Marek Gągolewski
Konstancja Bobecka-Wesołowska
Przemysław Grzegorzewski

Computer Statistics with R

7. Parametric Tests

Faculty of Mathematics and Information Science
Warsaw University of Technology
[Last update: December 9, 2012]

Contents

Info

These tutorials are likely to contain bugs and typos. In case you find any don't hesitate to *contact us*! Thanks in advance!

7.1. Preliminaries

Notation convention — quantiles:

- $z_\alpha = \texttt{qnorm}(\alpha)$,
- $t_\alpha^{[n]} = \texttt{qt}(\alpha, n)$,
- $\chi^2_{\alpha,n} = \texttt{qchisq}(\alpha, n)$,
- $F_\alpha^{[n_1,n_2]} = \texttt{qf}(\alpha, n_1, n_2)$.

7.1.1. Theory of hypothesis tests

Let $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{p_\theta, \theta \in \Theta\})$ be a identifiable statistical space. Consider the pair (H_0, K) of hypotheses, where

$$\begin{array}{lll} H_0: & \theta \in \Theta_{H_0} & \text{(null hypothesis)}, \\ K: & \theta \in \Theta_K & \text{(alternative hypothesis)}, \end{array} \tag{7.1}$$

$\Theta_{H_0} \subset \Theta$, $\Theta_K \subset \Theta$, $\Theta_{H_0} \cap \Theta_K = \emptyset$, $\Theta_{H_0} \cup \Theta_K = \Theta$.

Definition 1. *A statistical hypothesis test* is a function $\varphi : \mathcal{X} \to \{0, 1\}$, where 0 denotes the acceptation of the null hypothesis, and 1 — its rejection.

Definition 2. *An error of I kind* is a rejection of true null hypothesis.

Definition 3. *An error of II kind* is an acceptation of false null hypothesis.

Let us denote by α_φ and β_φ the probability of committing, respectively, an error of I and II kind, that is

$$\begin{array}{rcl} \alpha_\varphi & = & \mathrm{P}(\varphi = 1 | H_0), \\ \beta_\varphi & = & \mathrm{P}(\varphi = 0 | K). \end{array} \tag{7.2}$$

Definition 4. *The power function* of a test φ is a mapping $M : \Theta \to [0, 1]$:

$$M_\varphi(\theta) = \mathrm{P}_\theta(\varphi = 1). \tag{7.3}$$

It is easily seen that $\mathbb{E}_\theta\, \varphi = M_\varphi(\theta)$.

Definition 5. We say that a test φ has *significance level of α*, if $\mathbb{E}_\theta\, \varphi \le \alpha$ for all $\theta \in \Theta_{H_0}$.

Note that in practice we often set $\alpha = 0.05, 0.01$, or 0.1.

Definition 6. *The size* of a test φ is defined as

$$\mathrm{size}(\varphi) := \sup_{\theta \in \Theta_{H_0}} \mathbb{E}_\theta\, \varphi. \tag{7.4}$$

7.1.1.1. Testing statistical hypotheses in practice

Here is an algorithm for testing statistical hypotheses using computer programs such as R:

1. Specify statistical space (model).
2. Gather information on observed sample.
3. Specify a null hypothesis H_0 and an alternative hypothesis K.
4. Choose appropriate test.
5. Fix significance level α.
6. Calculate test statistic T.

7. Calculate p-value.

8. Make an appropriate decision: reject H_0 if p-value $< \alpha$, and do not reject H_0 otherwise.

If no computer program is available we may:

7'. Determine critical region R_α.

8'. Make an appropriate decision: reject H_0 if $T \in R_\alpha$, and do not reject H_0 otherwise.

7.1.2. Tests for mean μ

7.1.2.1. One-sample z-test

Model I $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu, \sigma)$, known σ. Null hypothesis: $H_0 : \mu = \mu_0$. Test statistic:

$$z = \frac{\overline{\mathbf{X}} - \mu_0}{\sigma}\sqrt{n} \quad \overset{H_0}{\sim} \quad \mathrm{N}(0,1). \tag{7.5}$$

R implementation: none.

