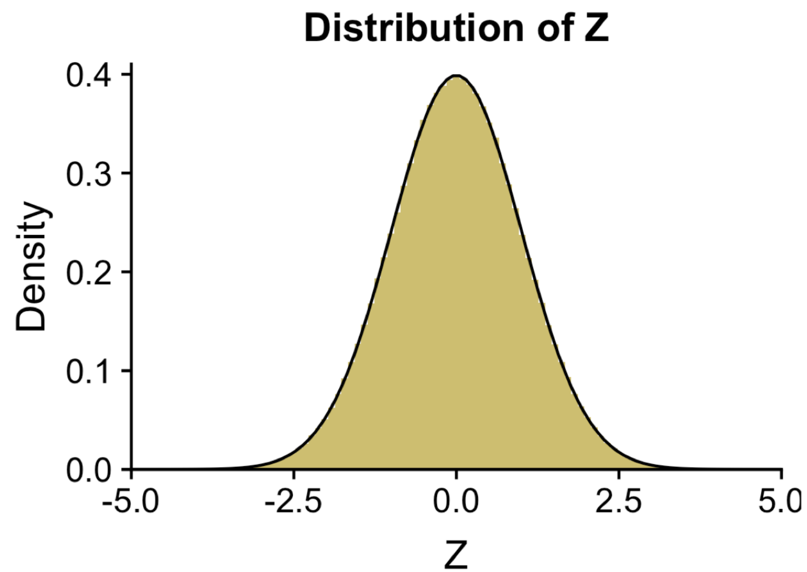
Null distribution for the deviation of the mean



Null distribution for the deviation of the mean

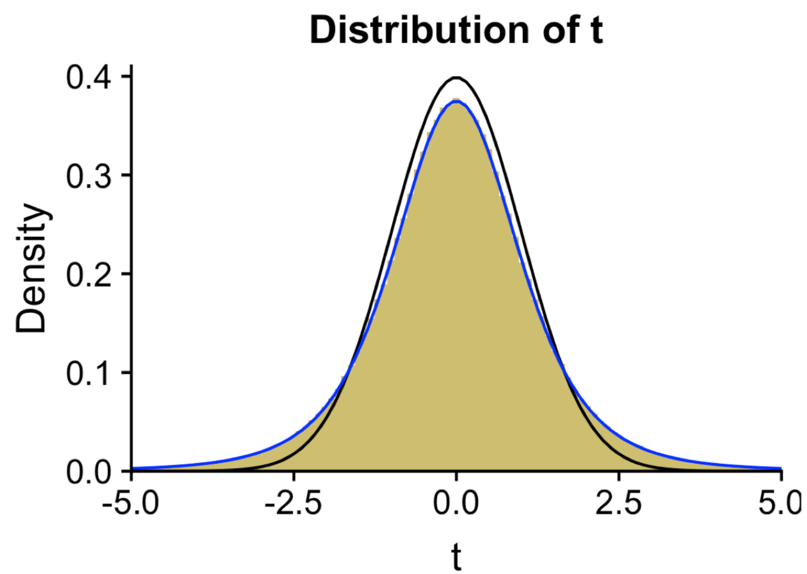
$$Z = \frac{M - \mu}{\sigma/\sqrt{n}}$$

σ - population parameter (unknown)



$$t = \frac{M - \mu}{SD/\sqrt{n}} = \frac{M - \mu}{SE}$$

SD - sample estimator (known)



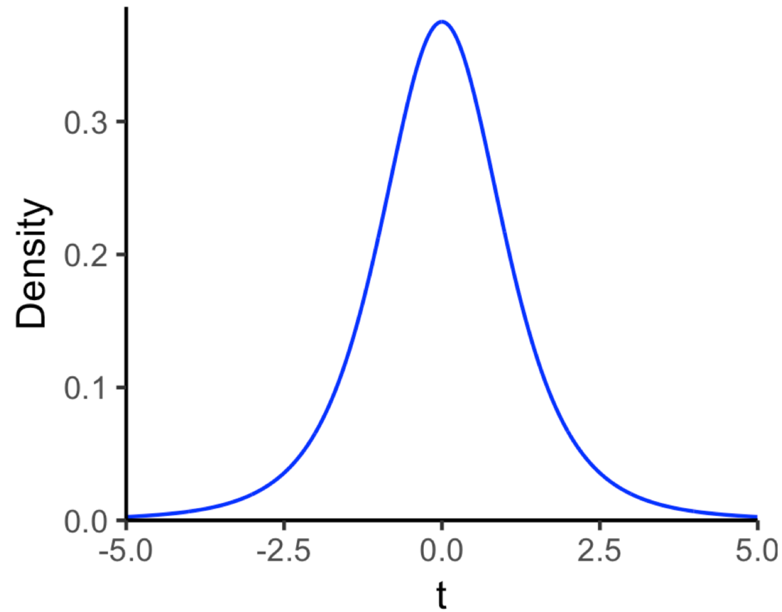
One-sample t-test

- Consider a sample of n measurements
  - M – sample mean
  - SD – sample standard deviation
  - $SE = SD/\sqrt{n}$ – sample standard error
- Null hypothesis: the sample comes from a population with mean μ
- Test statistic

$$t = \frac{M - \mu}{SE}$$

- is distributed with t-distribution with $n - 1$ degrees of freedom

null distribution
t-distribution with 4 d.o.f.



Null distribution represents all random samples when the null hypothesis is true

One-sample *t*-test: example

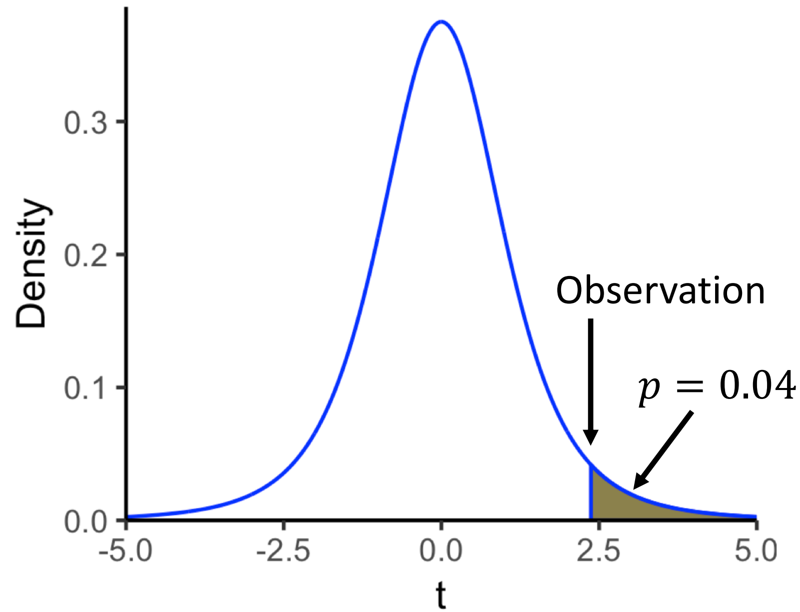
- H_0: $\mu = 20$ g
- 5 mice with body mass (g):
- 19.5, 26.7, 24.5, 21.9, 22.0

