Short review

Let us recall: we have a population characteristic X, whose pdf $f(x; \theta)$ depends on θ, the target parameter to be estimated. The estimation is done based on a sample of size n, i.e. sample variables $X_1, X_2, \ldots, X_n$ that are *iid*, with the same pdf as X.

We set up two hypotheses, the *null* hypothesis, always simple, i.e.

$$H_0 : \theta = \theta_0$$

and one of the *alternative* hypotheses

$$
\begin{aligned}
H_1 &: \quad \theta < \theta_0 \quad \text{(left-tailed test)}, \\
H_1 &: \quad \theta > \theta_0 \quad \text{(right-tailed test)}, \\
H_1 &: \quad \theta \neq \theta_0 \quad \text{(two-tailed test)}.
\end{aligned}
\tag{5.1}
$$

We want to decide if H_0 is *rejected* (in favor of H_1) or *not rejected* (accepted). We use a *test statistic* TS (with the same properties as the pivot in CI's) and a *rejection (critical) region* RR, such that for a given *significance level* $\alpha \in (0, 1)$,

$$P(\text{type I error}) \;=\; P(\text{ reject } H_0 \mid H_0) \;=\; P(TS \in RR \mid H_0) \;=\; \alpha. \tag{5.2}$$

The probability of a *type II error* is

$$P(\text{ type II error}) \;=\; P(\text{ not reject } H_0 \mid H_1) \;=\; P(TS \notin RR \mid H_1) \;=\; \beta.$$

In general, the significance level α is preset and a procedure is given for finding an appropriate rejection region, such that β is also reasonably small.

We considered the case where for a target parameter θ, $\overline{\theta}$ is an unbiased estimator ($E(\overline{\theta}) = \theta$), with standard error $\sigma_{\overline{\theta}}$, such that, under certain conditions, it is known that

$$Z \;=\; \frac{\overline{\theta} - \theta}{\sigma_{\overline{\theta}}} \left(= \frac{\overline{\theta} - E(\overline{\theta})}{\sigma(\overline{\theta})} \right) \tag{5.3}$$

has an approximately Standard Normal $N(0, 1)$ distribution. Using Z as a test statistic, we found

the rejection regions for the three alternatives (5.1) as

$$RR: \begin{cases} \{Z_0 \leq z_\alpha\} \\ \{Z_0 \geq z_{1-\alpha}\} \\ \{Z_0 \leq z_{\frac{\alpha}{2}} \text{ or } Z_0 \geq z_{1-\frac{\alpha}{2}}\} = \{|Z_0| \geq z_{1-\frac{\alpha}{2}}\}. \end{cases} \tag{5.4}$$

Alternatively, we perform a *significance test*. We compute the **P-value** of the test, the probability of observing a value *at least as extreme* (in the sense of the test conducted) of the test statistic TS as the value observed from the sample, TS_0, under the assumption that H_0 is true. In general, for the three alternatives (5.1), if TS_0 is the value of the test statistic TS under the assumption that H_0 is true and F is the cdf of TS, the P-value is computed by

$$P = \begin{cases} P(TS \leq TS_0 \mid H_0) & = & F(TS_0) \\ P(TS \geq TS_0 \mid H_0) & = & 1 - F(TS_0) \\ 2 \cdot \min\{P(TS \leq TS_0 \mid H_0), P(TS \geq TS_0 \mid H_0)\} & = & 2 \cdot \min\{F(TS_0), 1 - F(TS_0)\}. \end{cases} \tag{5.5}$$

Then the decision will be

$$\begin{aligned} &\text{if } P \leq \alpha, \text{ reject } H_0, \\ &\text{if } P > \alpha, \text{ do not reject } H_0. \end{aligned} \tag{5.6}$$

So, more precisely, the P-value of a test is the smallest level at which we could have preset α and still have been able to reject H_0, or the lowest significance level that *forces* rejection of H_0, i.e. the *minimum rejection level*.

5.5 Tests for the Parameters of One Population

Let X be a population characteristic, with pdf $f(x; \theta)$, mean $E(X) = \mu$ and variance $V(X) = \sigma^2$. Let $X_1, X_2, \ldots, X_n$ be sample variables.

Tests for the mean of a population, $\theta = \mu$

We test the hypotheses

$$H_0: \quad \mu = \mu_0, \text{ versus one of}$$

$$H_1: \begin{cases} \mu < \mu_0 \\ \mu > \mu_0 \\ \mu \neq \mu_0, \end{cases} \tag{5.7}$$

under the assumption that either X is approximately Normally $N(\mu, \sigma)$ distributed or that the sample is large $(n > 30)$.

