P-values and statistical tests
3. t-test

Marek Gierliński
Division of Computational Biology

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

Null hypothesis
\[ H_0: \text{no effect} \]

Significance level
\[ \alpha = 0.05 \]

Data

Statistic \( T \)

\( p \)-value

\[ p < \alpha \]
Reject \( H_0 \)

\[ p \geq \alpha \]
Insufficient evidence
One-sample t-test
Null hypothesis: the sample came from a population with mean $\mu = 20$ g
**t-statistic**

- Sample $x_1, x_2, ..., x_n$

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

- From these we can find

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

- more generic form:

  $t = \frac{\text{deviation}}{\text{standard error}}$
Student’s t-distribution

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

---

**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 the 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 tabulated, 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

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 \( t \)

\( \times 10^6 \)
Null distribution for the deviation of the mean

**Original distribution**

Gaussian ($\mu, \sigma$)

**Distribution of $Z$**

Gaussian (0, 1)

**Distribution of $M$**

Gaussian ($\mu, \sigma/\sqrt{n}$)

**Distribution of $t$**

- Gaussian (0, 1)
- $t$-distribution ($\nu$)
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
One-sample t-test: example

- **H₀**: \( \mu = 20 \) g

- 5 mice with body mass (g):
  - 19.5, 26.7, 24.5, 21.9, 22.0

\[
M = \frac{22.92 \text{ g}}{5} = 22.92 \text{ g}
\]
\[
SD = \frac{2.76 \text{ g}}{1.22} = 2.26 \text{ g}
\]
\[
SE = \frac{1.23 \text{ g}}{\sqrt{5}} = 1.23 \text{ g}
\]
\[
t = \frac{22.92 - 20}{1.22} = 2.37
\]
\[\nu = 4\]

\[p = 0.04\]
Normality of data

Original distribution

Distribution of $t$
Sidedness

One-sided test
\[ H_1: M > \mu \]

Observation
\[ p_1 = 0.04 \]

Two-sided test
\[ H_2: M \neq \mu \]

\[ p_2 = 2p_1 \]

\[ p_2 = 0.08 \]
# One-sample t-test: summary

<table>
<thead>
<tr>
<th>Input</th>
<th>sample of $n$ measurement 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>Limited usage (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
> 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?

\[ n_E = 12 \]
\[ M_E = 19.0 \text{ g} \]
\[ S_E = 4.6 \text{ g} \]

\[ n_S = 9 \]
\[ M_S = 24.0 \text{ g} \]
\[ S_S = 4.3 \text{ g} \]
The null distribution for the deviation between means

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 \( \Delta M \) and \( t \)

Normal population
\[ \mu = 20 \text{ g}, \sigma = 5 \text{ g} \]
Two-sample t-test

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

- Test statistic
  \[ t = \frac{M_1 - M_2}{SE} \]
  is distributed with t-distribution with $\nu$ degrees of freedom
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

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

\[
v = n_1 + n_2 - 2
\]

In case of equal samples sizes, \(n_1 = n_2\), these equations simplify:

\[
SD_{1,2}^2 = SD_1^2 + SD_2^2
\]

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

\[
v = 2n - 2
\]
Case 1: equal variances, example

\[ n_E = 12 \]
\[ M_E = 19.0 \text{ g} \]
\[ SD_E = 4.6 \text{ g} \]

\[ n_S = 9 \]
\[ M_S = 24.0 \text{ g} \]
\[ SD_S = 4.3 \text{ g} \]

\[ SD_{1,2} = 4.5 \text{ g} \]
\[ SE = 1.98 \text{ g} \]
\[ v = 19 \]
\[ t = 2.499 \]
\[ p = 0.011 \text{ (one-sided)} \]
\[ p = 0.022 \text{ (two-sided)} \]
Case 2: unequal variances

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

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

- 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

\[
\begin{array}{ll}
n_E = 12 & n_S = 9 \\
M_E = 19.0 \text{ g} & M_S = 24.0 \text{ g} \\
SD_E = 4.6 \text{ g} & SD_S = 4.3 \text{ g} \\
\end{array}
\]

\[
\begin{align*}
SE_E^2 &= 1.8 \text{ g}^2 \\
SE_S^2 &= 2.1 \text{ g}^2 \\
SE &= 1.96 \text{ g} \\
\nu &= 18 \\
t &= 2.524 \\
p &= 0.011 \text{ (one-sided)} \\
p &= 0.021 \text{ (two-sided)}
\end{align*}
\]
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