$$M = 22.92 \text{ g}$$
$$SD = 2.76 \text{ g}$$
$$SE = 1.23 \text{ g}$$

$$t = \frac{22.92 - 20}{1.23} = 2.37$$
$$\nu = 4$$

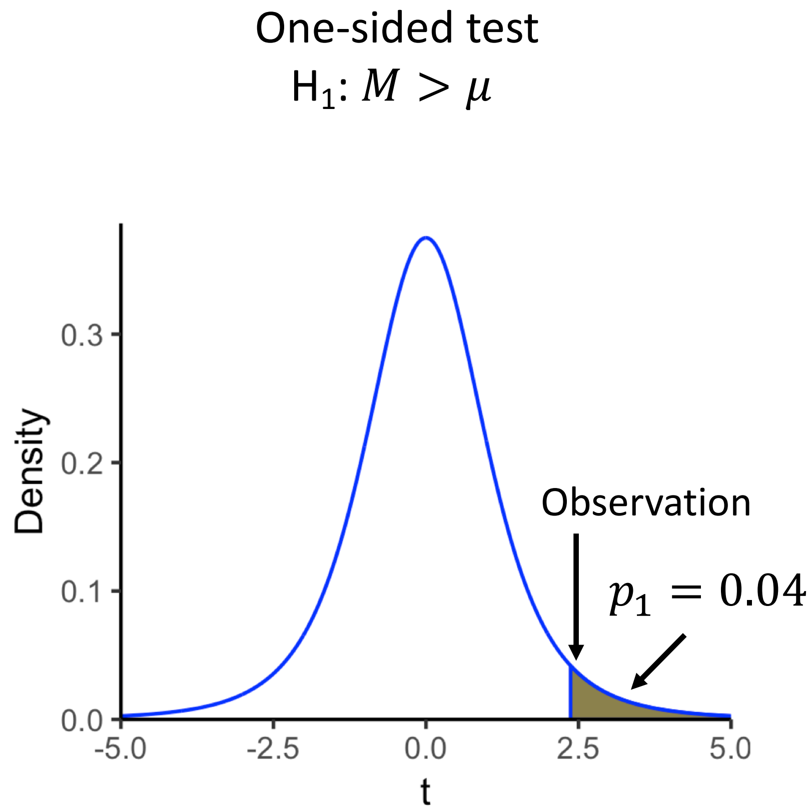
$$p = 0.04$$

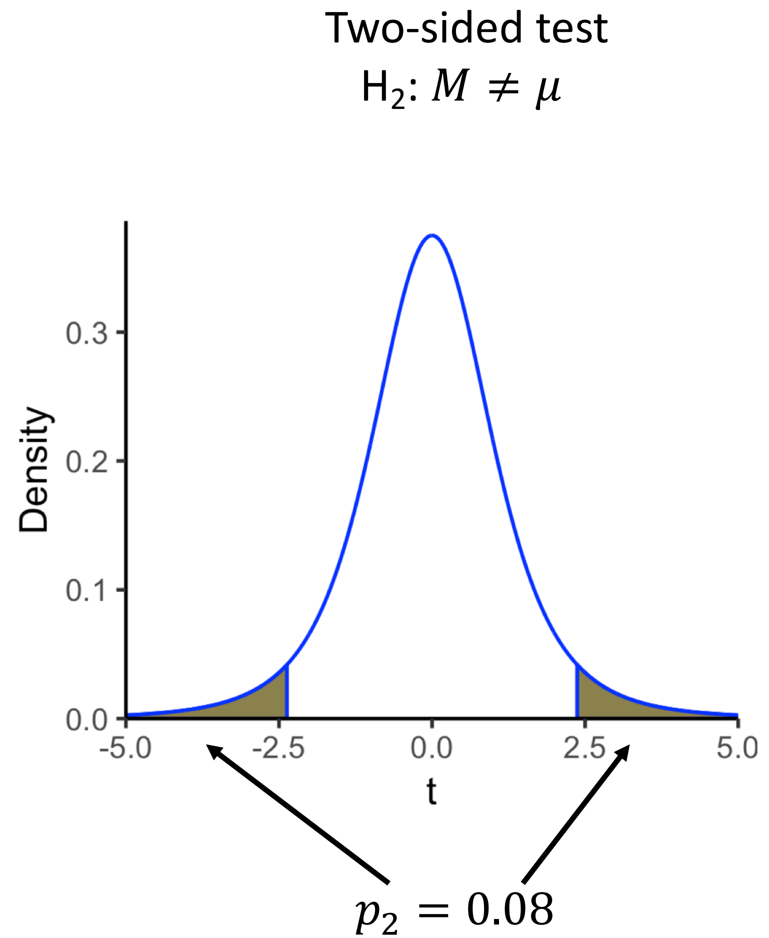
null distribution
t-distribution with 4 d.o.f.




```
> mass <- c(19.5, 26.7, 24.5, 21.9, 22.0)
> M <- mean(mass)
> n <- length(mass)
> SE <- sd(mass) / sqrt(n)
> t <- (M - 20) / SE
[1] 2.36968
> 1 - pt(t, n - 1)
[1] 0.03842385
```