Model II $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. ???, large n. Null hypothesis: $H_0 : \mu = \mu_0$. Test statistic (asymptotic):

$$z = \frac{\overline{\mathbf{X}} - \mu_0}{s_{\mathbf{X}}}\sqrt{n} \quad \overset{H_0}{\sim}_{n\to\infty} \mathrm{N}(0,1). \tag{7.6}$$

R implementation: `t.test(..., mu=...)`.

Alternative hypothesis:	Critical region:
$K_1 : \mu \neq \mu_0$	$R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$
$K_2 : \mu > \mu_0$	$R_2(\alpha) = [z_{1-\alpha}, +\infty)$
$K_3 : \mu < \mu_0$	$R_3(\alpha) = (-\infty, -z_{1-\alpha}]$

The type of alternative in `t.test()` is determined by `alternative="two.sided"` (K_1, default), `"greater"` (K_2) or `"less"` (K_3).

For example, given a numeric vector `x` of observed sample values,
`t.test(x, mu=1)` — tests $H_0 : \mu = 1$ versus $K : \mu \neq 1$.
`t.test(x, mu=1, alternative="less")` — tests $H_0 : \mu = 1$ versus $K : \mu < 1$.

7.1.2.2. One-sample t-test

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu, \sigma)$, unknown σ. Null hypothesis: $H_0 : \mu = \mu_0$. Test statistic:

$$t = \frac{\overline{\mathbf{X}} - \mu_0}{s_{\mathbf{X}}}\sqrt{n} \quad \overset{H_0}{\sim} \quad \mathrm{t}^{[n-1]}. \tag{7.7}$$

R implementation: `t.test(..., mu=...)`.

Alternative hypothesis:	Critical region:
$K_1 : \mu \neq \mu_0$	$R_1(\alpha) = \left(-\infty, -t^{[n-1]}_{1-0.5\alpha}\right] \cup \left[t^{[n-1]}_{1-0.5\alpha}, +\infty\right)$
$K_2 : \mu > \mu_0$	$R_2(\alpha) = \left[t^{[n-1]}_{1-\alpha}, +\infty\right)$
$K_3 : \mu < \mu_0$	$R_3(\alpha) = \left(-\infty, -t^{[n-1]}_{1-\alpha}\right]$

7.1.2.3. Two-sample z-test

Model I $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. $\mathrm{N}(\mu_X, \sigma_X)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. $\mathrm{N}(\mu_Y, \sigma_Y)$, known σ_X and σ_Y. Null hypothesis: $H_0 : \mu_X = \mu_Y$. Test statistic:

$$z = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{H_0}{\sim} \mathrm{N}(0, 1). \tag{7.8}$$

R implementation: none.

Model II $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. ???, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. ???, large n_X and n_Y. Null hypothesis: $H_0 : \mu_X = \mu_Y$. Test statistic (asymptotic):

$$z = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\frac{s_{\mathbf{X}}^2}{n_X} + \frac{s_{\mathbf{Y}}^2}{n_Y}}} \overset{H_0}{\sim}_{n_X, n_Y \to \infty} \mathrm{N}(0, 1). \tag{7.9}$$

R implementation: `t.test()`.

Alternative hypothesis:	Critical region:
$K_1 : \mu_X \neq \mu_Y$	$R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$
$K_2 : \mu_X > \mu_Y$	$R_2(\alpha) = [z_{1-\alpha}, +\infty)$
$K_3 : \mu_X < \mu_Y$	$R_3(\alpha) = (-\infty, -z_{1-\alpha}]$

