Table 6

sample (i)	Polarographic method (x_{1i})	Titration method (x_{2i})	Differences ($d_i = x_{1i} - x_{2i}$)
1	$18,60$	$18,58$	$0,02$
2	$27,60$	$27,37$	$0,23$
3	$27,50$	$27,27$	$0,23$
4	$25,00$	$24,64$	$0,36$
5	$24,50$	$24,10$	$0,40$
6	$26,80$	$26,33$	$0,47$
7	$29,70$	$29,33$	$0,37$
8	$26,63$	$26,50$	$0,13$

$$+(0,36-0,2763)^2+(0,40-0,2763)^2+(0,47-0,2763)^2+(0,37-0,2763)^2+$$

$$+(0,13-0,2763)^2\big]=0,0229 \Rightarrow s_d=\sqrt{s_d^2}=\sqrt{0,0229}=0,1513.$$

The value of the test statistic is then $t = 5,1684$. The quantile of the Student t-distribution $t_{1-\frac{\alpha}{2}}(n-1) = t_{0,995}(7) = 3,4995$ (Table 29). The critical region is

$$K_{0,01} = (-\infty; -3,4995) \cup (3,4995; \infty) \Rightarrow t \in K_{0,01}.$$

The value of the test statistic $t = 5,1684$ falls into the critical region $K_{0,01}$, therefore we reject the null hypothesis H_0 and accept the alternative hypothesis H_1. We can assume that, at the significance level $\alpha = 0,01$, the methods used are not equivalent.

5.4 Symmetry tests

Symmetry tests belong among the nonparametric tests. There are several such tests, but we will present the sign test and the Wilcoxon two-sample test.

5.4.1 Sign test

The *sign test (median test)* is one of the oldest and simplest nonparametric tests, which tests the symmetry of the distribution of a statistical variable about a constant x_0 as the median.

Consider a random sample V_n with a continuous distribution with an unknown median $\widetilde{x}$. The sample size is n.

1. Hypotheses:

$$H_0 : \widetilde{x} = x_0,$$
$$H_1 : \widetilde{x} \neq x_0.$$

2. The test statistic is

$$T = \min\{Z^+, Z^-\} \sim \text{Bi}\left(n_0, \frac{1}{2}\right), \tag{109}$$

where Z^+ is the number of values $d_i = x_i - x_0$ with a positive sign and Z^- is the number of values $d_i = x_i - x_0$ with a negative sign, n_0 is the number of nonzero differences $d_i = x_i - x_0$.

3. The critical region K_α is

$$K_\alpha = (-\infty; k_1(\alpha; n_0)) \cup (k_2(\alpha; n_0); \infty),$$

where $k_1(\alpha; n_0)$, $k_2(\alpha; n_0)$ are the critical values of the sign test, tabulated for $n_0 \leqq 100$ for $\alpha = 0,05$ and $\alpha = 0,01$ (Table 45).

Remark: Since the binomial distribution $\text{Bi}\left(n_0, \frac{1}{2}\right)$ is symmetric, $k_2 = n_0 - k_1$ holds.

For a large random sample size (it is sufficient that $n \geqq 20$), we can use the following test:

1. Hypotheses:

$$H_0 : \widetilde{x} = x_0,$$
$$H_1 : \widetilde{x} \neq x_0.$$

2. The test statistic is

$$Z = \frac{2T - n_0}{\sqrt{n_0}} \sim N(0, 1), \tag{110}$$

where $T = \min\{Z^+, Z^-\}$, Z^+ is the number of values $d_i = x_i - x_0$ with a positive sign and Z^- is the number of values $d_i = x_i - x_0$ with a negative sign, n_0 is the number of nonzero values $d_i = x_i - x_0$.

3. The critical region K_α is

$$K_\alpha = \left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right).$$

Remark: The sign test is also often used as a *paired sign test*, which is similar to the parametric paired t-test. In this case we calculate the differences d_i as differences of each pair

of observations, i. e. $d_i = x_i - y_i$ for $i = 1, 2, \ldots, n$, where x_i are values from the first random sample and y_i are values from the second random sample. We then proceed in the same way as for the sign test for one sample.

Example 5.10 *From a group of products, 12 pieces were selected for a weight check [in g]: 21.3; 18.5; 19.7; 26.5; 16.7; 23.1; 19.3; 17.5; 19.1; 24.3; 15.9; 20.8. At the significance level α=0.05, let us test whether the average weight of the product is 19 grams.*

Solution: Since we do not know the distribution of the population, we use the sign test. We will test

$$H_0 : \widetilde{x} = 19,$$
$$H_1 : \widetilde{x} \neq 19.$$

Table 7 shows the measured values and the differences $d_i = x_i - 19$.

