GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 8 • Number 1 • 2014

http://dx.doi.org/10.7494/geom.2014.8.1.21

Waldemar Krupiński*

Statistical Methods Applied in Evaluating the Reliability of Land Surveying Equipment

1. Introduction

In solving practical land surveying issues related to testing the reliability of surveying equipment a need arises to verify statistical hypotheses regarding both the parameters and the mathematical distribution of the probability of errors in geodetic observations.

Two separate examples should be mentioned:

1) a statistical hypothesis specifies both the type and the distribution of its parameters,
2) a hypothesis is applied only to the mathematical form of the distribution function while the parameters are estimated using one of the methods of estimation [5].

To solve the above questions, statistical tests of conformity and identity may be used. They consist in comparing the empirical distribution (land surveying) with an assumed theoretical distribution, usually – a normal distribution.

A null hypothesis H_0 is proposed, concerning the compliance of the compared distributions, against an alternative hypothesis H_1 stating non-compliance of the distributions [6].

The compared distributions – empirical and theoretical – will generally differ, but the differences, in case the hypothesis H_0 is correct, should not be too large.

For the purposes of research, a characteristic U [4, 5], is constructed, acting here as a measure of differences between the compared distributions. Then, the critical region S_{cr} is established that meets the following condition:

$$P(u \in S_{cr}) = P_{cr} \quad (1)$$

where P_{cr} is the probability of rejection of H_0 in case it is correct; usually $P_{cr} = 0.05$ is assumed.

* Rzeszów School of Engineering and Economics, Rzeszów, Poland

Also in the case when u calculated for a specified random sample is found within the critical area, hypothesis H_0 on the conformity of the compared distributions must be rejected.

When $u \notin S_{cr}$ – there are no grounds to reject H_0.

Among different statistical tests, the most well-known are the following conformity tests: χ^2 Pearson and λ Kolmogorov.

To test the accuracy of the measurements we performed by means of different instruments, some less well-known tests were used, namely, tests of identity:

- the Smirnov–Kolmogorov test,
- the rank sum test,
- the confidence intervals test.

2. The Smirnov–Kolmogorov Test

The Smirnov–Kolmogorov identity test is a statistical test of error distributions of two populations, based on two random samples with numbers n_1 and n_2. With empirical distribution functions marked as $S_{n1}(x)$ and $S_{n2}(x)$, respectively for the first and second sample – the tests statistic is:

$$D_{n1',n2} = \sup | S_{n1}(x) - S_{n2}(x) | \tag{2}$$

where:

$D_{n1',n2}$ – calculated test statistic,
n_1 – number of the first sample,
n_2 – number of the second sample,
x – observations.

The critical set is the interval:

$$[\mathrm{d}(\alpha, n_1, n_2); 1].$$

Critical values $\mathrm{d}(\alpha, n_1, n_2)$ multiplied by n_1 and n_2 for the significance level $\alpha = 0.05$ can be found in statistical tables.

If values $D_{n1',n2}$ are not within critical set, there are no grounds for rejecting the hypothesis on the normality of error distribution at significance level α.

Example 1

The test in this study will be used to verify the accuracy of angle measurements performed with the electronic total station Leica TS 02. The use of tests of identity in land surveying involves comparing the results of the measurements obtained with the tested device with the results from a previously verified instrument which operates correctly.

For this purpose, the test instrument, Leica TS 02, was used to measure horizontal angles (in two runs at different alignments of the telescope) 26 times in a triangle, thereby obtaining deviations of 26 “triangles” and comparing them with the deviations from measurements of these same angles prior tested by means of a total station Topcon GTS-220 [2]. Measurements by the total station Topcon GTS-220 were performed in three series at different alignments of the telescope, on the same triangle base, and the collected results did not establish the presence of systematic errors in these measurements.