7.1.2.4. Two-sample t-test

Model I $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. $\mathrm{N}(\mu_X, \sigma_X)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. $\mathrm{N}(\mu_Y, \sigma_Y)$, unknown σ_X and σ_Y, but $\sigma_X = \sigma_Y$ (pooled variance version). Null hypothesis: $H_0 : \mu_X = \mu_Y$. Let

$$d = n_X + n_Y - 2.$$

Test statistic:

$$t = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\frac{\mathring{s}^2}{n_X} + \frac{\mathring{s}^2}{n_Y}}} \overset{H_0}{\sim} \mathrm{t}^{[d]}. \tag{7.10}$$

where

$$\mathring{s}^2 = \frac{(n_X - 1)s_{\mathbf{X}}^2 + (n_Y - 1)s_{\mathbf{Y}}^2}{d}.$$

R implementation: `t.test(..., var.equal=T)`.

Model II $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. $\mathrm{N}(\mu_X, \sigma_X)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. $\mathrm{N}(\mu_Y, \sigma_Y)$, unknown σ_X and σ_Y. Null hypothesis: $H_0 : \mu_X = \mu_Y$. The Welch-Satterthwaite approximation for the degrees of freedom:

$$d = \frac{\left(\frac{s_{\mathbf{X}}^2}{n_X} + \frac{s_{\mathbf{Y}}^2}{n_Y}\right)^2}{\left(\frac{s_{\mathbf{X}}^2}{n_X}\right)^2 / (n_X - 1) + \left(\frac{s_{\mathbf{Y}}^2}{n_Y}\right)^2 / (n_Y - 1)}.$$

Test statistic (approximate):

$$t = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\frac{s_{\mathbf{X}}^2}{n_X} + \frac{s_{\mathbf{Y}}^2}{n_Y}}} \overset{H_0}{\sim} \mathrm{t}^{[d]}. \tag{7.11}$$

R implementation: `t.test()`.

Alternative hypothesis:	Critical region:
$K_1 : \mu_X \neq \mu_Y$	$R_1(\alpha) = \left(-\infty, -t^{[d]}_{1-0.5\alpha}\right] \cup \left[t^{[d]}_{1-0.5\alpha}, +\infty\right)$
$K_2 : \mu_X > \mu_Y$	$R_2(\alpha) = \left[t^{[d]}_{1-\alpha}, +\infty\right)$
$K_3 : \mu_X < \mu_Y$	$R_3(\alpha) = \left(-\infty, -t^{[d]}_{1-\alpha}\right]$

For example, given two vectors of observations `x` and `y`,
`t.test(x,y)` — tests $H_0 : \mu_X = \mu_Y$ versus $K : \mu_X \neq \mu_Y$.
`t.test(x,y,alternative="less",var.equal=TRUE)` — tests $H_0 : \mu_X = \mu_Y$ versus $K : \mu_X < \mu_Y$, assuming $\sigma_1^X = \sigma_2^Y$.

7.1.2.5. Paired t-test

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu_X, \sigma_X)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ i.i.d. $\mathrm{N}(\mu_Y, \sigma_Y)$, unknown σ_X and σ_Y, but observations are matched in pairs, i.e. (X_i, Y_i) are dependent for $i = 1, 2, \ldots, n$. Null hypothesis: $H_0 : \mu_X = \mu_Y$. Let $\mathbf{Z} = \mathbf{X} - \mathbf{Y}$. Test statistic:

$$t = \frac{\overline{\mathbf{Z}} - 0}{s_{\mathbf{Z}}} \sqrt{n} \overset{H_0}{\sim} \mathrm{t}^{[n-1]}. \tag{7.12}$$

R implementation: `t.test(..., paired=T)`.