Table 7

x_i	$21,3$	$18,5$	$19,7$	$26,5$	$16,7$	$23,1$	$19,3$	$17,5$	$19,1$	$24,3$	$15,9$	$20,8$
d_i	$2,3$	$-0,5$	$0,7$	$7,5$	$-2,3$	$4,1$	$0,3$	$-1,5$	$0,1$	$5,3$	$-3,1$	$1,8$
sign	$+$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	$+$

The number of differences Z^+ with a positive sign is 8, the number of differences Z^- with a negative sign is 4. We determine the value of the test statistic T using relation (109) $T = \min\{Z^+, Z^-\} = 4$. Since the number of nonzero values is $n_0 = 12$ and $\alpha = 0,05$, then $k_1(\alpha; n_0) = k_1(0,05; 12) = 2$ (Table 45) and $k_2(\alpha; n_0) = k_2(0,05; 12) = 10$. The critical region is

$$K_{0,05} = (-\infty; 2) \cup (10; \infty) \Rightarrow T \notin K_{0,05}.$$

The value of the test statistic $T = 4$ does not fall into the critical region $K_{0,05}$, therefore we do not reject the null hypothesis H_0. We can assume that, at the significance level $\alpha = 0,05$, the median weight of the products is 19 grams.

5.4.2 Wilcoxon two-sample test

The *Wilcoxon two-sample test (rank sum test, Mann-Whitney U test)* represents the nonparametric analogue of the two-sample t-test for independent samples. It is used instead of that test when the samples do not come from a normal distribution.

Consider two independent random samples V_{n_1} and V_{n_2}. The random sample V_{n_1} comes from a population with distribution function $F(x)$ with median $\widetilde{x}_1$ and the random sample

V_{n_2} comes from a population with distribution function $G(x)$ with median $\widetilde{x}_2$. We choose the samples such that the condition $n_1 \leqq n_2$ is satisfied.

First we arrange all $n_1 + n_2$ values of the random samples into a nondecreasing sequence, which forms the so-called *combined ranked series*. Then we assign rank numbers to the individual values in the combined ranked series, i. e. we number the values from smallest to largest with natural numbers, whereby equal values are assigned the same average rank (the arithmetic mean of their ranks).

For a random sample of small size, we can use the following test:

1. Hypotheses:

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2,$$

resp.

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$U = \min\{U_1; U_2\}, \tag{111}$$

where

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1, \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_2, \tag{112}$$

where T_1 is the sum of the ranks of the values from the sample V_{n_1} and T_2 is the sum of the ranks of the values from the sample V_{n_2}. For verification we can use the relations

$$U_1 + U_2 = n_1\, n_2 \quad \text{and} \quad T_1 + T_2 = \frac{1}{2}(n_1 + n_2)(n_1 + n_2 + 1).$$

3. The critical region K_α is

$$K_\alpha = \left(-\infty; U_\alpha(n_1, n_2)\right),$$

where $U_\alpha(n_1, n_2)$ is the tabulated critical value of the Wilcoxon two-sample test (Table 46).

For a random sample of large size, we can use the following test:

1. Hypotheses:

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2,$$

resp.

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$U_0 = \frac{U_1 - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \sim N(0,1), \tag{113}$$

where

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1, \tag{114}$$

where T_1 is the sum of the ranks of the values from the sample V_{n_1}.

3. The critical region K_α is

$$K_\alpha = \left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right).$$

Example 5.11 *A pharmaceutical company compares the adverse side effects of two manufactured drugs. The percentage [in %] of patients in whom these effects were recorded is monitored. The obtained values are given in the table. Using the Wilcoxon two-sample test, let us test at the significance level α=0.05 whether both drugs have the same percentage of adverse effects.*

Drug A	*31*	*33*	*29*	*35*		
Drug B	*25*	*34*	*28*	*24*	*33*	*40*

Solution: We will test

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2.$$

The null hypothesis states that the percentage of patients in whom adverse effects appeared is the same for both drugs. Against the two-sided alternative hypothesis, which states that the percentage of patients in whom adverse effects appeared is different for both drugs.

The size of the first sample is $n_1 = 4$ and of the second $n_2 = 6$. We arrange all values, regardless of which group they belong to, into a nondecreasing sequence and assign a rank to each value. If we have equal values, we assign them the same average rank (Table 8).

We calculate the rank sums T_1, T_2 and the values U_1, U_2 using relations (112).

$$T_1 = 4 + 5 + 6,5 + 9 = 24,5,$$

Table 8

Values	24	25	28	29	31	33	33	34	35	40
Drug	B	B	B	A	A	A	B	B	A	B
Rank	1	2	3	4	5	6,5	6,5	8	9	10

$$T_2 = 1 + 2 + 3 + 6,5 + 8 + 10 = 30,5,$$
$$U_1 = 4 \cdot 6 + \frac{4(4+1)}{2} - 24,5 = 9,5,$$
$$U_2 = 4 \cdot 6 + \frac{6(6+1)}{2} - 30,5 = 14,5.$$

We determine the value of the test statistic U using relation (111)

$$U = \min\{9,5; 14,5\} = 9,5.$$

The critical value of the Wilcoxon two-sample test is $U_\alpha(n_1, n_2) = U_{0,05}(4,6) = 2$ (Table 46). The critical region is

$$K_{0,05} = (-\infty; 2) \Rightarrow U \notin K_{0,05}.$$

The value of the test statistic $U = 9,5$ does not fall into the critical region $K_{0,05}$, therefore at the significance level $\alpha = 0,05$ we do not reject the null hypothesis H_0. We can assume that the percentage of patients in whom adverse effects appeared is the same for both drugs.