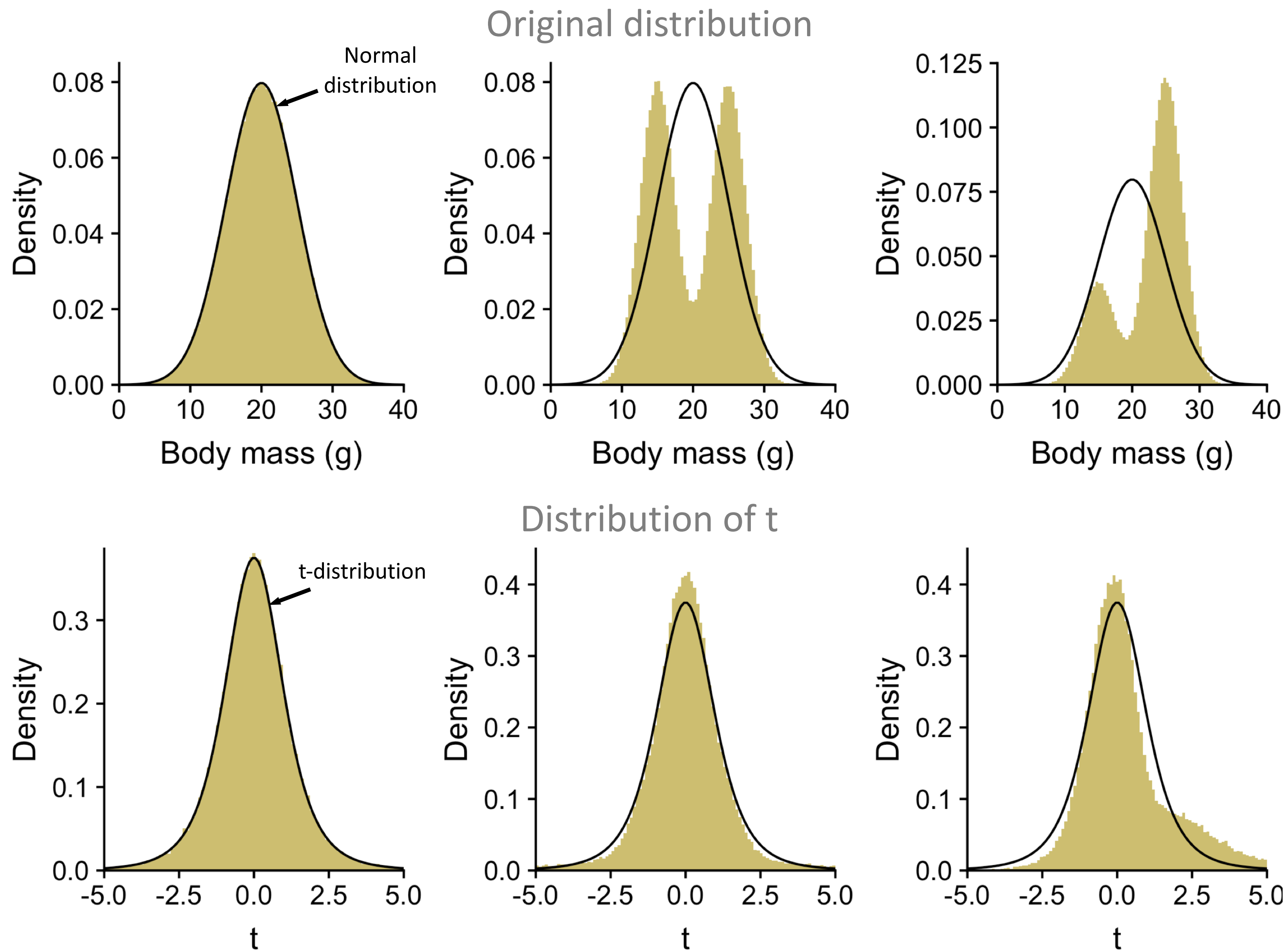
Sidedness





$$p_2 = 2p_1$$

Normality of data



One-sample t-test: summary

Input	sample of n measurements theoretical value μ (population mean)
Assumptions	Observations are random and independent Data are normally distributed
Usage	Examine if the sample is consistent with the population mean
Null hypothesis	Sample came from a population with mean μ
Comments	Use for differences and ratios (e.g. SILAC) Works well for non-normal distribution, as long as it is symmetric

How to do it in R?

```
# One-sided t-test
> mass = c(19.5, 26.7, 24.5, 21.9, 22.0)
> t.test(mass, mu=20, alternative="greater")

        One Sample t-test

data:  mass
t = 2.3697, df = 4, p-value = 0.03842
alternative hypothesis: true mean is greater than 20
95 percent confidence interval: 20.29307      Inf
sample estimates:
mean of x
     22.92
```

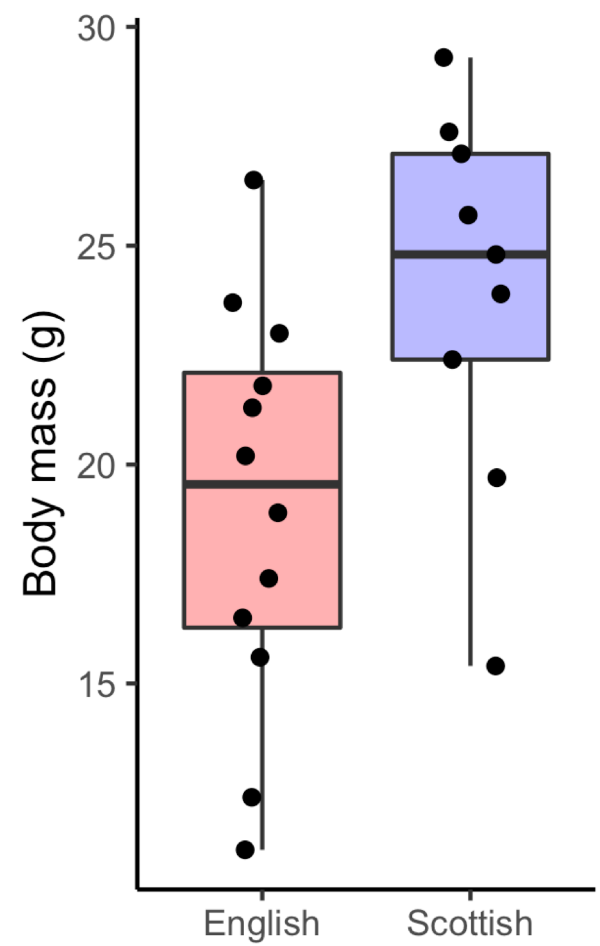
Two-sample *t*-test

Two samples

- Consider two samples (different sizes)
- Are they different?
- Are their means different?
- Do they come from populations with different means?