Alternative hypothesis:	Critical region:
$K_1 : \mu_X \neq \mu_Y$	$R_1(\alpha) = \left(-\infty, -t^{[n-1]}_{1-0.5\alpha}\right] \cup \left[t^{[n-1]}_{1-0.5\alpha}, +\infty\right)$
$K_2 : \mu_X > \mu_Y$	$R_2(\alpha) = \left[t^{[n-1]}_{1-\alpha}, +\infty\right)$
$K_3 : \mu_X < \mu_Y$	$R_3(\alpha) = \left(-\infty, -t^{[n-1]}_{1-\alpha}\right]$

7.1.3. Tests for variance σ^2

7.1.3.1. One-sample test for variance

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu, \sigma)$, unknown μ. Null hypothesis: $H_0 : \sigma^2 = \sigma_0^2$. Test statistic:

$$x = \frac{(n-1)s^2_{\mathbf{X}}}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{n-1}. \tag{7.13}$$

R implementation: none.

Alternative hypothesis:	Critical region:
$K_1 : \sigma^2 \neq \sigma_0^2$	$R_1(\alpha) = \left(0, \chi^2_{0.5\alpha,n-1}\right] \cup \left[\chi^2_{1-0.5\alpha,n-1}, +\infty\right)$
$K_2 : \sigma^2 > \sigma_0^2$	$R_2(\alpha) = \left[\chi^2_{1-\alpha,n-1}, +\infty\right)$
$K_3 : \sigma^2 < \sigma_0^2$	$R_3(\alpha) = \left(0, \chi^2_{\alpha,n-1}\right]$

7.1.3.2. F-test

$\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. $\mathrm{N}(\mu_X, \sigma_X)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. $\mathrm{N}(\mu_Y, \sigma_Y)$, unknown μ_X and μ_Y. Null hypothesis: $H_0 : \sigma_X^2 = \sigma_Y^2$. Test statistic:

$$F = \frac{s^2_{\mathbf{X}}}{s^2_{\mathbf{Y}}} \overset{H_0}{\sim} \mathrm{F}^{[n_1-1, n_2-2]}. \tag{7.14}$$

R implementation: `var.test()`.

Alternative hypothesis: $K : \sigma_X^2 > \sigma_Y^2$. $R = \left[F^{[n_1-1,n_2-1]}_{1-\alpha}, +\infty\right)$.

See also: Bartlett's test, Levene's test.

7.1.4. Tests for proportion

7.1.4.1. One-sample z-test

Model I $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. Bern(p), large n (≥ 100). Null hypothesis: $H_0 : p = p_0$. Test statistic (asymptotic):

$$z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)}} \sqrt{n} \quad \overset{H_0}{\sim}_{n \to \infty} \mathrm{N}(0, 1). \tag{7.15}$$

Model II $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. Bern(p), small n. Null hypothesis: $H_0 : p = p_0$. Test statistic (approximate):

$$z = 2\left(\arcsin\sqrt{\widehat{p}} - \arcsin\sqrt{p_0}\right)\sqrt{n} \quad \overset{H_0}{\sim} \quad \mathrm{N}(0, 1). \tag{7.16}$$

Alternative hypothesis:	Critical region:
$K_1 : p \neq p_0$	$R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$
$K_2 : p > p_0$	$R_2(\alpha) = [z_{1-\alpha}, +\infty)$
$K_3 : p < p_0$	$R_3(\alpha) = (-\infty, -z_{1-\alpha}]$

R implementation (similar tests, more accurate): `binom.test()`, `prop.test()`.

7.1.4.2. Two-sample z-test

Model I $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. Bern(p_X), $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. Bern(p_Y), large n_X and n_Y (≥ 100). Null hypothesis: $H_0 : p_X = p_Y$. Let $k_X = \sum_{i=1}^{n_X} X_i$, $k_Y = \sum_{i=1}^{n_Y} Y_i$. Test statistic (asymptotic):

$$z = \frac{\widehat{p}_X - \widehat{p}_Y}{\sqrt{\mathring{p}(1 - \mathring{p})}} \sqrt{\mathring{n}} \quad \overset{H_0}{\sim}_{n_X, n_Y \to \infty} \quad \mathrm{N}(0, 1), \tag{7.17}$$

where $\mathring{p} = \frac{k_X + k_Y}{n_X + n_Y}$ and $\mathring{n} = \frac{n_X n_Y}{n_X + n_Y}$.