The structured results obtained in both tests are presented in Table 1

Table 1. Structured results of both samples

x [^{g}]	Number		Cumulative number		$S_{n1}(x)$	$S_{n2}(x)$	$\lvert S_{n1}(x) - S_{n2}(x) \rvert$
	1 sample Topcon	2 sample Leica	1 sample Topcon	2 sample Leica			
– 0.00180	0	1	0	1	0.0000	0.0556	0.0556
– 0.00175	0	1	0	2	0.0000	0.1111	0.1111
– 0.00160	1	0	1	2	0.0556	0.1111	0.0556
– 0.00145	1	0	2	2	0.1111	0.1111	0.0000
– 0.00140	1	1	3	3	0.1667	0.1667	0.0000
– 0.00125	1	0	4	3	0.2222	0.1667	0.0556
– 0.00090	1	0	5	3	0.2778	0.1667	0.1111
– 0.00075	0	1	5	4	0.2778	0.2222	0.0556
– 0.00050	1	0	6	4	0.3333	0.2222	0.1111
– 0.00035	1	1	7	5	0.3889	0.2778	0.1111
– 0.00025	1	0	8	5	0.4444	0.2778	0.1667
0.00015	2	0	10	5	0.5556	0.2778	**0.2778***
0.00030	0	2	10	7	0.5556	0.3889	0.1667
0.00060	0	1	10	8	0.5556	0.4444	0.1111
0.00075	1	1	11	9	0.6111	0.5000	0.1111
0.00085	0	1	11	10	0.6111	0.5556	0.0556
0.00100	2	3	13	13	0.7222	0.7222	0.0000
0.00115	1	0	14	13	0.7778	0.7222	0.0556
0.00120	1	0	15	13	0.8333	0.7222	0.1111
0.00130	1	1	16	14	0.8889	0.7778	0.1111
0.00145	0	1	16	15	0.8889	0.8333	0.0556
0.00150	1	0	17	15	0.9444	0.8333	0.1111
0.00175	0	1	17	16	0.9444	0.8889	0.0556
0.00185	1	0	18	16	1.0000	0.8889	0.1111
0.00195	0	1	18	17	1.0000	0.9444	0.0556
0.00215	0	1	18	18	1.0000	1.0000	0.0000
total	18	18					

* maximum difference of distribution

Theoretical distribution leaps: 1:18 = 0.0556.
Smirnov–Kolmogorov critical value tables provide as follows:

$$n_1 \cdot n_2 \cdot \mathrm{d}(\alpha, n_{1,} n_2) = 18 \cdot 18 \cdot \mathrm{d}(0.05, 18, 18) = 162,$$

it results, therefore, that d(0.05, 18, 18) = 0.5.

Thus, the critical interval is [0.5, 1], and the maximum difference of distribution is 0.2778, which is not within the critical interval, which means that there is no reason to reject the hypothesis of conformity of distributions at the significance level $\alpha = 0.05$.

The final conclusion of the test is, therefore, the absence of significant measurement errors, i.e. the tested total station Leica TS 02 is suitable for measuring angles.

The correctness of angle measurements made by Leica TS 02 was also verified by the Shapiro–Wilk test. According to [1], Kolmogorov and Pearson test do not use all the information that can be obtained from the present sample. In both tests, information is lost due to grouping the observations into classes. Moreover, χ^2 Pearson's test does not take into account the differences ($n_i - np_i$) which are included in the test statistics, χ^2 while the Kolmogorov test is based on only one difference, namely, the maximum absolute value of the difference D_n – the test statistics. In contrast, the conformity test of random variable distribution from the sample with a normal distribution using the full information from the sample is the Shapiro–Wilk test [1].

Considering the above, to control the distribution of angular measurement errors made by Leica Total Station TS 02 – an additional Shapiro–Wilk statistical test of conformity will be applied.

In this test, the Shapiro–Wilk test statistic W_d shall be used for the verification of the hypothesis of normal distribution of measurement errors – a random variable defined by the following formula [6]:

$$W_d = \frac{\left(\sum_i a_i(n) \cdot \left(X_{n-i+1} - X_i\right)\right)^2}{\sum_i (X_j - \overline{X})^2} \quad i = 1, \ldots, \frac{n}{2} \tag{3}$$

where:

- X_j) – sample element values,
- $a_i(n)$ – constants dependent on sample size and the value i (summarized in Shapiro–Wilk tables).

If the statistical value $W_d W$ is within a critical area in the Shapiro–Wilk test, the hypothesis of normality is rejected at the significance level α.

Otherwise, there is no reason to reject the hypothesis H_0. Structured results for “triangles” and calculations (for Example 1) are shown in Table 2.