Case σ known ($\boxed{\textbf{ztest}}$)

We use the test statistic

$$TS \;=\; Z \;=\; \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \;\in\; N(0, 1), \tag{5.8}$$

with observed value

$$Z_0 \;=\; \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}. \tag{5.9}$$

Then, as before, at the $\alpha \in (0, 1)$ significance level, the rejection region for each test will be given by

$$RR: \quad \begin{cases} \{Z_0 \leq z_\alpha\} \\ \{Z_0 \geq z_{1-\alpha}\} \\ \{|Z_0| \geq z_{1-\frac{\alpha}{2}}\} \end{cases} \tag{5.10}$$

and the P-value will be computed as

$$P = \begin{cases} P(Z \leq Z_0 \mid H_0) & = & \Phi(Z_0) \\ P(Z \geq Z_0 \mid H_0) & = & 1 - \Phi(Z_0) \\ P(|Z| \geq |Z_0| \mid H_0) & = & 2\left(1 - \Phi(|Z_0|)\right), \end{cases} \tag{5.11}$$

since $N(0, 1)$ is symmetric, where

$$\Phi(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

is Laplace's function, the cdf for the Standard Normal $N(0, 1)$ distribution.

Case σ unknown ($\boxed{\textbf{ttest}}$)

In this case, we use the test statistic

$$TS \;=\; T \;=\; \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \;\in\; T(n - 1), \tag{5.12}$$

3

with observed value

$$T_0 = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}.$$

(5.13)

Similarly to the previous case, we find the rejection region for the three alternatives as

$$RR: \begin{cases} \{T_0 \le t_\alpha\} \\ \{T_0 \ge t_{1-\alpha}\} \\ \{|T_0| \ge t_{1-\frac{\alpha}{2}}\}, \end{cases}$$

(5.14)

and compute the P-value by

$$P = \begin{cases} P(T \le T_0 \mid H_0) &= F(T_0) \\ P(T \ge T_0 \mid H_0) &= 1 - F(T_0) \\ P(|T| \ge |T_0| \mid H_0) &= 2\left(1 - F(|T_0|)\right), \end{cases}$$

(5.15)

where the cdf F and the quantiles refer to the $T(n-1)$ distribution.

Tests for the variance of a population, $\theta = \sigma^2$ ($\boxed{\text{vartest}}$)

Assuming that X has a Normal $N(\mu, \sigma)$ distribution, we test the hypotheses

$$H_0: \quad \sigma^2 = \sigma_0^2, \qquad\qquad H_0: \quad \sigma = \sigma_0,$$
$$H_1: \begin{cases} \sigma^2 < \sigma_0^2 \\ \sigma^2 > \sigma_0^2 \\ \sigma^2 \ne \sigma_0^2, \end{cases} \text{equivalent to} \quad H_1: \begin{cases} \sigma < \sigma_0 \\ \sigma > \sigma_0 \\ \sigma \ne \sigma_0. \end{cases}$$

(5.16)

The test statistic will be

$$TS = V = \frac{(n-1)s^2}{\sigma^2} \in \chi^2(n-1),$$

(5.17)

with observed value

$$V_0 = \frac{(n-1)s^2}{\sigma_0^2}.$$

(5.18)

Even though the $\chi^2(n-1)$ distribution is not symmetric, we use the same line of reasoning and computations to find the rejection region for the three alternatives:

$$RR: \begin{cases} \{V_0 \leq \chi^2_\alpha\} \\ \{V_0 \geq \chi^2_{1-\alpha}\} \\ \{V_0 \leq \chi^2_{\frac{\alpha}{2}} \text{ or } V_0 \geq \chi^2_{1-\frac{\alpha}{2}}\}. \end{cases} \quad (5.19)$$

Same goes for the computation of the P-values:

$$P = \begin{cases} P(V \leq V_0 \mid H_0) & = & F(V_0) \\ P(V \geq V_0 \mid H_0) & = & 1 - F(V_0) \\ 2 \cdot \min\{P(V \leq V_0 \mid H_0), P(V \geq V_0 \mid H_0)\} & = & 2 \cdot \min\{F(V_0), 1 - F(V_0)\}, \end{cases}$$
$$(5.20)$$

where the cdf F and the quantiles refer to the $\chi^2(n-1)$ distribution.