5.5 Goodness-of-fit tests

In most of the tests we have discussed so far, we assumed that the population follows a certain distribution. In practice, such a statement must be verified. This is the purpose of so-called *goodness-of-fit tests*. There are many such tests; we will present the most commonly used ones: Pearson's goodness-of-fit test, the Kolmogorov–Smirnov test, the Anderson–Darling test, and the Shapiro–Wilk normality test.

Consider a random sample V_n that comes from a population with a probability distribution with an unknown distribution function $F(x)$. Also given is a known distribution function $G(x)$ of a theoretical probability distribution.

5.5.1 Pearson's goodness-of-fit test

Pearson's goodness-of-fit test (the χ^2 goodness-of-fit test) is one of the oldest statistical tests (dating from 1900). It is a universal test that can be used for distributions of both discrete

and continuous random variables. In doing so, we assume that the values of the random sample are sorted into k classes (intervals).

1. Hypotheses:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} \sim \chi^2(k-r-1), \tag{115}$$

where n_j is the number of values of the observed random sample in the j-th class (interval), p_j is the probability measure of the j-th class (interval) determined by the distribution function G, np_j is the theoretical frequency in the j-th class (interval), r is the number of estimated parameters when determining G, k is the number of classes and n is the size of the random sample (at least $n \geqq 30$).

3. The critical region K_α is

$$K_\alpha = \left(\chi^2_{1-\alpha}(k-r-1); \infty\right).$$

In practical use of the test, satisfaction of the so-called *Cochran's rule* is required, i. e. $np_j \geqq 5$ for $j = 1, 2, \ldots, k$. If this condition is not satisfied, i. e. $np_j < 5$, we merge adjacent classes so that the condition is subsequently satisfied. However, strict adherence to it is only necessary for a small number of degrees of freedom of the test statistic χ^2. For a larger number of degrees of freedom of the test statistic $k - r - 1 \geqq 3$, it is sufficient to require the condition $np_j \geqq 4$, and for $k-r-1 \geqq 6$ it is sufficient to require the condition $np_j \geqq 1$ for $j = 1, 2, \ldots, k$.

Example 5.12 *At the significance level α=0.05, let us determine whether the random variable given by the table comes from a population with an exponential distribution. Let us use Pearson's goodness-of-fit test.*

I_j	⟨0;1⟩	⟨1;2⟩	⟨2;3⟩	⟨3;4⟩	⟨4;5⟩
n_j	60	20	11	5	4

Solution: The sample size is $n = 100$. The values of the random sample are divided into $k = 5$ classes (intervals). The number of estimated parameters when determining $G(x)$ is $r = 1$ (the exponential distribution has one parameter λ). To estimate the parameter λ we

Table 9

I_j	$(0;1\rangle$	$(1;2\rangle$	$(2;3\rangle$	$(3;4\rangle$	$(4;5\rangle$
z_j	$0,5$	$1,5$	$2,5$	$3,5$	$4,5$
n_j	60	20	11	5	4

use the sample arithmetic mean $\overline{x}$, which we calculate using relation (72), where z_j are the values of the class mark calculated using relation (4) and given in Table 9.
Then for the parameter λ we have

$$\lambda \approx \overline{x} = \frac{1}{100}\left(0,5 \cdot 60 + 1,5 \cdot 20 + 2,5 \cdot 11 + 3,5 \cdot 5 + 4,5 \cdot 4\right) = 1,23.$$

We will test

$$H_0 : F(x) = G(x), \text{ where } G(x) \sim Ex(\lambda),$$
$$H_1 : F(x) \neq G(x).$$

We are testing whether the distribution function $F(x)$ of the distribution from which the random sample comes is equal to the distribution function $G(x)$ of the exponential distribution. We determine the value of the test statistic χ^2 using relation (115), and to calculate its value we use Table 10.

Table 10

j	t_j	$G(t_j)$	$G(t_{j+1})$	p_j	$n\,p_j$	n_j
1	0	0	$0,5565$	$0,5565$	$55,65$	60
2	1	$0,5565$	$0,8033$	$0,2468$	$24,68$	20
3	2	$0,8033$	$0,9128$	$0,1095$	$10,95$	11
4	3	$0,9128$	$0,9613$	$0,0485$	$4,85$	5
5	4	$0,9613$	1	$0,0387$	**3,87**	4
6	∞	1	–	–	–	–

In column 2 of Table 10 are the boundaries t_j of the class intervals, where the lower boundary of the first interval, i. e. 0, remains unchanged in this case and the upper boundary of the last interval is replaced by ∞ (see the domain of the distribution function of the exponential distribution).

In column 3 of Table 10 are the values of the distribution function of the exponential distribution $G(t_j)$, which we calculate using the relation for the distribution function of the exponential distribution (60)

$$G(t_j) = 1 - e^{-\frac{t_j}{\lambda}} = 1 - e^{-\frac{t_j}{1,23}}.$$

In column 5 of Table 10 are the theoretical probabilities p_j, which we calculate based on the difference of the values of the distribution function, i. e. $p_j = G(t_{j+1}) - G(t_j)$, where the values $G(t_{j+1})$ are given in column 4 of Table 10.