$n_E = 12$
$M_E = 19.0$ g
$S_E = 4.6$ g

$n_S = 9$
$M_S = 24.0$ g
$S_S = 4.3$ g



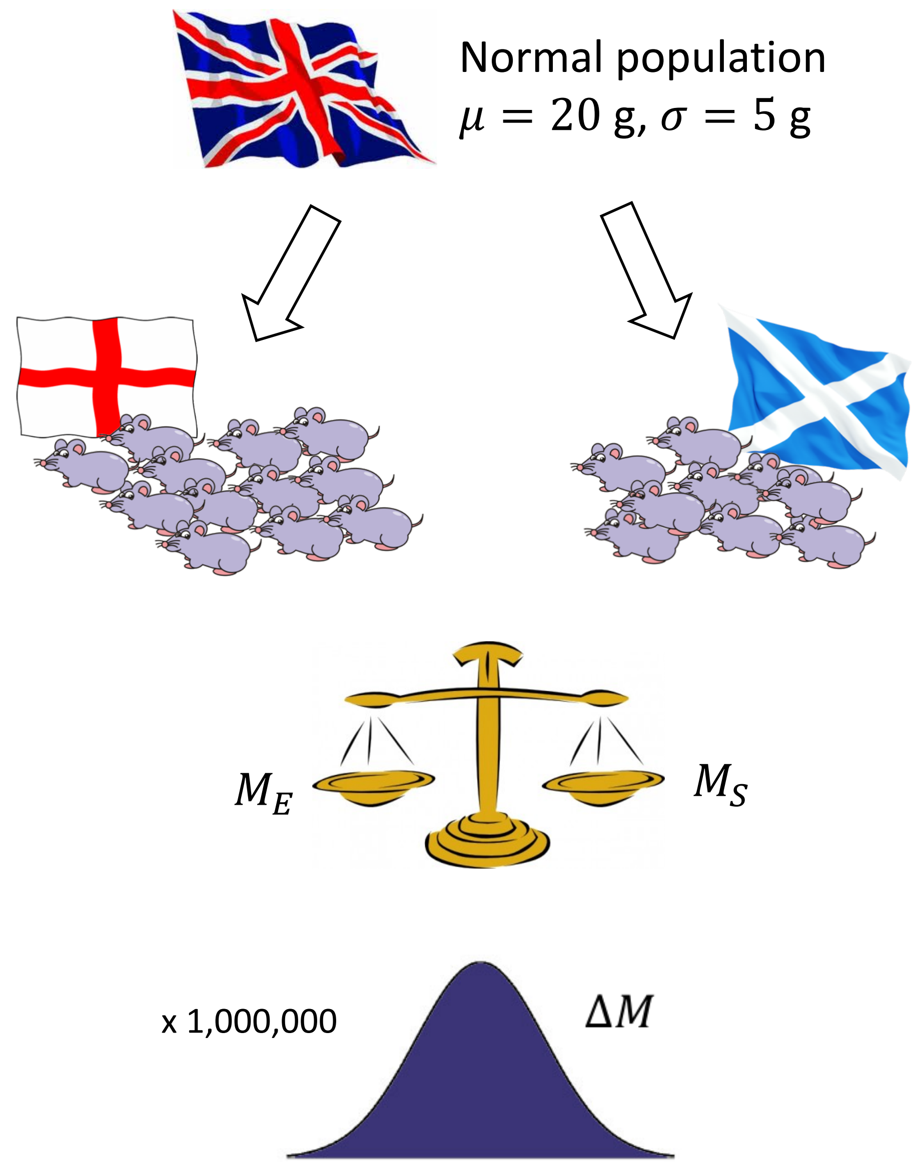
Gedankenexperiment: null distribution

Population
of British mice
$\mu = 20$ g, $\sigma = 5$

Select two samples
size 12 and 9

$$t = \frac{M_E - M_S}{SE}$$

Build distribution
of ΔM and t

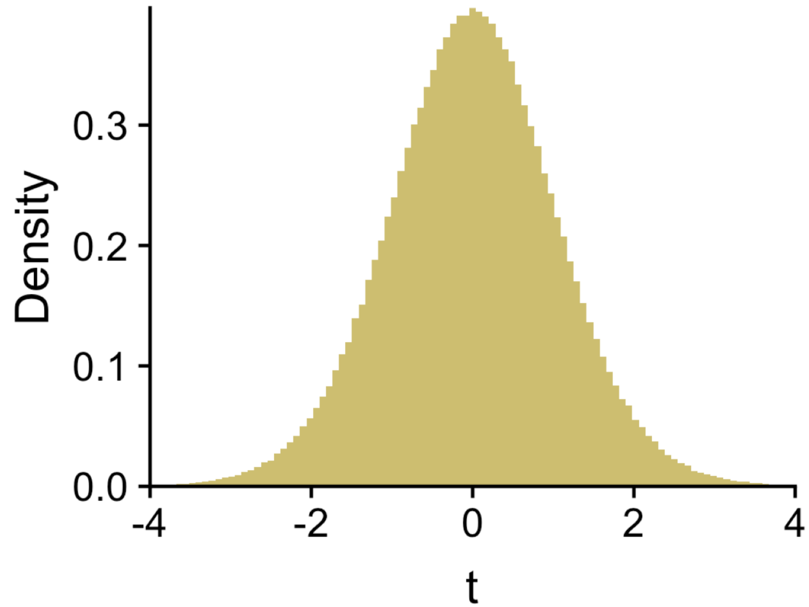


Null distribution

- *Gedankenexperiment*
- Test statistic

$$t = \frac{M_1 - M_2}{SE}$$

is distributed with t-distribution with ν degrees of freedom



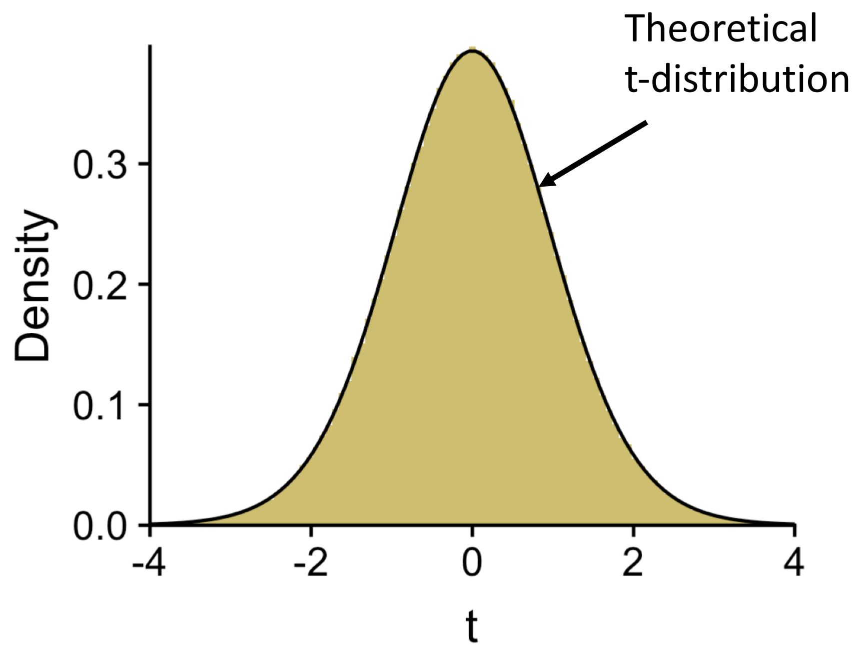
Null distribution represents all random samples when the null hypothesis is true

Null distribution

- *Gedankenexperiment*
- Test statistic

$$t = \frac{M_1 - M_2}{SE}$$

is distributed with t-distribution with ν degrees of freedom



Null distribution represents all random samples when the null hypothesis is true

Two-sample t-test

- Two samples of size n_1 and n_2

- Null hypothesis: both samples come from populations of the same mean
- H_0: $\mu_1 = \mu_2$

- Find M_1, M_2 and SE
- Test statistic

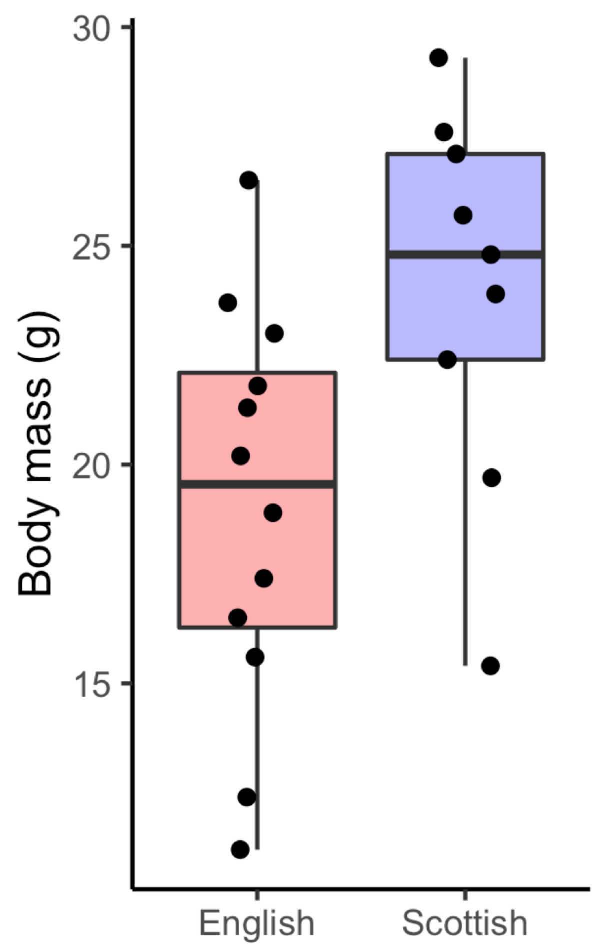
$$t = \frac{M_1 - M_2}{SE}$$

is distributed with t-distribution with ν degrees of freedom

- How do we find SE from two samples?

$n_E = 12$
$M_E = 19.0$ g
$SD_E = 4.6$ g

$n_S = 9$
$M_S = 24.0$ g
$SD_S = 4.3$ g



Case 1: equal variances

- Assume that both distributions have the same variance (or standard deviation)
- Use pooled variance estimator:

$$SD_{1,2}^2 = \frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}$$

- And then the standard error and the number of degrees of freedom are

$$SE = SD_{1,2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\nu = n_1 + n_2 - 2$$