Model II $\mathbf{X} = (X_1, X_2, \ldots, X_{n_X})$ i.i.d. Bern(p_X), $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ i.i.d. Bern(p_Y), small n_X and n_Y. Null hypothesis: $H_0 : p_X = p_Y$. Let $k_X = \sum_{i=1}^{n_X} X_i$, $k_Y = \sum_{i=1}^{n_Y} Y_i$. Test statistic (approximate):

$$z = 2\left(\arcsin\sqrt{\widehat{p}_X} - \arcsin\sqrt{\widehat{p}_Y}\right)\sqrt{\mathring{n}} \quad \overset{H_0}{\sim} \quad \mathrm{N}(0, 1), \tag{7.18}$$

where $\mathring{n} = \frac{n_X n_Y}{n_X + n_Y}$.

Alternative hypothesis:	Critical region:
$K_1 : p_X \neq p_Y$	$R_1(\alpha) = (-\infty, -z_{1-0.5\alpha}] \cup [z_{1-0.5\alpha}, +\infty)$
$K_2 : p_X > p_Y$	$R_2(\alpha) = [z_{1-\alpha}, +\infty)$
$K_3 : p_X < p_Y$	$R_3(\alpha) = (-\infty, -z_{1-\alpha}]$

R implementation (more accurate): `prop.test()`.

7.2. Examples

Ex. 7.1. The internal pressure strength of glass bottles is an important characteristic of their quality. The bottler wants to know whether the expected pressure strength exceeds

1.20 N/mm^2. From previous experience, he knows that the pressure strength is normally distributed with standard deviation of 0.07 N/mm^2.

A random sample of 20 bottles is selected and the bottles are placed on a hydrostatic pressure-testing machine that increases the pressure in the bottle until it breaks. Here are the results [N/mm^2]:

> 1.36, 1.14, 1.27, 1.15, 1.20, 1.29, 1.27, 1.18, 1.23, 1.36, 1.38, 1.37, 1.30, 1.21, 1.33, 1.28, 1.32, 1.29, 1.33, 1.25.

Test the hypothesis of interest at significance level of $\alpha = 0.04$.

Solution.
Our statistical model in this task is as follows. We are given a realization of a sample $(X_1, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu, \sigma)$ with unknown μ and known $\sigma = 0{,}07$.

We shall use the one-sample z-test to verify $H_0 : \mu = \mu_0$ against $K : \mu > \mu_0$. This tool is not available in R, so we will perform the calculations "manually".

The null hypothesis H_0 should be rejected, if the test statistic

$$T = \frac{\overline{\mathbf{X}} - \mu_0}{\sigma} \sqrt{n}. \tag{7.19}$$

falls into the critical region (a.k.a. rejection region) $R_\alpha = [z_{1-\alpha}, +\infty)$, where $z_{1-\alpha}$ denotes the $(1 - \alpha)$-standardized normal quantile.

Here we have: $n = 20$, $\alpha = 0.04$ and $\mu_0 = 1.2$.

Let us calculate the test statistic:

```
x <- c(1.36, 1.14, 1.27, 1.15, 1.2, 1.29, 1.27, 1.18, 1.23, 1.36, 1.38, 1.37,
    1.3, 1.21, 1.33, 1.28, 1.32, 1.29, 1.33, 1.25)
mu0 <- 1.2
sigma <- 0.07
(T <- (mean(x) - mu0)/sigma * sqrt(length(x)))
## [1] 4.824
```

The quantile is equal to:

```
qnorm(0.96)
## [1] 1.751
```

We note that the value of the test statistic T falls into the rejection region $R_{0.04} = [1.7507, +\infty)$, therefore the null hypothesis should be rejected. Thus, the expected bottle pressure strength significantly exceeds (at significance level $\alpha = 0.04$) 1.2 N/mm^2.