When using the test, satisfaction of Cochran's rule is required. In our case the number of classes is $k = 5$ and the number of estimated parameters is $r = 1$, so $k - r - 1 = 5 - 1 - 1 = 3 \geqq 3$. It is sufficient to require the condition $np_j \geqq 4$.

In column 6, row five of Table 10, the condition is not satisfied. We merge the fourth and fifth rows in column 6 of Table 10, giving us a number of classes $k = 4$ (Table 11). The number of estimated parameters $r = 1$, so $k - r - 1 = 4 - 1 - 1 = 2$. Now the condition $np_j \geqq 5$ must hold, which is satisfied as we see in Table 11.

Table 11

j	t_j	$G(t_j)$	p_j	$n\,p_j$	n_j	$\frac{(n_j - np_j)^2}{np_j}$
1	0	0	0,5565	55,65	60	0,3400
2	1	0,5565	0,2468	24,68	20	0,8875
3	2	0,8033	0,1095	10,95	11	0,0002
4	3	0,9128	0,0485	8,72	9	0,0090
5	4	0,9613	0,0387			
6	∞	1	–	–	–	$\sum_{j=1}^{5} \frac{(n_j - np_j)^2}{np_j} = 1,2367$

The value of the test statistic is $\chi^2 = 1,2367$ and the quantile of the χ^2-distribution is $\chi^2_{1-\alpha}(k - r - 1) = \chi^2_{0,95}(2) = 5,99$ (Table 30). The critical region is

$$K_{0,05} = (5,99; \infty) \Rightarrow \chi^2 \notin K_{0,05}.$$

The value of the test statistic $\chi^2 = 1,2367$ does not fall into the critical region $K_{0,05}$, therefore we do not reject the null hypothesis H_0. We can assume that, at the significance level $\alpha = 0,05$, the random sample comes from an exponential distribution, although we do not know its parameter λ.

5.5.2 Kolmogorov–Smirnov test

The *Kolmogorov–Smirnov test* is a nonparametric test used for testing distributions of continuous-type random variables. Moreover, the distribution must be given uniquely, i. e. without unknown parameters. Unlike Pearson's goodness-of-fit test, the test can also be used for a random sample of small size (i.e., it does not require sorting into intervals).

1. Hypotheses:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$D = \sup_{x \in R} |F_n(x) - G(x)|, \tag{116}$$

where $G(x)$ is the considered theoretical distribution function of a continuous distribution, $F_n(x)$ is the empirical distribution function of the random sample, for which

$$F_n(x) = \frac{1}{n} \sum_{x_j \leqq x} n_j, \tag{117}$$

where n_j is the frequency of the value x_j in the random sample and n is the size of the random sample.

Given this definition of the empirical distribution function $F_n(x)$, it is possible to calculate the test statistic D using the relation

$$D = \max\{\max(d_j^-); \max(d_j^+)\}, \tag{118}$$

where $d_j^- = |F_n(x_{j-1}) - G(x_j)|$, $d_j^+ = |F_n(x_j) - G(x_j)|$ for $j = 1, 2, \ldots, n$.

3. The critical region K_α is

$$K_\alpha = (D_\alpha(n); \infty),$$

where $D_\alpha(n)$ is the tabulated critical value for the Kolmogorov–Smirnov test (Table 47). For large sample sizes ($n > 100$), the critical values $D_\alpha(n)$ are approximated by the relation

$$D_\alpha(n) \doteq \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}. \tag{119}$$

Remark: If all measured values $x_1, x_2, \ldots, x_n$ are pairwise distinct, we can use a special form of the empirical distribution function

$$F_n(x) = \begin{cases} 0 & \text{for } -\infty < x < x_{(1)}, \\ \frac{i}{n} & \text{for } x_{(i)} \leqq x < x_{(i+1)},\ i = 1, 2, \ldots, n-1, \\ 1 & \text{for } x_{(n)} < x < \infty, \end{cases}$$

where the values of the random sample are arranged into a nondecreasing sequence according to size (ranked series), i.e.

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

Example 5.13 *At the significance level α=0.05, let us verify whether the random variable given by the table comes from a population with an exponential distribution with parameter λ=1.23. Let us use the Kolmogorov–Smirnov test.*

x_j	0,5	1,5	2,5	3,5	4,5
n_j	60	20	11	5	4

Solution: The sample size is $n = 100$, the parameter of the exponential distribution is $\lambda = 1,23$. We will test

$$H_0 : F(x) = G(x), \text{ where } G(x) \sim Ex(1,23),$$
$$H_1 : F(x) \neq G(x).$$

We are testing whether the distribution function $F(x)$ of the distribution from which the random sample comes is equal to the distribution function $G(x)$ of the exponential distribution with parameter $\lambda = 1,23$. We determine the value of the test statistic D using relation (118). To calculate its value we use Table 12.