Example 5.1. Let us consider again the problem in Example 5.4 (Lecture 10): The number of monthly sales at a firm is known to have a mean of 20 and a standard deviation of 4 and all salary, tax and bonus figures are based on these values. However, in times of economical recession, a sales manager fears that his employees do not average 20 sales per month, but less, which could seriously hurt the company. For a number of 36 randomly selected salespeople, it was found that in one month they averaged 19 sales.

Now, suppose that for the sample considered, the standard deviation is found to be $s = 4.5$. Assuming that the number of monthly sales at that firm is Normally distributed, at the 5% significance level, does the assumption on σ seem to be correct?

Solution. We are now testing the variance. We want to know if the value $\sigma = 4$ is correct *or not*, so, this will be a *two-tailed* test.

$$\begin{aligned} H_0: & \quad \sigma = 4 \\ H_1: & \quad \sigma \neq 4, \end{aligned}$$

i.e.,

$$\begin{aligned} H_0: & \quad \sigma^2 = 16 = \sigma_0^2 \\ H_1: & \quad \sigma^2 \neq 16 = \sigma_0^2, \end{aligned}$$

We have $n = 36$ and $s^2 = (4.5)^2 = 20.25$. The observed value of the test statistic is

$$V_0 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{35 \cdot 20.25}{16} = 44.2969.$$

5

The significance level is $\alpha = 0.05$ and the two quantiles for the $\chi^2(35)$ distribution are

$$\chi^2_{0.025} = 20.5694,$$
$$\chi^2_{0.975} = 53.2033.$$

Then the rejection region is

$$RR = (-\infty, 20.5694] \cup [53.2033, \infty),$$

which *does not* include the value V_0. Therefore, the decision is to *not reject* the null hypothesis, i.e. to conclude that the assumption $\sigma = 4$ is correct.

On the other hand, the P-value is

$$P = 2 \cdot \min\{P(V \le V_0), P(V \ge V_0)\} = 2 \cdot \min\{0.8652, 0.1348\} = 0.2697.$$

Since

$$\alpha = 0.05 < 0.2697 = P,$$

the decision is to *not reject* the null hypothesis.

Notice, again, that the significance test tells us more! Since the P-value is so large (remember, it is comparable to a probability of an *error*, so a *small* quantity), not only at the 5% significance level we decide to accept H_0, but at *any* reasonable significance level the decision would be the same. That means that the data *strongly* suggests that H_0 is true and should not be rejected. Even though the *sample* standard deviation *is not* equal to 4, still, statistically, the data strongly suggests that the *population* standard deviation *is* 4. We should be careful not to extrapolate the property of one sample to the entire population (data from a sample may be misleading, if it is not used properly ...)

∎

5.6 Tests for Comparing the Parameters of Two Populations

Assume we have two population characteristics $X_{(1)}$ and $X_{(2)}$, with means and variances $E(X_{(1)}) = \mu_1, V(X_{(1)}) = \sigma_1^2$ and $E(X_{(2)}) = \mu_2, V(X_{(2)}) = \sigma_2^2$, respectively. We draw two independent random samples $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$, with sample means $\overline{X}_1, \overline{X}_2$, sample variances s_1^2, s_2^2, respectively and *pooled* variance

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Tests for the difference of means, $\theta = \mu_1 - \mu_2$

We test the hypotheses

$$
\begin{array}{ll}
H_0: \quad \mu_1 - \mu_2 = 0, & H_0: \quad \mu_1 = \mu_2, \\[2mm]
H_1: \left\{ \begin{array}{l} \mu_1 - \mu_2 < 0 \\ \mu_1 - \mu_2 > 0 \\ \mu_1 - \mu_2 \neq 0, \end{array} \right. \quad \text{equivalent to} \quad & H_1: \left\{ \begin{array}{l} \mu_1 < \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 \neq \mu_2, \end{array} \right.
\end{array} \tag{5.21}
$$

under the assumption that either $X_{(1)}$ and $X_{(2)}$ have approximately Normal distributions or that the samples are large enough ($n_1 + n_2 > 40$).

Case σ_1, σ_2 known

We use the test statistic

$$
TS \;=\; Z \;=\; \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \;\in\; N(0,1), \tag{5.22}
$$

with observed value

$$
Z_0 \;=\; \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}. \tag{5.23}
$$

The rejection regions and P-values for the three alternatives are then given by equations (5.10)-(5.11), with Z_0 from (5.23).