In case of equal samples sizes, $n_1 = n_2 = n$, these equations simplify:

$$SD_{1,2}^2 = \frac{1}{2}(SD_1^2 + SD_2^2)$$

$$SE = \frac{SD_{1,2}}{\sqrt{n}}$$

$$\nu = 2n - 2$$

Case 1: equal variances, example

$n_E = 12$
$M_E = 19.04 \text{ g}$
$SD_E = 4.61 \text{ g}$

$n_S = 9$
$M_S = 23.99 \text{ g}$
$SD_S = 4.32 \text{ g}$

$SD_{1,2} = 4.49 \text{ g}$

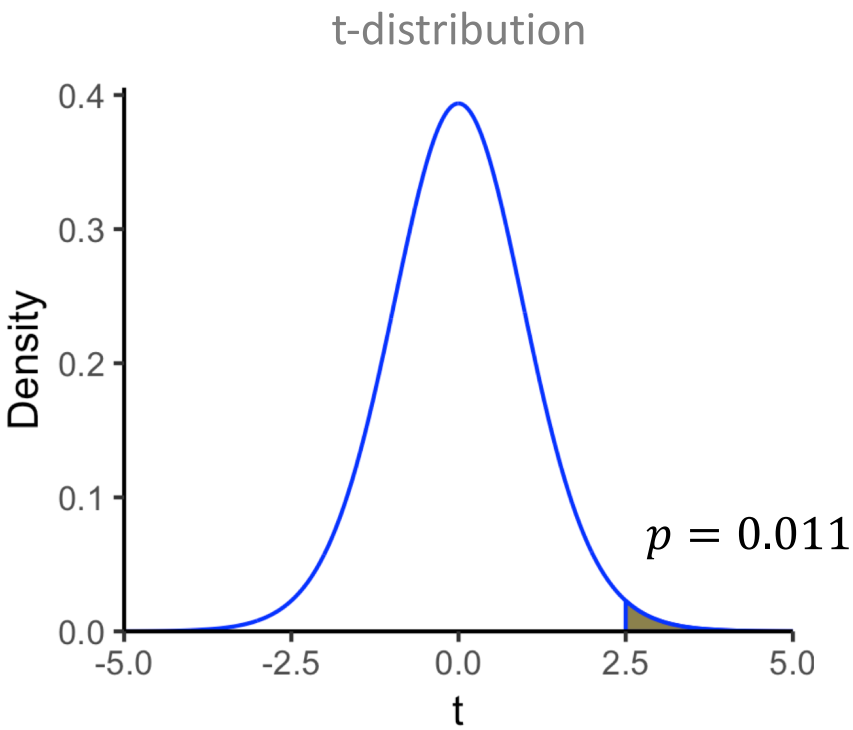
$SE = 1.98 \text{ g}$

$\nu = 19$

$$t = \frac{23.99 - 19.04}{1.98} = 2.499$$

$p = 0.011$ (one-sided)

$p = 0.022$ (two-sided)




```
> 1 - pt(2.499, 19)
[1] 0.01089314
```

Case 2: unequal variances

- Assume that distributions have different variances
- Welch's t-test
- Find individual standard errors (squared):

$$SE_1^2 = \frac{SD_1^2}{n_1} \qquad SE_2^2 = \frac{SD_2^2}{n_2}$$

- Find the common standard error:

$$SE = \sqrt{SE_1^2 + SE_2^2}$$

- Number of degrees of freedom

$$\nu \approx \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}}$$

Case 2: unequal variances, example

$n_E = 12$
$M_E = 19.04$ g
$SD_E = 4.61$ g

$n_S = 9$
$M_S = 23.99$ g
$SD_S = 4.32$ g

$SE_E^2 = 1.77\ \mathrm{g}^2$

$SE_S^2 = 2.07\ \mathrm{g}^2$

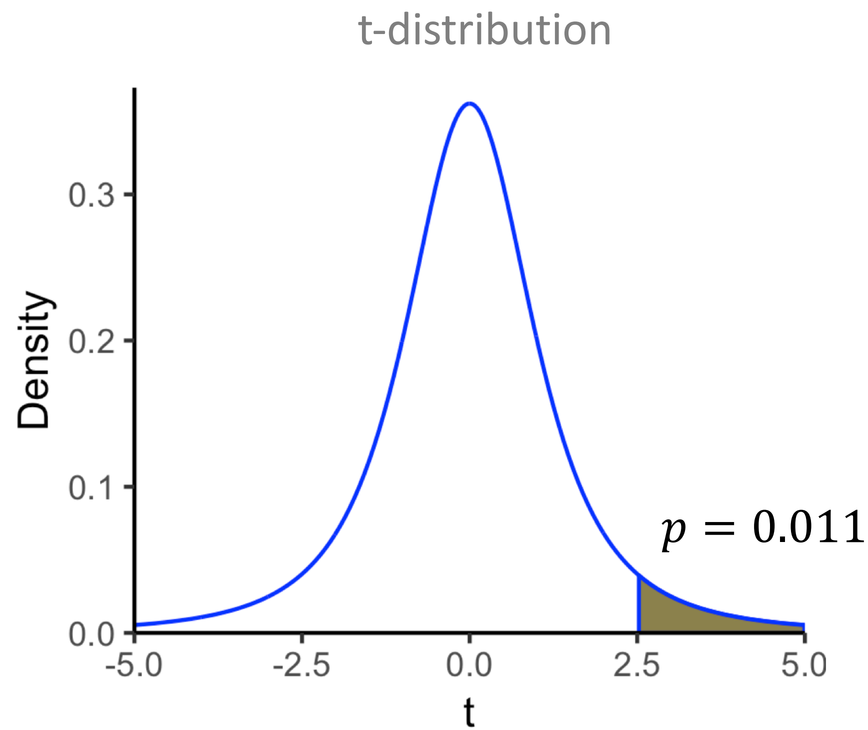
$SE = 1.96$ g

$\nu = 18.0$

$$t = \frac{23.99 - 19.04}{1.96} = 2.524$$

$p = 0.011$ (one-sided)

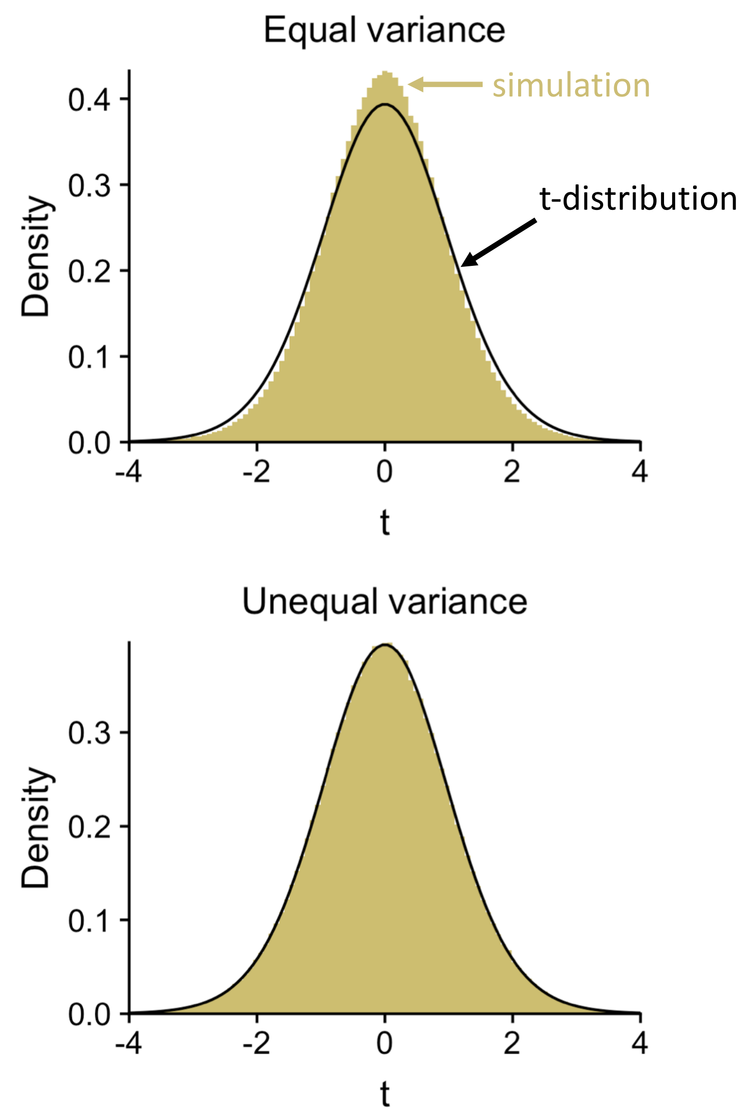
$p = 0.021$ (two-sided)




```
> 1 - pt(2.524, 18)
[1] 0.01061046
```

What if variances are not equal?