Table 12

j	x_j	n_j	$G(x_j)$	$F_n(x_{j-1})$	$F_n(x_j)$	d_j^-	d_j^+
1	0,5	60	0,3340	0,00	0,60	**0,3340**	**0,2660**
2	1,5	20	0,7046	0,60	0,80	0,1046	0,0954
3	2,5	11	0,8600	0,80	0,91	0,0690	0,0410
4	3,5	5	0,9419	0,91	0,96	0,0319	0,0181
5	4,5	4	0,9742	0,96	1,00	0,0142	0,0258

In column 4 of Table 12 are the values $G(x_j)$ of the distribution function of the exponential distribution, which we calculate using the relation for the distribution function of the exponential distribution (60)

$$G(x_j) = 1 - e^{-\frac{x_j}{\lambda}} = 1 - e^{-\frac{x_j}{1,23}}.$$

In columns 5 and 6 of Table 12 are the values of the empirical distribution function of the random sample $F_n(x_{j-1})$ and $F_n(x_j)$ calculated using relation (117).

In column 7 of Table 12 are the values d_j^-, where $d_j^- = |F_n(x_{j-1}) - G(x_j)|$, and in column 8 of Table 12 are the values d_j^+, where $d_j^+ = |F_n(x_j) - G(x_j)|$.

The value of the test statistic is the largest number from the last two columns, i. e. $D = \max\{0,3340; 0,2660\} = 0,3340$.

We calculate the critical value $D_\alpha(n)$ using relation (119). We have $D_\alpha(n) = D_{0,05}(100) = 0,1358$ (Table 47). The critical region is

$$K_{0,05} = (D_\alpha(n); \infty) = (0,1358; \infty) \Rightarrow D \in K_{0,05}.$$

The value of the test statistic $D = 0,334$ falls into the critical region $K_{0,05}$, therefore we reject the null hypothesis H_0 and accept the alternative hypothesis H_1. We can assume that, at the significance level $\alpha = 0,05$, the random sample does not come from an exponential distribution with parameter $\lambda = 1,23$. However, this does not mean that it does not come from an exponential distribution with a different parameter.

5.5.3 Anderson–Darling test

The *Anderson–Darling test* makes it possible to test the agreement of the empirical distribution function with the theoretical distribution function of the normal, exponential, lognormal, or Weibull distribution. Unlike the previous test, the critical values depend on the theoretical distribution.

1. Hypotheses:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$AD = \frac{1}{n}\sum_{i=1}^{n}(1-2i)\left[\ln G(x_{(i)}) + \ln\left(1 - G(x_{(n+1-i)})\right)\right] - n, \tag{120}$$

   where n is the size of the random sample. The condition is that the values of the random sample must be arranged into a nondecreasing sequence according to size (ranked series), i. e.

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

3. The critical region K_α is

$$K_\alpha = (AD_\alpha; \infty),$$

   where AD_α is the tabulated critical value of the test.

If the theoretical distribution function $G(x)$ comes from a normal distribution with a small sample size ($n \leqq 10$), the value AD_α is determined using the relation

$$AD_\alpha \approx \frac{k(\alpha)}{1 + \frac{0,75}{n} + \frac{2,25}{n^2}}, \tag{121}$$

Table 13: Values of the constant $k(\alpha)$ for various significance levels α

α	$0,005$	$0,01$	$0,0025$	$0,05$	$0,1$
$k(\alpha)$	$1,159$	$1,035$	$0,873$	$0,752$	$0,631$

where $k(\alpha)$ is a constant for the significance level α given in Table 13.

If the theoretical distribution function $G(x)$ comes from an exponential distribution, the value AD_α for the significance level α is given in Table 14.

Table 14: Values of AD_α for various significance levels α

α	$0,005$	$0,01$	$0,0025$	$0,05$	$0,1$
AD_α	$2,244$	$1,959$	$1,591$	$1,321$	$1,162$

Remark: If the theoretical distribution function $G(x)$ comes from a distribution other than the normal or exponential distribution, the values AD_α for the given significance level α are available on the internet or in other specialized literature. We will not present them here, since we will not deal with these distributions in the Anderson–Darling test.

Example 5.14 *Using the Anderson–Darling test at the significance level α=0.05, let us verify whether the values 10, 20, 35, 4, 30, 1 of a random sample come from an exponential distribution.*

Solution: The sample size is $n = 6$. To estimate the parameter λ we use the sample arithmetic mean $\overline{x}$, which we calculate using relation (71). Thus

$$\lambda = \overline{x} = \frac{1}{6}\,(10 + 20 + 35 + 4 + 30 + 1) = 16,67.$$

We will test

$$H_0 : F(x) = G(x), \text{ where } G(x) \sim Ex(\lambda),$$
$$H_1 : F(x) \neq G(x).$$

We are testing whether the distribution function $F(x)$ of the distribution from which the random sample comes is equal to the distribution function $G(x)$ of the exponential distribution. We determine the value of the test statistic AD using relation (120). To calculate its value we use Table 15, in which we arrange the values of the random sample into a nondecreasing

Table 15

(i)	$x_{(i)}$	$G(x_{(i)})$	$\ln G(x_{(i)})$	$\ln\left[1-G(x_{(n+1-i)})\right]$	$\ln G(x_{(i)})+\ln\left[1-G(x_{(n+1-i)})\right]$
1	1	$0,0582$	$-2,8433$	$-2,0996$	$-4,9432$
2	4	$0,2134$	$-1,5447$	$-1,8000$	$-3,3447$
3	10	$0,4512$	$-0,7959$	$-1,2000$	$-1,9959$
4	20	$0,6988$	$-0,3584$	$-0,6000$	$-0,9584$
5	30	$0,8347$	$-0,1807$	$-0,2400$	$-0,4207$
6	35	$0,8775$	$-0,1306$	$-0,0600$	$-0,1906$

sequence (ranked series).