Table 2. Structured results for ″triangles″ and calculations for Example 1

No.	X [cc]	$X_{n-i+1}-X_i$	$a_i(n)$	$a_i(n)\cdot(X_{n-i+1}-X_i)$	$X_j-\bar{X}$	$(X_j-\overline{X})^2$
1	–18.0	39.5	0.4407	17.40765	–20.2885	411.6232
2	–17.5	37.0	0.3043	11.25910	–19.7885	391.5847
3	–16.0	34.5	0.2533	8.73885	–18.2885	334.4692
4	–14.5	32.0	0.2151	6.88320	–16.7885	281.8537
5	–14.0	29.0	0.1836	5.32440	–16.2885	265.3152
6	–12.5	27.0	0.1563	4.22010	–14.7885	218.6997
7	–9.0	22.0	0.1316	2.89520	–11.2885	127.4302
8	–7.5	19.5	0.1089	2.12355	–9.7885	95.8147
9	–5.0	16.5	0.0876	1.44540	–7.2885	53.1222
10	–3.5	13.5	0.0672	0.90720	–5.7885	33.5067
11	–2.5	11.0	0.0476	0.52360	–4.7885	22.9297
12	1.5	6.0	0.0284	0.17040	–0.7885	0.6217
13	3.0	3.0	0.0094	0.02820	0.7115	0.5062
14	6.0	–	–	–	3.7115	13.7752
15	7.5	–	–	–	5.2115	27.1597
16	8.5	–	–	–	6.2115	38.5827
17	10.0	–	–	–	7.7115	59.4672
18	11.5	–	–	–	9.2115	84.8517
19	12.0	–	–	–	9.7115	94.3132
20	13.0	–	–	–	10.7115	114.7362
21	14.5	–	–	–	12.2115	149.1207
22	15.0	–	–	–	12.7115	161.5822
23	17.5	–	–	–	15.2115	231.3897
24	18.5	–	–	–	16.2115	262.8127
25	19.5	–	–	–	17.2115	296.2357
26	21.5	–	–	–	19.2115	369.0817

$\bar{X} = 2.2885$ $\sum_{i=1}^{13} 61.92685$ $\sum (X_j - \overline{X})^2 = 4140.5857$

$$\left[\sum_i a_i(n)\cdot\left(X_{n-i+1}-X_i\right)\right]^2 = 3834.9348,$$

$$W_d = \frac{3834.9348}{4140.5857} = 0.926.$$

The critical range in the Shapiro–Wilk test is the interval [0; $W_{\alpha,n}$]. In this case, $W_{\alpha,n} = W_{0.05;26} = 0.920$ which can be found in Shapiro–Wilk statistical tables.

Since the calculated statistical value $W_d = 0.926$ is outside the interval [0; 920], there is no reason for rejection H_0, with a normal distribution of measurement errors, i.e. the tested instrument is suitable for use in measurments.

This result confirms the conclusion from the Smirnov–Kolmogorov test.

3. T-test Checksum

In order to verify whether the two random samples of the following observation numbers: the first one n_1, the second n_2, are characterized by the same distribution, the following steps are performed:

1. All observations from the two samples are set in a range of non-decreasing values and numbered, thus giving them rank. One can, then, calculate the sum of ranks of the elements for each sample separately. So, the value T_m is acquired as the sum of consecutive numbers for the sample with a smaller size.
2. The checked null hypothesis was the assumption that there is no difference between the distributions of both general populations from which the samples were taken, so that the sum of the ranks of the smaller sample (T_m) is to the sum of the ranks with a larger sample size, as the number of observations from a smaller sample (n_1) is to the number of observations in the greater sample (n_2) [8].
3. To test this hypothesis, we compare the resulting sum of the ranks of the smaller sample (T_m) to the value T which is found in the tables of critical values for the T-test, at a significance level $\alpha = 0.05$. The critical region is the test interval T (0; T_α) – in this case (0; $T_{0.05}$).

Example 2

In order to verify whether the height measurements made by the tested level Topcon AT G7N are not subject to systematic errors, 12 leveling measurements were carried out "from the middle" at a closed string of the length of approximately 1 km [3].

Thus, $n_1 = 12$ mesh values were obtained.

If the value T_m is larger than T_α there is no reason to rejetc H_0.

In order to verify H_0 with equal distributions of errors of both series, in like manner $n_2 = 15$ height measurements were carried out with an electronic level Leica Sprinter 150 prior tested to be operating correctly.

The results of observation (mesh) acquired are presented in Table 3.