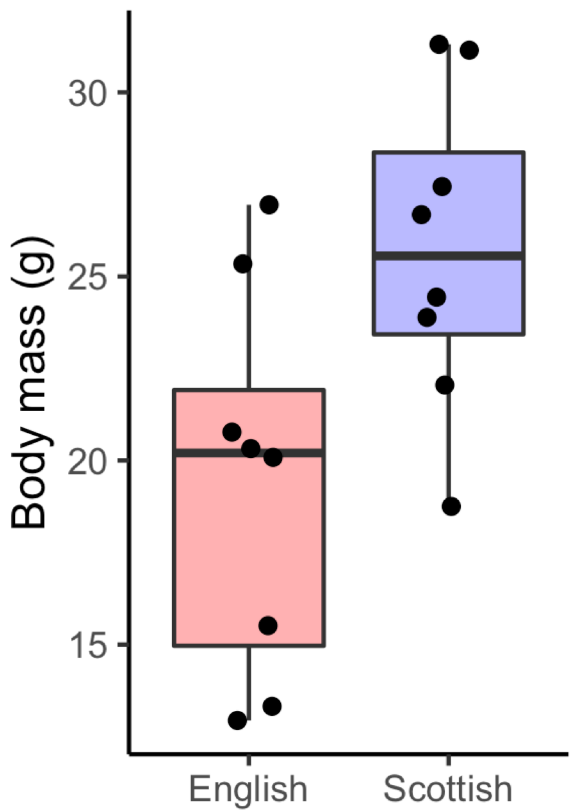
- Say, our samples come from two populations:
  - English: $\mu = 20\text{ g},\quad \sigma = 5\text{ g}$
  - Scottish: $\mu = 20\text{ g},\quad \sigma = 2.5\text{ g}$
- 'Equal variance' t-statistic does not represent the null hypothesis
- Unless you are certain that the variances are equal, use the Welch's test

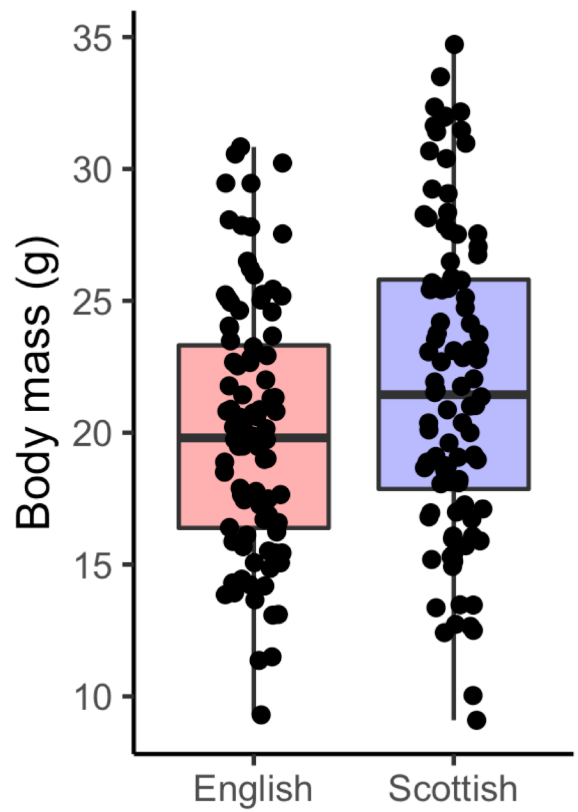


P-values vs. effect size

$n = 8$
$\Delta M = 6.3$ g
$p = 0.02$

$n = 100$
$\Delta M = 1.8$ g
$p = 0.02$





P-value is not a measure of biological significance

Two-sample t test: summary

Input	Two samples of n_1 and n_2 measurements
Assumptions	Observations are random and independent (no before/after data) Data are normally distributed
Usage	Compare sample means
Null hypothesis	Samples came from populations with the same means
Comments	Works well for non-normal distribution, as long as it is symmetric Two versions: equal and unequal variances; if unsure, use the unequal variance test

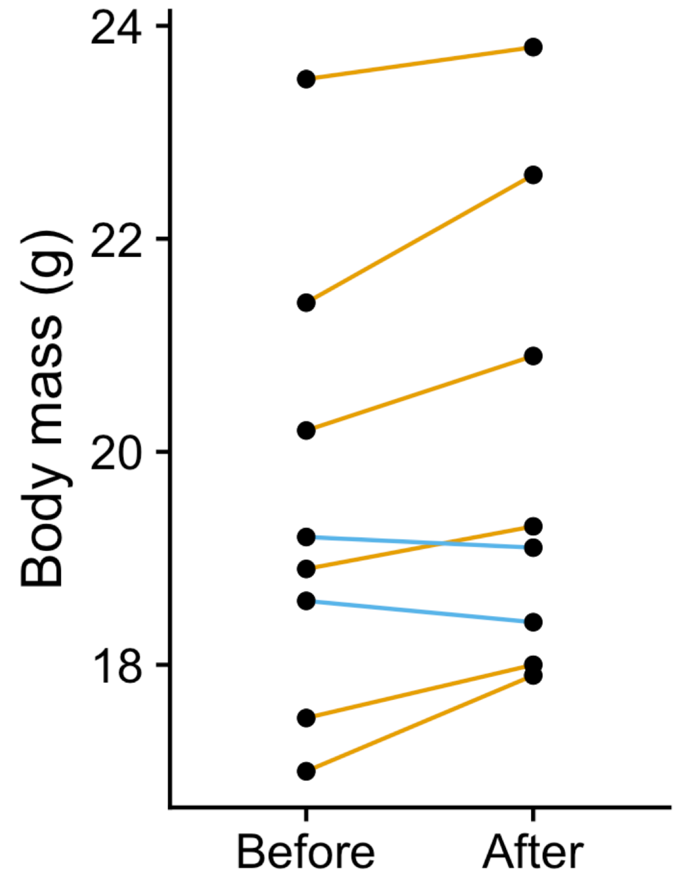
How to do it in R?

```
> English <- c(16.5, 21.3, 12.4, 11.2, 23.7, 20.2, 17.4, 23, 15.6, 26.5, 21.8, 18.9)
> Scottish <- c(19.7, 29.3, 27.1, 24.8, 22.4, 27.6, 25.7, 23.9, 15.4)
# One-sided t-test, equal variances
> t.test(Scottish, English, var.equal=TRUE, alternative="greater")

        Two Sample t-test

data:  Scottish and English
t = 2.4993, df = 19, p-value = 0.01089
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.524438      Inf
sample estimates:
mean of x mean of y
 23.98889  19.04167
# One-sided t-test, unequal variances
> t.test(Scottish, English, var.equal=FALSE, alternative="greater")

        Welch Two Sample t-test

data:  Scottish and English
t = 2.5238, df = 17.969, p-value = 0.01062
```

Paired t-test

Paired t-test

- Samples are paired
- For example: mouse weight before and after obesity treatment
- Null hypothesis: there is no difference between before and after
- M_Δ - the mean of the individual differences
- Example: mouse body mass (g)



Before:	21.4	20.2	23.5	17.5	18.6	17.0	18.9	19.2
After:	22.6	20.9	23.8	18.0	18.4	17.9	19.3	19.1

Paired t-test

- Samples are paired
- Find the differences:

$$\Delta_i = x_i - y_i$$

then

M_Δ - mean

SD_Δ - standard deviation

$SE_\Delta = SD_\Delta / \sqrt{n}$ - standard error

- The test statistic is

$$t = \frac{M_\Delta}{SE_\Delta}$$

- t-distribution with $n - 1$ degrees of freedom

Non-paired t-test (Welch)

$M_{\mathrm{after}} - M_{\mathrm{before}} = 0.46\ \mathrm{g}$

$SE = 1.08\ \mathrm{g}$

$t = 0.426$

$p = 0.34$

Paired test

$M_\Delta = 0.28\ \mathrm{g}$

$SE_\Delta = 0.17\ \mathrm{g}$

$t = 2.75$

$p = 0.014$

How to do it in R?

