8. t-test

“Aggregate statistics can sometimes mask important information”

Ben Bernanke
Statistical model

Null hypothesis
$H_0$: no effect

All other assumptions

Significance level
$\alpha = 0.05$

$\textbf{Statistical test against } H_0$

$p < \alpha$?

- yes
  - Reject $H_0$
  - (at your own risk)
  - Effect is real

- no
  - 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**

\[ t = \frac{M - \mu}{SE} \]

**Sample mean**  
**Test mean**  
**Standard error**

Generic form

\[ t = \frac{\text{deviation}}{\text{standard error}} \]

In both cases \( t = 2 \)
Reminder: Student’s t-distribution

- t-statistic is distributed with t-distribution
- Standardized
- One parameter: degrees of freedom, $\nu$
- For large $\nu$ approaches normal distribution

![Student’s t-distribution](image)
Null distribution for the deviation of the mean

Population of mice
$\mu = 20\, \text{g}, \sigma = 5$

Select sample size 5

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

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

Build distributions of $M$, $Z$ and $t$
Null distribution for the deviation of the mean
Null distribution for the deviation of the mean

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

\( \sigma \) - 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 $\mu$

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

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

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

![null distribution](null_distribution.png)

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

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

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

$p = 0.04$

```r
> 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

One-sided test
$H_1: M > \mu$

Two-sided test
$H_2: M \neq \mu$

$p_1 = 0.04$

$p_2 = 2p_1 = 0.08$

$p_2 = 2p_1$
Normality of data

Original distribution

Distribution of t
Statistical test vs confidence interval

**Confidence interval**

\[ n, M, SE \]

\[ n - 1 \text{ d.o.f.} \]

\[ t = \frac{M - \mu}{SE} \]

\[ tc <- qt(0.975, df = n - 1) \]

\[ \text{lower} <- M - tc \times SE \]

\[ \text{upper} <- M + tc \times SE \]

**t-test**

\[ t <- (M - 20) / SE \]

\[ p <- 2 \times (1 - pt(t, df = n-1)) \]
Statistical test vs confidence interval

When 95% CI touches $\mu$, then $p = 0.05$

```r
> t.test(x5, mu = 20)

One Sample t-test

data:  x5
t = 2.7766, df = 4, p-value = 0.04999
95 percent confidence interval:  
20.00035 34.81073
sample estimates:  
mean of x 27.40554
```
### One-sample t-test: summary

<table>
<thead>
<tr>
<th>Input</th>
<th>sample of $n$ measurements theoretical value $\mu$ (population mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>Observations are random and independent</td>
</tr>
<tr>
<td></td>
<td>Data are normally distributed</td>
</tr>
<tr>
<td>Usage</td>
<td>Examine if the sample is consistent with the population mean</td>
</tr>
<tr>
<td>Null hypothesis</td>
<td>Sample came from a population with mean $\mu$</td>
</tr>
<tr>
<td>Comments</td>
<td>Use for differences and ratios (e.g. SILAC)</td>
</tr>
<tr>
<td></td>
<td>Works well for non-normal distribution, as long as it is symmetric</td>
</tr>
</tbody>
</table>
How to do it in R?

# One-sided t-test

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

```r
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
Consider two samples (different sizes)

Are they different?

Are their means different?

Do they come from populations with different means?

\[ n_E = 12 \]
\[ M_E = 19.0 \text{ g} \]
\[ SD_E = 4.6 \text{ g} \]
\[ n_S = 9 \]
\[ M_S = 24.0 \text{ g} \]
\[ SD_S = 4.3 \text{ g} \]
Gedankenexperiment: null distribution

Population of British mice
\[ \mu = 20 \, \text{g}, \, \sigma = 5 \, \text{g} \]

Select two samples size 12 and 9

\[ t = \frac{M_E - M_S}{SE} \]

Build distribution of \( t \)

Normal population
\[ \mu = 20 \, \text{g}, \, \sigma = 5 \, \text{g} \]

Population of British mice
\[ \mu = 20 \, \text{g}, \, \sigma = 5 \, \text{g} \]

Select two samples size 12 and 9

\[ t = \frac{M_E - M_S}{SE} \]

Build distribution of \( t \)

Normal population
\[ \mu = 20 \, \text{g}, \, \sigma = 5 \, \text{g} \]
Null distribution

- *Gedankenexperiment*

- Test statistic

\[ t = \frac{M_1 - M_2}{SE} \]

is distributed with t-distribution with \( \nu \) 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 \( \nu \) degrees of freedom

Generic form

\[ t = \frac{\text{deviation}}{\text{standard error}} \]

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 $\nu$ degrees of freedom

- How do we find $SE$ and $\nu$ from two samples?