In column 2 of Table 15 are the values of the random sample arranged into a nondecreasing sequence.

In column 3 of Table 15 are the values $G(x_{(i)})$ of the distribution function of the exponential distribution, which we calculate using the relation for the distribution function of the exponential distribution (60)

$$G(x_{(i)}) = 1 - e^{-\frac{x_{(i)}}{\lambda}} = 1 - e^{-\frac{x_{(i)}}{16,67}}.$$

In column 4 of Table 15 are the values $\ln G(x_{(i)})$ and in column 5 are the values $\ln\left[1-G(x_{(n+1-i)})\right]$.

In column 6 of Table 15 is the sum of the values of columns 4 and 5, i. e. $\ln G(x_{(i)}) + \ln\left[1-G(x_{(n+1-i)})\right]$.

The value of the test statistic AD is determined using relation (120)

$$AD = \frac{1}{6}\Big[-1\cdot(-4,9432) - 3\cdot(-3,3447) - 5\cdot(-1,9959) - 7\cdot(-0,9584) -$$

$$-9\cdot(-0,4207) - 11\cdot(-0,1906)\Big] - 6 = 0,2580.$$

The critical value of the test $AD_\alpha = AD_{0,05} = 1,321$ (Table 14).
The critical region is

$$K_{0,05} = (AD_\alpha;\infty) = (1,321;\infty) \Rightarrow AD \notin K_{0,05}.$$

The value of the test statistic $AD = 0,2580$ does not fall into the critical region $K_{0,05}$, therefore we do not reject the null hypothesis H_0. We can assume that, at the significance level $\alpha = 0,05$, the random sample comes from an exponential distribution.

5.5.4 Shapiro–Wilk normality test

The *Shapiro–Wilk normality test* is the most commonly used test for normality. It is based on determining whether the points of a constructed quantile-quantile plot (Q–Q plot) differ significantly from the regression line fitted through these points. It is used mainly for random samples of sizes $7 \leqq n \leqq 2000$. Consider a random sample V_n, whose values we arrange according to size into a nondecreasing sequence (ranked series)

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

1. Hypotheses:

$$H_0 : F(x) = G(x), \text{ where } G(x) \sim N(\mu, \sigma^2),$$
$$H_1 : F(x) \neq G(x).$$

2. The test statistic is

$$W = \frac{\left(\sum_{i=1}^{m} a_{i,n} \left(x_{(n-i+1)} - x_{(i)} \right) \right)^2}{\sum_{i=1}^{n} \left(x_{(i)} - \overline{x} \right)^2}, \tag{122}$$

   where $a_{i,n}$ are tabulated weights (Table 48), $\overline{x}$ is the sample arithmetic mean and $m = \frac{n}{2}$ for n even, resp. $m = \frac{n-1}{2}$ for n odd.

3. The critical region K_α is

$$K_\alpha = \left(-\infty; W_\alpha(n)\right),$$

   where $W_\alpha(n)$ is the tabulated critical value for the Shapiro–Wilk normality test (Table 49).

Remark: The closer the value of the test statistic W is to 1, the better the agreement between the theoretical and empirical distribution.

Example 5.15 *Using the Shapiro–Wilk normality test at the significance level α=0.05, let us verify whether the values 4.05; 4.15; 4.2; 4.25; 4.3; 4.35; 4.45; 4.55 of a random sample come from a normal distribution.*

Solution: We will test

$$H_0 : F(x) = G(x), \text{ where } G(x) \sim N(\mu, \sigma^2),$$
$$H_1 : F(x) \neq G(x).$$

We are testing whether the distribution function $F(x)$ of the distribution from which the random sample comes is equal to the distribution function $G(x)$ of the normal distribution.