- t-test on samples:
  - ‘equal’: $t = 4.17, \quad p = 2.8 \times 10^{-4}$
  - ‘unequal’: $t = 4.56, \quad p = 1.4 \times 10^{-4}$

- Unless you are certain that the variances are equal, use the Welch’s test
P-values vs. effect size

\[ n_E = 12 \quad n_S = 9 \]
\[ \log FC = 0.33 \]
\[ p = 0.022 \]

\[ n_E = 100 \quad n_S = 100 \]
\[ \log FC = 0.11 \]
\[ p = 0.022 \]
P-value is not a measure of biological significance
# 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 |
How to do it in R?

```r
# Two-sided t-test, equal variances
> Scottish = c(19.7, 29.3, 27.1, 24.8, 22.4, 27.6, 25.7, 23.9, 15.4)
> t.test(English, Scottish, var.equal=T)

Two Sample t-test

data:  English and Scottish
t = -2.4993, df = 19, p-value = 0.02177
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  -9.0903223  -0.8041221
sample estimates:
mean of x  mean of y
  19.04167  23.98889

# Two-sided t-test, unequal variances
> t.test(English, Scottish, var.equal=F)

Welch Two Sample t-test

data:  English and Scottish
t = -2.5238, df = 17.969, p-value = 0.02125
```
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

- $M_\Delta$ - the mean of the individual differences

- Example: mouse body mass (g)

<table>
<thead>
<tr>
<th>Before</th>
<th>21.4</th>
<th>20.2</th>
<th>23.5</th>
<th>17.5</th>
<th>18.6</th>
<th>17.0</th>
<th>18.9</th>
<th>19.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>22.6</td>
<td>20.9</td>
<td>23.8</td>
<td>18.0</td>
<td>18.4</td>
<td>17.9</td>
<td>19.3</td>
<td>19.1</td>
</tr>
</tbody>
</table>
Paired t-test

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

then

- \( M_\Delta \): mean
- \( SD_\Delta \): 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)

- \( M_a - M_b = 0.46 \text{ g} \)
- \( SE = 1.08 \text{ g} \)
- \( t = 0.426 \)
- \( p = 0.34 \)

Paired test

- \( M_\Delta = 0.28 \text{ g} \)
- \( SE_\Delta = 0.17 \text{ g} \)
- \( t = 2.75 \)
- \( p = 0.03 \)
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(before, after, paired=T)

Paired t-test

data:  before and after
t = -2.7545, df = 7, p-value = 0.02832
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  
-0.85953136 -0.06546864
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
  - $\nu$ - 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

<table>
<thead>
<tr>
<th></th>
<th>( n_E = 12 )</th>
<th>( n_S = 9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( SD_E^2 )</td>
<td>21 ( g^2 )</td>
<td>19 ( g^2 )</td>
</tr>
</tbody>
</table>

Body mass (g)
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

Normalized frequency

0.0
0.2
0.4
0.6
0.8

0
2
4
6

\( F \)
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

\[ \text{F-distribution, } \nu_1 = 11, \nu_2 = 8 \]

---

**Remainder**

Test statistic for two-sample t-test:

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

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

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

- Test statistic:
  \[
  F = \frac{4.61^2}{4.32^2} = 1.139
  \]
  \[
  \nu_E = 11
  \]
  \[
  \nu_S = 8
  \]
  \[
  p = 0.44
  \]
## 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><strong>Usage</strong></td>
<td>compare sample variances</td>
</tr>
<tr>
<td><strong>Null hypothesis</strong></td>
<td>samples came from populations with the same variance</td>
</tr>
<tr>
<td><strong>Comments</strong></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?

# 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 to 1
95 percent confidence interval:
  0.3437867 Inf
sample estimates:
  ratio of variances
  1.138948
Hand-outs available at http://tiny.cc/statlec