- Sample data:
  - English: $n_E = 12$, $M_E = 19.0$ g, $SD_E = 4.6$ g
  - Scottish: $n_S = 9$, $M_S = 24.0$ g, $SD_S = 4.3$ g
Case 1: equal variances – pool data together

- 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
  
  $$SE = SD_{1,2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
  
  $$\nu = n_1 + n_2 - 2$$
Case 1: equal variances, example

\[
\begin{align*}
  n_E &= 12 \\
  M_E &= 19.04 \text{ g} \\
  SD_E &= 4.61 \text{ g} \\
  n_S &= 9 \\
  M_S &= 23.99 \text{ g} \\
  SD_S &= 4.32 \text{ g}
\end{align*}
\]

\[SD_{1,2} = 4.49 \text{ g}\]

\[SE = 1.98 \text{ g}\]

\[v = 19\]

\[t = \frac{23.99 - 19.04}{1.98} = 2.499\]

\[p = 0.011 \text{ (one-sided)}\]

\[p = 0.022 \text{ (two-sided)}\]

\[> 1 - \text{pt}(2.499, 19)\]

[1] 0.01089314
Case 2: unequal variances - approximate

- Assume that distributions have different variances
- Welch’s t-test

- Find individual standard errors (squared):
  \[ SE_1^2 = \frac{SD_1^2}{n_1} \quad SE_2^2 = \frac{SD_2^2}{n_2} \]

- 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

\[ n_E = 12 \quad M_E = 19.04 \text{ g} \quad SD_E = 4.61 \text{ g} \]
\[ n_S = 9 \quad M_S = 23.99 \text{ g} \quad SD_S = 4.32 \text{ g} \]

\[ SE_E^2 = 1.77 \text{ g}^2 \]
\[ SE_S^2 = 2.07 \text{ g}^2 \]
\[ SE = 1.96 \text{ g} \]
\[ v = 18.0 \]
\[ t = \frac{23.99 - 19.04}{1.96} = 2.524 \]

\[ p = 0.011 \text{ (one-sided)} \]
\[ p = 0.021 \text{ (two-sided)} \]
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 variances’ $t$-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

\[ n = 8 \]
\[ \Delta M = 6.3 \text{ g} \]
\[ p = 0.02 \]

\[ n = 100 \]
\[ \Delta M = 1.8 \text{ g} \]
\[ p = 0.02 \]
P-value is not a measure of biological relevance
Overlapping 95% confidence intervals

If 95% CI don’t overlap, a two-sample t-test is highly significant
## Two-sample t test: summary

<table>
<thead>
<tr>
<th>Input</th>
<th>Two samples of $n_1$ and $n_2$ measurements</th>
</tr>
</thead>
</table>
| **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?

```r
> 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")

        Welch Two Sample t-test

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

```r
# One-sided t-test, unequal variances
> t.test(Scottish, English, var.equal = FALSE, 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
```
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

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

- Find the differences:
  \[ \Delta_i = x_i - y_i \]

- Then
  \[ M_\Delta \text{ - mean} \]
  \[ SD_\Delta \text{ - standard deviation} \]
  \[ SE_\Delta = \frac{SD_\Delta}{\sqrt{n}} \text{ - standard error} \]

- The test statistic is
  \[ t = \frac{M_\Delta}{SE_\Delta} \]

- t-distribution with \( n - 1 \) degrees of freedom

Paired test

One sample t-test against \( \mu = 0 \)
Paired t-test

**Paired test**

\[ M_\Delta = 0.28 \text{ g} \]
\[ SE_\Delta = 0.17 \text{ g} \]
\[ t = 2.75 \]
\[ p = 0.014 \]

**Non-paired t-test (Welch)**

\[ M_{\text{after}} - M_{\text{before}} = 0.46 \text{ g} \]
\[ SE = 1.08 \text{ g} \]
\[ t = 0.426 \]
\[ p = 0.34 \]
# Paired t-test

```r
> 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

```r
> t.test(after - before, mu = 0, alternative = "greater")
```

One Sample t-test

data:  after - before
t = 2.7545, df = 7, p-value = 0.01416
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
  - $\nu$ - 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, \( \sigma^2 \)

- Null hypothesis: \( \sigma_1^2 = \sigma_2^2 \)

- We need a test statistic with known distribution
Gedankenexperiment

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

Select two samples size 12 and 9

$$F = \frac{SD_E^2}{SD_S^2}$$

Build distribution of $F$

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, $\nu_1 = 11, \nu_2 = 8$

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

<table>
<thead>
<tr>
<th>Input</th>
<th>two samples of $n_1$ and $n_2$ measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>compare sample variances</td>
</tr>
<tr>
<td>Null hypothesis</td>
<td>samples came from populations with the same variance</td>
</tr>
<tr>
<td>Comments</td>
<td>requires normality of data right now, it might look pointless, but is necessary in ANOVA. Very important test!</td>
</tr>
</tbody>
</table>
How to do it in R?

```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
```
Slides available at
https://dag.compbio.dundee.ac.uk/training/Statistics_lectures.html