The sample size is $n = 8$. All values are arranged into a nondecreasing sequence. We determine the value of the test statistic W using relation (122), where $m = \frac{n}{2} = 4$, the sample arithmetic mean $\overline{x}$ is calculated using relation (71)

$$\overline{x} = \frac{1}{8}\big(4,05 + 4,15 + 4,2 + 4,25 + 4,3 + 4,3 + 4,45 + 4,55\big) = 4,2875$$

and the tabulated values $a_{i,n}$ are $a_{1,8} = 0,6052$, $a_{2,8} = 0,3164$, $a_{3,8} = 0,1743$, $a_{4,8} = 0,0561$ (Table 48). Then the value of the test statistic W is

$$W = \frac{[0,6052(4,55 - 4,05) + 0,3164(4,45 - 4,15) + 0,1743(4,35 - 4,2) + 0,0561(4,3 - 4,25)]^2}{\sum_{i=1}^{n} \big(x_{(i)} - \overline{x}\big)^2},$$

where

$$\sum_{i=1}^{n} \big(x_{(i)} - \overline{x}\big)^2 = (4,05 - 4,2875)^2 + (4,15 - 4,2875)^2 + (4,2 - 4,2875)^2 + (4,25 - 4,2875)^2 +$$

$$+(4,3 - 4,2875)^2 + (4,35 - 4,2875)^2 + (4,45 - 4,2875)^2 + (4,55 - 4,2875)^2 = 0,1838.$$

Thus $W = 0,99$. The critical value for the Shapiro–Wilk normality test is $W_{0,05}(8) = 0,818$ (Table 49). The critical region is

$$K_{0,05} = (-\infty; 0,818) \Rightarrow W \notin K_{0,05}.$$

The value of the test statistic $W = 0,99$ does not fall into the critical region $K_{0,05}$, therefore at the significance level $\alpha = 0,05$ we do not reject the null hypothesis H_0. We can assume that the values of the random sample come from a normal distribution.

5.6 Tests for extreme values

Sometimes, among the measured values there occur values that differ considerably from the rest. Such values are called *extreme (outlying) values*. Extreme values distort the calculated characteristics of the random sample, which can lead to an erroneous conclusion in the analysis of the given sample. Therefore it is good to detect them and subsequently exclude them from the analyzed sample. A box plot (boxplot) (chapter 2.2) serves for the graphical detection of extreme values. In order to be able to exclude an extreme value from the analyzed sample, it is necessary to statistically test its deviation from the other values. The most commonly used tests for extreme values include Grubbs' test and Dixon's test. Both assume that the distribution of the population is normal and can be used for a small random sample size ($n \leqq 30$).

Consider a random sample V_n that comes from a population with a normal distribution with mean value μ and standard deviation σ. We arrange the values of the random sample according to size into a nondecreasing sequence (ranked series)

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

5.6.1 Grubbs' test

Grubbs' test has two forms depending on whether we are testing the extremeness of the minimum or the maximum value.

Alternative for testing the extremeness of the minimum value:

1. Hypotheses:

$$H_0 : \text{the value } x_{(1)} \text{ is not extreme,}$$
$$H_1 : \text{the value } x_{(1)} \text{ is extreme.}$$

2. The test statistic is

$$T(1) = \frac{\overline{x} - x_{(1)}}{s}\sqrt{\frac{n}{n-1}}, \tag{123}$$

where $\overline{x}$ is the sample arithmetic mean, s is the sample standard deviation and n is the size of the random sample.

3. The critical region K_α is

$$K_\alpha = (T_\alpha(n); \infty),$$

where $T_\alpha(n)$ is the critical value of Grubbs' test, tabulated for $n \leqq 30$ for $\alpha = 0,05$ and $\alpha = 0,01$ (Table 50).

Alternative for testing the extremeness of the maximum value:

1. Hypotheses:

$$H_0 : \text{the value } x_{(n)} \text{ is not extreme,}$$
$$H_1 : \text{the value } x_{(n)} \text{ is extreme.}$$

2. The test statistic is

$$T(n) = \frac{x_{(n)} - \overline{x}}{s}\sqrt{\frac{n}{n-1}}, \tag{124}$$

where $\overline{x}$ is the sample arithmetic mean, s is the sample standard deviation and n is the size of the random sample.

3. The critical region K_α is

$$K_\alpha = (T_\alpha(n); \infty)\,,$$

where $T_\alpha(n)$ is the critical value of Grubbs' test, tabulated for $n \leqq 30$ for $\alpha = 0,05$ and $\alpha = 0,01$ (Table 50).

5.6.2 Dixon's test

Like Grubbs' test, *Dixon's test (Q-test)* also has two forms depending on whether we are testing the extremeness of the minimum or the maximum value. In this case it is not necessary to calculate the sample arithmetic mean and the sample standard deviation of the random sample.

Alternative for testing the extremeness of the minimum value:

1. Hypotheses:

$$H_0 : \text{the value } x_{(1)} \text{ is not extreme,}$$
$$H_1 : \text{the value } x_{(1)} \text{ is extreme.}$$

2. The test statistic is

$$Q(1) = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}. \tag{125}$$

3. The critical region K_α is

$$K_\alpha = (Q_\alpha(n); \infty)\,,$$

where $Q_\alpha(n)$ is the critical value of Dixon's test, tabulated for $n \leqq 30$ for $\alpha = 0,05$ and $\alpha = 0,01$ (Table 51).

Alternative for testing the extremeness of the maximum value:

1. Hypotheses:

$$H_0 : \text{the value } x_{(n)} \text{ is not extreme,}$$
$$H_1 : \text{the value } x_{(n)} \text{ is extreme.}$$

2. The test statistic is
$$Q(n) = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}. \tag{126}$$

3. The critical region K_α is
$$K_\alpha = (Q_\alpha(n); \infty),$$
where $Q_\alpha(n)$ is the critical value of Dixon's test, tabulated for $n \leqq 30$ for $\alpha = 0,05$ and $\alpha = 0,01$ (Table 51).