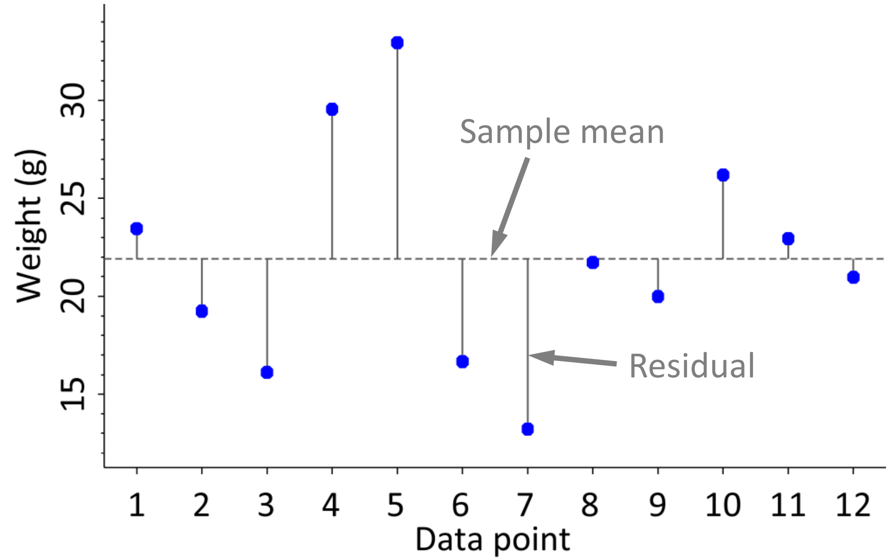
```
# Paired t-test
> before <- c(21.4, 20.2, 23.5, 17.5, 18.6, 17.0, 18.9, 19.2)
> after <- c(22.6, 20.9, 23.8, 18.0, 18.4, 17.9, 19.3, 19.1)
> t.test(after, before, paired=TRUE, alternative="greater")

        Paired t-test

data:  after and before
t = 2.7545, df = 7, p-value = 0.01416
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1443915       Inf
sample estimates:
mean of the differences
                 0.4625
```

F-test

Variance

- One sample of size n
- Sample variance

$$SD_{n-1}^2 = \frac{1}{n-1}\sum_i (x_i - M)^2$$



- Generalized variance: mean square

$$MS = \frac{SS}{\nu}$$

- where
  - SS - sum of squared residuals
  - ν - number of degrees of freedom

Comparison of variance

- Consider two samples
  - English mice, $n_E = 12$
  - Scottish mice $n_S = 9$

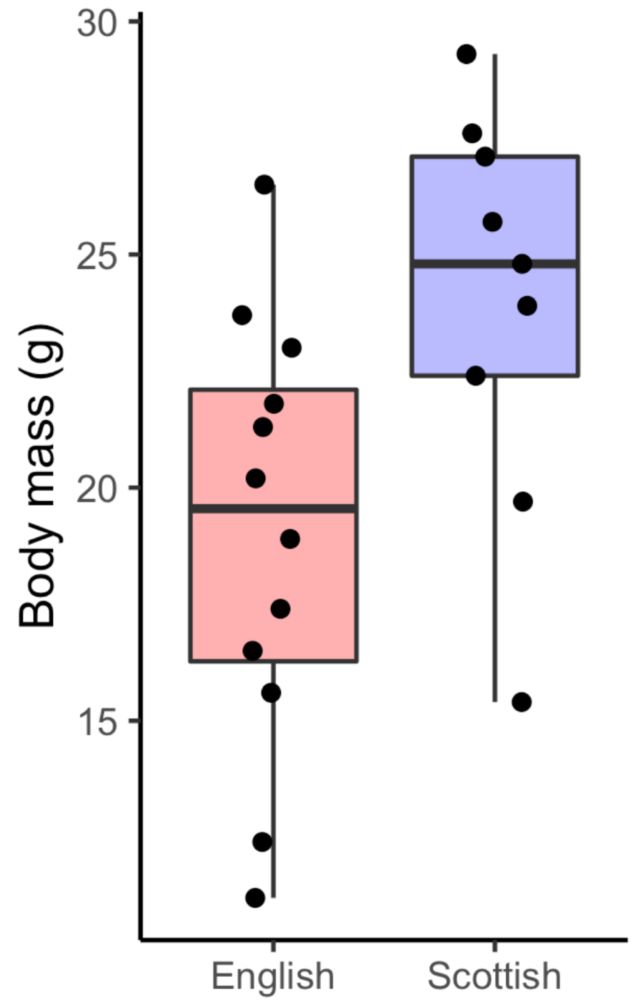
- We want to test if they come from the populations with the same variance, σ^2

- Null hypothesis: $\sigma_1^2 = \sigma_2^2$

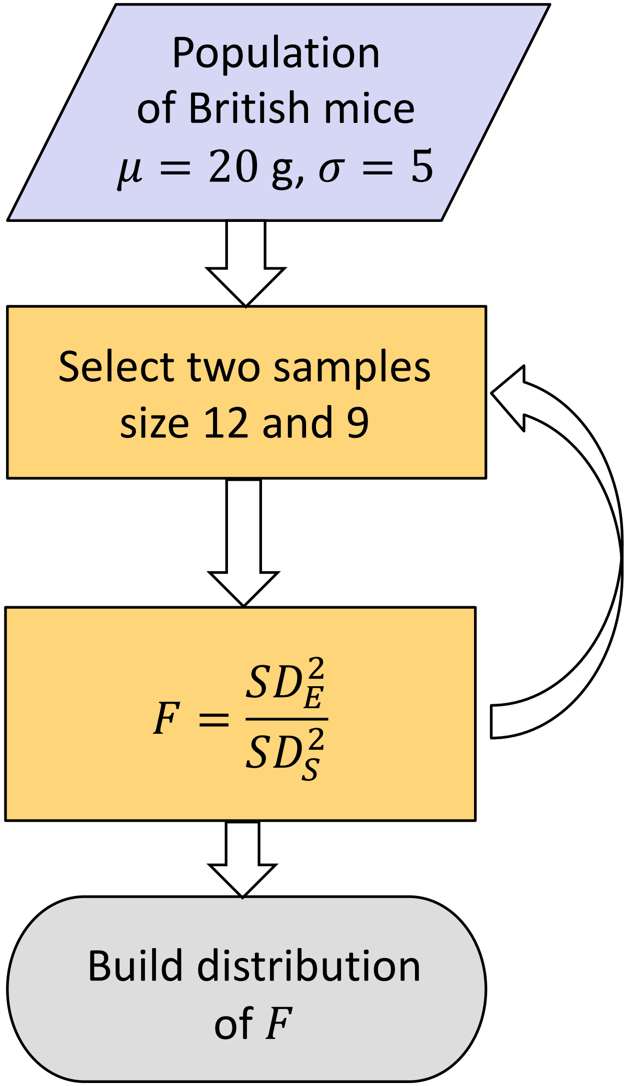
- We need a test statistic with known distribution