Case $\sigma_1 = \sigma_2$ unknown (ttest2)

The test statistic is

$$
TS \;=\; T \;=\; \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \;\in\; T(n_1 + n_2 - 2), \tag{5.24}
$$

with observed value

$$
T_0 \;=\; \frac{\overline{X}_1 - \overline{X}_2}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}. \tag{5.25}
$$

The rejection regions and P-values for the three alternatives are then given by equations (5.14)-

(5.15), where T_0 is given in (5.25) and the cdf F and the quantiles refer to the $T(n_1 + n_2 - 2)$ distribution.

Case σ_1, σ_2 unknown ($\boxed{\textbf{ttest2}}$)

We now use the test statistic

$$TS \;=\; T^* \;=\; \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \;\in\; T(n), \tag{5.26}$$

where $\dfrac{1}{n} = \dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}$ and $c = \dfrac{\dfrac{s_1^2}{n_1}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}.$

The observed value of the test statistic is

$$T_0^* \;=\; \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}. \tag{5.27}$$

The rejection regions and P-values for the three alternatives are again as in equations (5.14)-(5.15), with T_0 replaced by T_0^* from (5.27). The cdf F and the quantiles refer to the $T(n)$ distribution.

Remark 5.2. The same Matlab command **ttest2** performs a T-test for the difference of two population means, when the variances are *not* assumed equal, with the option *vartype* set on "unequal" (the default being "equal", when it can be omitted).

Tests for the ratio of variances, $\theta = \dfrac{\sigma_1^2}{\sigma_2^2}$ ($\boxed{\textbf{vartest2}}$)

Assuming that both $X_{(1)}$ and $X_{(1)}$ have Normal distributions, we test the hypotheses

$$
H_0 : \; \frac{\sigma_1^2}{\sigma_2^2} = 1,
$$

$$
H_1 : \;
\begin{cases}
\dfrac{\sigma_1^2}{\sigma_2^2} < 1 \\[2mm]
\dfrac{\sigma_1^2}{\sigma_2^2} > 1 \\[2mm]
\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1,
\end{cases}
\quad <=> \quad
\begin{aligned}
& H_0 : \; \sigma_1^2 = \sigma_2^2, \\[2mm]
& H_1 :
\begin{cases}
\sigma_1^2 < \sigma_2^2 \\
\sigma_1^2 > \sigma_2^2 \\
\sigma_1^2 \neq \sigma_2^2,
\end{cases}
\end{aligned}
\quad <=> \quad
\begin{aligned}
& H_0 : \; \sigma_1 = \sigma_2, \\[2mm]
& H_1 :
\begin{cases}
\sigma_1 < \sigma_2 \\
\sigma_1 > \sigma_2 \\
\sigma_1 \neq \sigma_2.
\end{cases}
\end{aligned}
\tag{5.28}
$$

The test statistic used is

$$TS = F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \in F(n_1 - 1, n_2 - 1), \qquad (5.29)$$

with observed value

$$F_0 = \frac{s_1^2}{s_2^2}. \qquad (5.30)$$

Again, just like in the case of one population variance, the $F(n_1 - 1, n_2 - 1)$ distribution is not symmetric, but proceeding as before, we find the rejection region for the three alternatives as

$$RR: \begin{cases} \{F_0 \le f_\alpha\} \\ \{F_0 \ge f_{1-\alpha}\} \\ \{F_0 \le f_{\frac{\alpha}{2}} \text{ or } F_0 \ge f_{1-\frac{\alpha}{2}}\}. \end{cases} \qquad (5.31)$$

and the P-values given by

$$P = \begin{cases} P(F \le F_0 \mid H_0) & = F(F_0) \\ P(F \ge F_0 \mid H_0) & = 1 - F(F_0) \\ 2 \cdot \min\{P(F \le F_0 \mid H_0), P(F \ge F_0 \mid H_0)\} & = 2 \cdot \min\{F(F_0), 1 - F(F_0)\}, \end{cases} \qquad (5.32)$$

where the cdf F and the quantiles refer to the $F(n_1 - 1, n_2 - 1)$ distribution.

Example 5.3. Suppose the strengths to a certain load of two types of material, $M1$ and $M2$, are studied, knowing that they are approximately Normally distributed. The more weight they can resist to, the stronger they are. Two independent random samples are drawn and they yield the following data.

	$M1$		$M2$
$n_1 =$	25	$n_2 =$	16
$\overline{X}_1 =$	380	$\overline{X}_2 =$	370
$s_1^2 =$	537	$s_2^2 =$	196

a) At the 5% significance level, do the variances of the two populations seem to be equal or not?

b) At the same significance level, does the data suggest that on average, $M1$ is stronger than $M2$? (In both parts, perform both hypothesis and significance testing).