$\boxdot$

Ex. 7.2. Coffee cans in some factory are filled (at least theoretically) to a net weight of 150 g, but some variability is observed. In fact, a random sample of seven cans of this particular brand gave net weights of 142, 151, 148, 151, 145, 150, 141. Assuming normality of cans' net weights, test whether expected value of the weight is actually 150 g, with maximal probability of making type I error equal to 0.05.

Solution.
We are given a realization of a sample $(X_1, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu, \sigma)$ with two unknown parameters μ and σ. Here, one-sample t-test should be used to verify the null hypothesis $H_0 : \mu = 150$ against $K : \mu \neq 150$.

We fix significance level $\alpha = 0.05$.

```r
coffee <- c(142, 151, 148, 151, 145, 150, 141)
mean(coffee)  # let's check the mean weight
```

```
## [1] 146.9
```

```r
t.test(coffee, mu = 150)
```

```
##
##  One Sample t-test
##
## data:  coffee
## t = -1.97, df = 6, p-value = 0.0963
## alternative hypothesis: true mean is not equal to 150
## 95 percent confidence interval:
##  143.0 150.8
## sample estimates:
## mean of x
##     146.9
```

The value of the test statistic is $T = -1.97$. We compare the p-value $= 0.0963$ to the assumed significance level $\alpha = 0.05$. As p-value $> \alpha$, therefore we do not reject H_0. Thus, the expected net weight does not differ significantly (at $\alpha = 0{,}05$) from 150 g.

⊡

Ex. 7.3. 7 randomly selected academic teachers and 15 office workers were asked about for how many years have they been employed at WUT. The following results were obtained:

Teachers	$14, 17, 7, 33, 2, 24, 26, 22, 12$
Office workers	$13, 15, 3, 2, 25, 4, 1, 18, 6, 9, 20, 11, 5, 1, 7$

It is assumed that both distributions are normal. Verify a hypothesis ($\alpha = 0{,}05$) that the average seniority of the teachers is significantly greater than that of the office workers.

Solution.
We are given two independent random samples: $(X_1, \ldots, X_{n_1})$ i.i.d. $\mathrm{N}(\mu_1, \sigma_1)$, and $(Y_1, \ldots, Y_{n_2})$ i.i.d. $\mathrm{N}(\mu_2, \sigma_2)$.

We shall verify $H_0 : \mu_1 = \mu_2$ against $K : \mu_1 > \mu_2$ using two-sample t-test at significance level $\alpha = 0.05$. As the t-test has two variants, first we check whether the sample variances are equal. The F-test verifies $H_0' : \sigma_1^2 = \sigma_2^2$ against $K' : \sigma_1^2 \neq \sigma_2^2$:

```r
academ <- c(14, 17, 7, 33, 2, 24, 26, 22, 12)
office <- c(13, 15, 3, 2, 25, 4, 1, 18, 6, 9, 20, 11, 5, 1, 7)
var.test(academ, office)
```

```
##
##  F test to compare two variances
##
## data:  academ and office
## F = 1.725, num df = 8, denom df = 14, p-value = 0.3557
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5251 7.1239
## sample estimates:
## ratio of variances
##              1.725
```

Therefore, we may assume that $\sigma_1^2 = \sigma_2^2$, because p-value= ... $> \alpha$.

Note

In research reports, we should always also provide the reader with the value of test statistic $F = ...$

Let us now perform the t-test:

```
t.test(academ, office, alternative = "greater", var.equal = T)
##
##  Two Sample t-test
##
## data:  academ and office
## t = 2.294, df = 22, p-value = 0.01587
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.039   Inf
## sample estimates:
## mean of x mean of y
##    17.444     9.333
```

Hence, the average seniority of the academic teachers is significantly ($\alpha = 0.05$) greater that the seniority of the office workers, because p-value= ... $< \alpha$ (with test statistic $t = ...$). ⊡

Ex. 7.4. A memory test was performed on a group of 10 children (a pre-test). After a few months, during which these children were subjected to some regular development activities, the test was performed once again (a post-test). Assuming normality of the results and the significance level of 1%, determine whether the exercises significantly improved their memory skills.