$n_E = 12$
$SD_E^2 = 21\ \mathrm{g}^2$

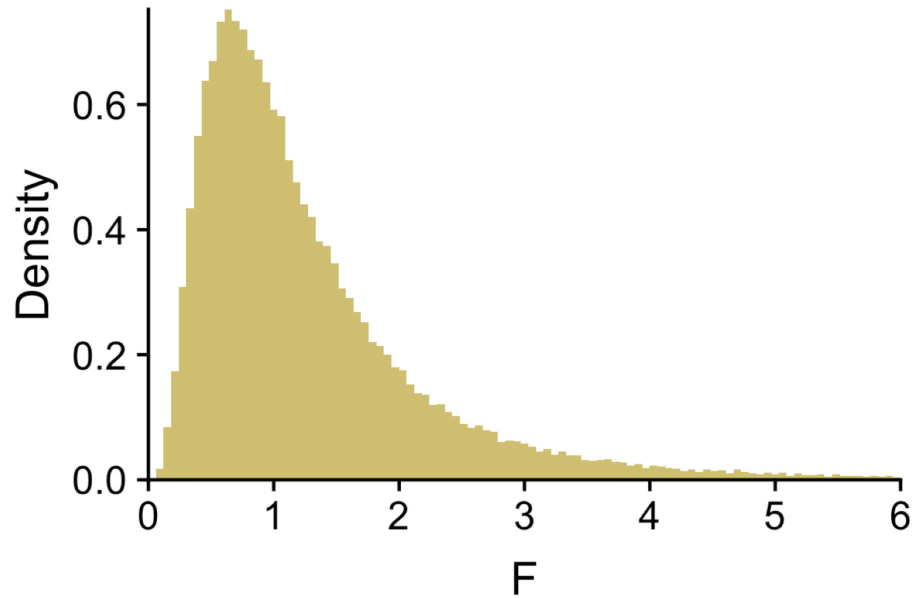
$n_S = 9$
$SD_S^2 = 19\ \mathrm{g}^2$



Gedankenexperiment



null distribution



Null distribution represents all random samples when the null hypothesis is true

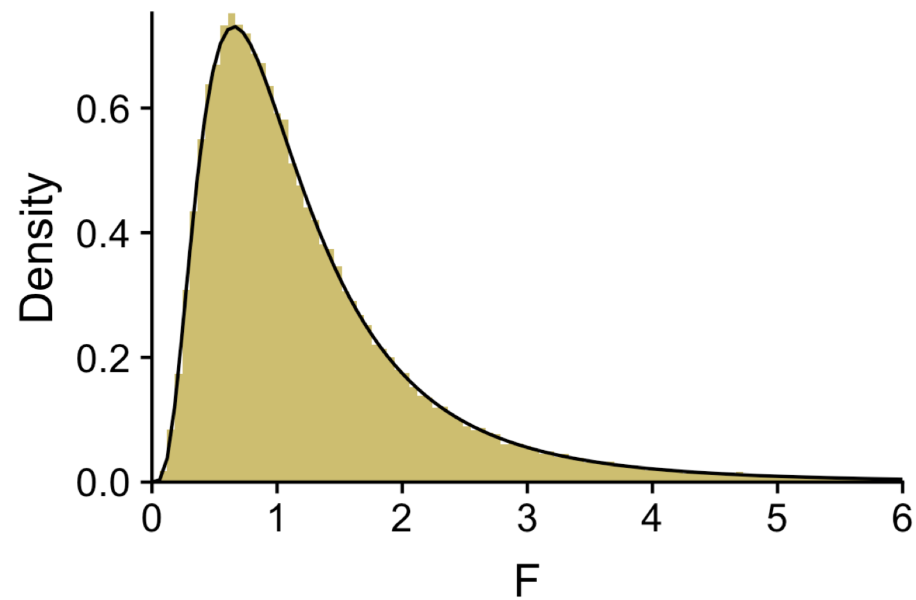
Test to compare two variances

- Consider two samples, sized n_1 and n_2
- Null hypothesis: they come from distributions with the same variance
- H_0: $\sigma_1^2 = \sigma_2^2$
- Test statistic:

$$F = \frac{SD_1^2}{SD_2^2}$$

is distributed with F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

F-distribution, $\nu_1 = 11, \nu_2 = 8$



Reminder
Test statistic for two-sample t-test:

$$t = \frac{M_1 - M_2}{SE}$$

Null distribution represents all random samples when the null hypothesis is true

F-test

- English mice: $SD_E = 4.61$ g, $n_E = 12$
- Scottish mice: $SD_S = 4.32$ g, $n_E = 9$

- Null hypothesis: they come from distributions with the same variance

- Test statistic:

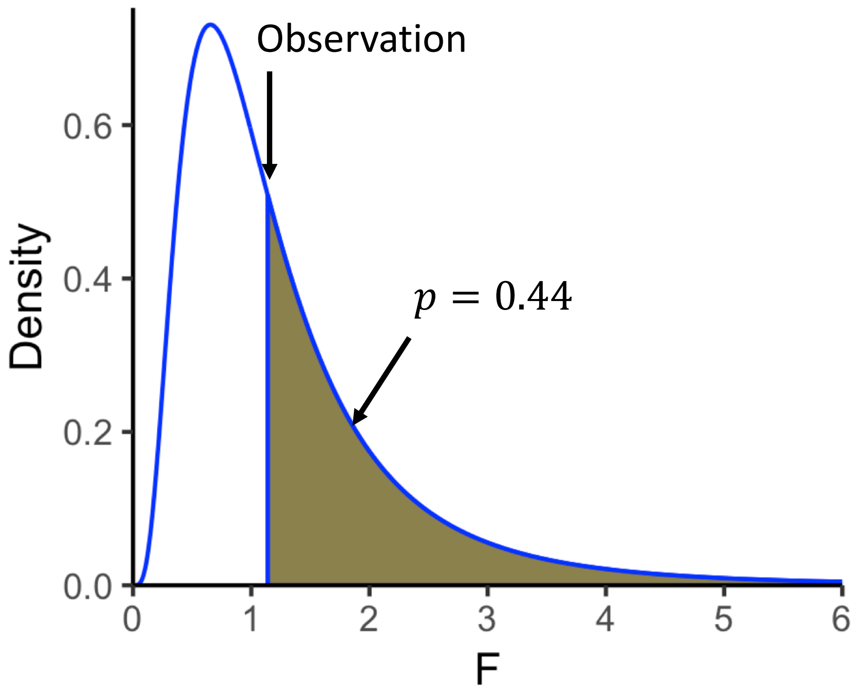
$$F = \frac{4.61^2}{4.32^2} = 1.139$$

$$\nu_E = 11$$

$$\nu_S = 8$$

$$p = 0.44$$

F-distribution, $\nu_1 = 11, \nu_2 = 8$




```
> 1 - pf(1.139, 11, 8)
[1] 0.4375845
```

Two-sample variance test (F-test): summary

Input	two samples of n_1 and n_2 measurements
Usage	compare sample variances
Null hypothesis	samples came from populations with the same variance
Comments	requires normality of data right now, it might look pointless, but is necessary in ANOVA. Very important test!

How to do it in R?

```
# Two-sample variance test
> var.test(English, Scottish, alternative="greater")

         F test to compare two variances

data:  English and Scottish
F = 1.1389, num df = 11, denom df = 8, p-value = 0.4376
alternative hypothesis: true ratio of variances is greater than 1
95 percent confidence interval:
 0.3437867       Inf
sample estimates:
ratio of variances
          1.138948
```

Hand-outs available at http://tiny.cc/statlec