Solution.

a) First, we compare the variances of the two populations, so we know which way to proceed for

comparing the means. We want to know if they are equal or not, so it is a two-tailed test. Hence, our hypotheses are

$$H_0: \quad \sigma_1^2 = \sigma_2^2$$
$$H_1: \quad \sigma_1^2 \neq \sigma_2^2.$$

The observed value of the test statistic is

$$F_0 \;=\; \frac{s_1^2}{s_2^2} \;=\; \frac{537}{196} = 2.7398.$$

For $\alpha = 0.05$, $n_1 = 25$ and $n_2 = 16$, the quantiles for the $F(24, 15)$ distribution are

$$f_{\frac{\alpha}{2}} \;=\; f_{0.025} \;=\; 0.4103$$
$$f_{1-\frac{\alpha}{2}} \;=\; f_{0.975} \;=\; 2.7006.$$

Thus, the rejection region for our test is

$$RR \;=\; (-\infty,\; 0.4103] \,\cup\, [2.7006, \infty)$$

and clearly, $F_0 \in RR$. Thus we reject H_0 in favor of H_1, i.e. we conclude that the data suggests that the population variances are *different*.

Let us also perform a significance test. The P-value of this (two-tailed) test is

$$P \;=\; 2 \cdot \min\{P(F \leq F_0), P(F \geq F_0)\} \;=\; 2 \cdot \min\{0.9765, 0.0235\} \;=\; 0.0469.$$

Since our $\alpha > P$, the "minimum rejection significance level", we reject H_0.

Note. We now know that for instance, at 1% significance level (or any level less than 4.69%), we would have *not* rejected the null hypothesis. This goes to show that the data can be "misleading". Simply comparing the values of the sample functions does not necessarily mean that the same thing will be true for the corresponding population parameters. Here, s_1^2 is *much* larger than s_2^2, yet at 1% significance level, we would have concluded that the population variances seem to be equal.

b) Next we want to compare the population means. If $M1$ is to be *stronger* than $M2$ on average, than we must perform a *right*-tailed test:

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 > \mu_2$$

Which one of the tests for the difference of means should we use? The answer is in part a). At this significance level, the variances are unknown and *different*.

Then the value of the test statistic is, by (5.27)

$$T_0^* = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{380 - 370}{\sqrt{\dfrac{537}{25} + \dfrac{196}{16}}} = 1.7218.$$

To find the rejection region, we compute

$$c = 0.6368, \quad n = 38.9244 \approx 39$$

and the quantile for the $T(39)$ distribution

$$t_{1-\alpha} = t_{0.95} = 1.6849.$$

Then the rejection region of the test is

$$RR = [1.6849, \infty),$$

which includes the value T_0^*, so we *reject* H_0 in favor of H_1. So we conclude that yes, the data suggests that material $M1$ is, on average, stronger than material $M2$.

On the other hand, the P-value of this test is

$$P = P(T^* \geq T_0^*) = 1 - F(T_0^*) = 1 - F(1.7218) = 0.0465,$$

where F is the cdf of the $T(39)$ distribution. Again, the P-value is lower than $\alpha = 0.05$, which forces the rejection of H_0.

∎

Tests for the difference of means, paired data, $\theta = \mu_1 - \mu_2$

Recall that in some applications, we want to compare the means of two populations, when two random samples (one from each population) are available, which *are not* independent, where each observation in one sample is naturally or by design *paired* with an observation in the other sample (usually cases best described by "before and after" situations).

In such cases, both samples have the same length, n:

$$X_{11}, \ldots, X_{1n} \text{ and } X_{21}, \ldots, X_{2n}$$

and we consider the sample of their *differences*,

$$D_1, \ldots, D_n,$$

where

$$D_i = X_{1i} - X_{2i}, \; i = \overline{1, n}.$$

For this sample, we have

$$\overline{X}_d \;=\; \frac{1}{n} \sum_{i=1}^{n} D_i, \text{ the sample mean and}$$

$$s_d^2 \;=\; \frac{1}{n-1} \sum_{i=1}^{n} \left(D_i - \overline{X}_d \right)^2, \text{ the sample variance.}$$