Child ID	1,	2,	3,	4,	5,	6,	7,	8,	9,	10.
Pre-test	27,	21,	34,	24,	30,	27,	33,	31,	22,	27.
Post-test	29,	32,	29,	27,	31,	26,	35,	30,	29,	28.

Solution.
This time the pre- and post-test results $X_1, \ldots, X_{10}$ are $Y_1, \ldots, Y_{10}$ *pairwise dependent*.

```
pre <- c(27, 21, 34, 24, 30, 27, 33, 31, 22, 27)
post <- c(29, 32, 29, 27, 31, 26, 35, 30, 29, 28)
```

We have $(X_1, \ldots, X_n)$ i.i.d. $\mathrm{N}(\mu_1, \sigma_1)$, and $(Y_1, \ldots, Y_n)$ i.i.d. $\mathrm{N}(\mu_2, \sigma_2)$ with unknown expectations and variances. The paired t-test shall be used to verify $H_0 : \mu_1 = \mu_2$ against $K : \mu_1 < \mu_2$ (improvement?).

```
t.test(pre, post, alternative = "less", paired = T)
##
##  Paired t-test
##
## data:  pre and post
## t = -1.43, df = 9, p-value = 0.09322
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf 0.5634
## sample estimates:
## mean of the differences
##                      -2
```

As the p-value= ... $> \alpha = 0.01$, we do not reject H_0. The "development" activities did not significantly improve their memory skills.

⊡

Ex. 7.5. If some industrial device works properly, the amount of faulty items it outputs does not exceed 4%. If the number is greater than 4%, one should adjust the production process. A quality controller picked up a random sample of 200 items and found 14 faulty products. Shall the process be adjusted? Assume significance level of 0.05.

Solution.
We are given a random sample $(X_1, \ldots, X_n)$ i.i.d. Bern(p). We verify $H_0 : p = 0.04$ against $K > 0.04$ using one-sample test for proportion implemented in the `binom.test()` function ($\alpha = 0.05$).

```
binom.test(14, 200, p = 0.04, alternative = "greater")
```

```
##
##  Exact binomial test
##
## data:  14 and 200
## number of successes = 14, number of trials = 200, p-value = 0.03121
## alternative hypothesis: true probability of success is greater than 0.04
## 95 percent confidence interval:
##  0.04281 1.00000
## sample estimates:
## probability of success
##                   0.07
```

The null hypothesis therefore shall be rejected at $\alpha = 0.05$. The production process should be adjusted.

Note the result given by another function (however, the `binom.test()` procedure is preferred, because it has greater power):

```
prop.test(14, 200, p = 0.04, alternative = "greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  14 out of 200, null probability 0.04
## X-squared = 3.939, df = 1, p-value = 0.02359
## alternative hypothesis: true p is greater than 0.04
## 95 percent confidence interval:
##  0.04372 1.00000
## sample estimates:
##    p
## 0.07
```

⊡

Ex. 7.6. 455 of 700 M.Sc. students and 517 of 1320 B.Sc. students work for money during their studies. Do the proportions differ significantly?

Solution.
Here we have two independent samples $(X_1, \ldots, X_{n_1})$ i.i.d. Bern(p_1), and $(Y_1, \ldots, Y_{n_2})$ i.i.d. Bern(p_2). We verify $H_0 : p_1 = p_2$ versus $K : p_1 \neq p_2$ with the `prop.test()` function.

As significance level is not given, we may set e.g. $\alpha = 0.05$.

```
prop.test(c(455, 517), c(700, 1320))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(455, 517) out of c(700, 1320)
## X-squared = 121.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2132 0.3035
## sample estimates:
## prop 1 prop 2
## 0.6500 0.3917
```

Obviously, we conclude that the proportions of students working for money differ significantly between the two groups.

⊡