Example 5.16 *Among the values obtained from a certain measurement 5.83; 5.80; 5.85; 5.88; 5.84; 5.83; 5.96; 5.78; 5.82; 5.81; 5.86; 5.82, let us verify at the significance level α=0.05 whether the value 5.96 is not extreme. We assume that the values follow a normal distribution. Let us use*
a) Grubbs' test,
b) Dixon's test.

Solution: We arrange all values into a nondecreasing sequence: 5, 78; 5, 80; 5, 81; 5, 82; 5, 82; 5, 83; 5, 83; 5, 84; 5, 85; 5, 86; 5, 88; 5, 96. For a quick check of whether this is an extreme value, we can also use a boxplot (Fig. 24).

In this case we are going to test the extremeness of the maximum value. The sample size is $n = 12$.
We will test
$$H_0 : \text{the value } x_{(12)} = 5,96 \text{ is not extreme,}$$
$$H_1 : \text{the value } x_{(12)} = 5,96 \text{ is extreme.}$$
a) We determine the value of the test statistic $T(12)$ for Grubbs' test using relation (124)
$$T(12) = \frac{x_{(12)} - \overline{x}}{s}\sqrt{\frac{12}{12-1}},$$
where the sample arithmetic mean $\overline{x}$ is calculated using relation (71)
$$\overline{x} = \frac{1}{12}(5,78 + 5,80 + 5,81 + 5,82 + 5,82 + 5,83 + 5,83+$$
$$+5,84 + 5,85 + 5,86 + 5,88 + 5,96) = 5,84$$

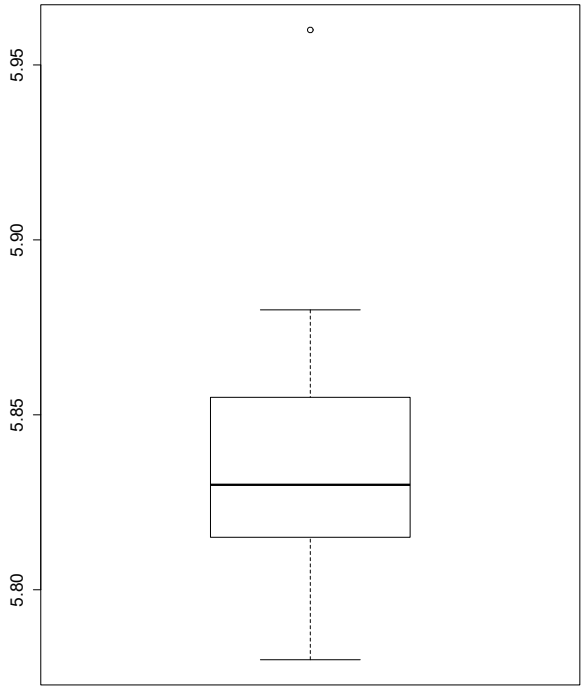


Figure 24: Boxplot for Example 5.16

and the sample standard deviation s is calculated using relation (77). For the sample variance s^2 we have (73)

$$s^2 = \frac{1}{11}\big[(5,78-5,84)^2+(5,80-5,84)^2+(5,81-5,84)^2+2.(5,82-5,84)^2+$$
$$+2.(5,83-5,84)^2+(5,84-5,84)^2+(5,85-5,84)^2+(5,86-5,84)^2+$$
$$+(5,88-5,84)^2+(5,96-5,84)^2\big]=0,00214 \Rightarrow s=\sqrt{s^2}=0,0463.$$

The value of the test statistic for Grubbs' test is $T(12) = 2,7059$. The critical value is $T_{0,05}(12) = 2,387$ (Table 50). The critical region is

$$K_{0,05} = (2,387;\infty) \Rightarrow T(12) \in K_{0,05}.$$

The value of the test statistic $T(12) = 2,7059$ falls into the critical region $K_{0,05}$, therefore at the significance level $\alpha = 0,05$ we reject the null hypothesis H_0 and accept the alternative hypothesis H_1. Based on Grubbs' test we can assume that the value $x_{(12)} = 5,96$ is extreme.

b) We determine the value of the test statistic $Q(12)$ for Dixon's test using relation (126)

$$Q(12) = \frac{x_{(12)} - x_{(11)}}{x_{(12)} - x_{(1)}} = \frac{5,96-5,88}{5,96-5,78} = 0,4444.$$

The critical value of Dixon's test is $Q_{0,05}(12) = 0,376$ (Table 51). The critical region is

$$K_{0,05} = (0,376; \infty) \Rightarrow Q(12) \in K_{0,05}.$$

The value of the test statistic $Q(12) = 0,4444$ falls into the critical region $K_{0,05}$, therefore at the significance level $\alpha = 0,05$ we reject the null hypothesis H_0 and accept the alternative hypothesis H_1. Based on Dixon's test we can assume that the value $x_{(12)} = 5,96$ is extreme.