Then, it is known that when n is large enough ($n > 30$) or the two populations that the samples are drawn from have approximately Normal distributions $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)$, the statistic

$$T \;=\; \frac{\overline{X}_d - (\mu_1 - \mu_2)}{\frac{s_d}{\sqrt{n}}} \tag{5.33}$$

has a Student $T(n-1)$ distribution, so we can use it as a test statistic for testing the hypotheses (5.21). Its observed value is

$$T_0 \;=\; \frac{\overline{X}_d}{\frac{s_d}{\sqrt{n}}}. \tag{5.34}$$

Then, as before, we determine the rejection region corresponding to the three alternatives to be

$$RR: \quad \begin{cases} \{T_0 \leq t_\alpha\} \\ \{T_0 \geq t_{1-\alpha}\} \\ \{|T_0| \geq |t_{1-\frac{\alpha}{2}}|\} \end{cases} \tag{5.35}$$

and compute the P-value by

$$
P = \begin{cases}
P(T \leq T_0 \mid H_0) & = & F(T_0) \\
P(T \geq T_0 \mid H_0) & = & 1 - F(T_0) \\
P(|T| \geq |T_0| \mid H_0) & = & 2\left(1 - F(|T_0|)\right),
\end{cases}
\tag{5.36}
$$

where the quantiles and the cdf F refer to the $T(n-1)$ distribution.

Example 5.4. Information about ocean weather can be extracted from radar returns with the aid of special algorithms. A study is conducted to estimate the difference in wind speed as measured on the ground, at 12 specified times, using two methods simultaneously. These data result:

Times	1	2	3	4	5	6	7	8	9	10	11	12
Method I	4.46	3.99	3.73	3.29	4.82	6.71	4.61	3.87	3.17	4.42	3.76	3.3
Method II	4.08	3.94	5.00	5.2	3.92	6.21	5.95	3.07	4.76	3.25	4.89	4.8

Assuming the measurements taken by the two methods are approximately Normally distributed, at the 1% significance level, does the data suggest that, on average, the two sets of measurements differ?

Solution. By looking at the data, we see that at some times the measurement taken by the first method is higher, at others, the one given by the second. So we cannot say if, on average, these differences will cancel each other, to yield about the same mean value.

So, we want to test

$$
\begin{aligned}
H_0 : & \quad \mu_1 = \mu_2 \\
H_1 : & \quad \mu_1 \neq \mu_2,
\end{aligned}
$$

a two-tailed alternative. The samples yield the following data: sample size $n = 12$, sample mean $\overline{X}_d = -0.4117$ and sample variance $s_d^2 = 1.2973$, so $s_d = 1.139$.

The observed value of the test statistic from (5.34) is

$$
T_0 = \frac{\overline{X}_d}{\frac{s_d}{\sqrt{n}}} = -1.2521.
$$

For $\alpha = 0.01$, the quantiles for the $T(11)$ distribution are

$$
\begin{aligned}
t_{\alpha/2} & = & t_{0.005} & = & -3.1058, \\
t_{1-\alpha/2} & = & -t_{\alpha/2} & = & 3.1058,
\end{aligned}
$$

so the rejection region is

$$RR = (-\infty, -3.1058] \cup [3.1058, \infty).$$

Since $T_0 \notin RR$, we cannot reject the null hypothesis, which means we decide that the two population means are approximately equal.

On the other hand, the P-value of this test is

$$P = 2\left(1 - F(|T_0|)\right) = 0.2365,$$

We have

$$\alpha = 0.01 < 0.2365 = P,$$

the minimum rejection level, so the decision is to *not reject* the null hypothesis. Notice that, again, the P-value is much larger than any conceivable significance level α, so that means that the data strongly suggests that H_0 should not be rejected, i.e., that the two population means *do not* differ.

∎

Remark 5.5.

1. As mentioned before, both hypothesis and significance testing lead to the same conclusion. From the implementation point of view, significance testing is more efficient, since it avoids the inversion of a cdf (i.e. computation of quantiles), which is often a complicated improper integral. This is the reason why, although the main tests *are* implemented in Matlab, the rejection regions *are not* computed.

2. Many tests (and formulas for CI's) work under the assumption of Normality of the population from which the sample was drawn. In practice, when there are outliers in the data, that is rarely the case. How important is this assumption of Normality and how affected are the results of these tests by small departures from model assumptions? Z-tests and T-tests work well even when the underlying population is not quite Normally distributed. From this point of view, they are called **robust** tests. χ^2-tests and F-tests, however, are *not* robust, they perform very poorly when the assumption of Normality is breached. In modern Statistics there is an ongoing search for finding robust methods of estimation for variances.