Table 3. The results of observation for Example 2

No.	X_i [mm] Topcon	X_i [mm] Leica Sprinter
1	–4.0	+3.0
2	+2.5	–1.5
3	+2.0	0.0
4	–1.0	+2.5
5	–3.0	–2.0
6	+0.5	–1.0
7	–3.5	–0.5
8	+3.0	+1.0
9	–1.0	–1.5
10	0.0	+3.0
11	–3.5	–3.0
12	+3.0	0.0
13	–	–2.5
14	–	–2.0
15	–	+1.5

Ascending results of both tests are as follows (middle row contains the results of measurements, the top row – marking a random sample, the lower row – the rank given):

I	I	I	I	II	II	II	II	II	II	I	I	II	II
–4.0	–3.5	–3.5	–3.0	–3.0	–2.5	–2.0	–2.0	–1.5	–1.5	–1.0	–1.0	–1.0	–0.5
1	2	3	4.5	4.5	6	7	8	9	10	12	12	12	14
I	II	II	I	II	II	I	I	II	I	I	II	II	
0.0	0.0	0.0	0.5	1.0	1.5	2.0	2.5	2.5	3.0	3.0	3.0	3.0	
16	16	16	18	19	20	21	22.5	22.5	25.5	25.5	25.5	25.5	

Deviation –3.5 occurs in two cases, but because both are from the same series of measurements – they are numbered sequentially (numbers 2 and 3). Deviation –1.0 occurs in three cases, but it occurs twice in series 1 and once with series 2. These three results should be assigned numbers 11, 12 and 13. Since, however, it does not matter in what order we put the results of series 1 and 2 – we computed a "common" number for these three results:

$$\frac{11+12+13}{3}=\frac{36}{3}=12.$$

Total rank for the first (smaller) sample is:

$$T_m = 1 + 2 + 3 + 4.5 + 12 + 12 + 16 + 18 + 21 + 22.5 + 25.5 + 25.5 = 163.$$

Total rank for the second sample is:

$T_2 = 4.5 + 6 + 7 + 8 + 9 + 10 + 12 + 14 + 16 + 16 + 19 + 20 + 22.5 + 25.5 + 25.5 = 215.$

The sum of the two total ranks is:

$$163 + 215 = 378.$$

The sum of consecutive numbers from 1 to 27 is equal to:

$$s = \frac{n(n+1)}{2} = \frac{27 \cdot 28}{2} = 378.$$

The compatibility of these sums is a checksum of the calculations made.

Sum T_m i.e. the sum of ranks for the smaller sample is 163, and the value T from the test tables for $n_1 = 12$ and $n_2 = 15$, at a significance level $\alpha = 0.05$ is 127.

Therefore:

$$T_{0.05} = 127 < T_m = 163.$$

Thus, here is no basis to reject the null hypothesis H_0, i.e. the distributions of populations represented by the two samples are identical. If T_m was less than 127, the sum of ranks of the larger sample would be greater than 251, as 378 – 127 = 251.

The difference of these sums would be too large, because the sum of the ranks should be in the same ratio as the number of tests, namely:

$$\frac{n_1}{n_2} = \frac{12}{15} = 0.80 \qquad \text{and} \qquad \frac{T_m}{T_2} = \frac{163}{215} \cong 0.76.$$

which is to a very large extent fulfilled.

In the event that $T_{0.05}$ was higher than T_m, would be:

$$\frac{T_m}{T_2} = \frac{127}{215} \cong 0.51,$$

which is significantly different from 0.80.

Thence the inequality:

$$T_{0.05} = 127 < T_m = 163$$

indicates a lack of evidence to reject H_0, i.e. the measurements do not include systematic errors and the tested level Topcon AT G7N is suitable for measurements without special rectification actions.

4. Detection of Systematic Errors Using Confidence Intervals

In order to check whether the results of measurements made by the tested equipment are not subject to systematic errors, we compared the results of measurements taken with the instruments with the results of measurements made by previously proven, correctly operating instrument.

Analyzing the results of the two series of measurements one can determine whether the tested differences do not exceed levels that would indicate the presence of systematic errors in the test equipment [4].

To this end, we constructed confidence intervals for a selected confidence level for the average value. On the basis of the value Q calculated from the two series of measurements, we created two confidence intervals separately for each series, as their boundaries would be different due to the different average values x_1 and + x_2 (the ranges would be the same if the average of both series is equal; $x_1 = x_2$).

For the calculations, we assume the null hypothesis H_0 that the average values are homogeneous (derived from the general population).

If the intervals partially overlap, the hypothesis H_0 can be considered true. In the absence of common areas in the intervals, the alternative hypothesis H_1 should be deemed as true, consisting in that the averages are not homogeneous, and hence that there are systematic errors in measurements.

Example 3

In two series, the distance between two stabilized points was measured 30 times. The first series of results was obtained from a previously proved Topcon GTS-220 total station, the second series – from the testes Leica TS 02 level [2]. The results are presented in the Table 4.

Table 4. Results of measurements for both series

No.	Series I distance; Topcon GTS-220			Series II distance; Leica TS 02		
	[m]	$x_i - \bar{x}$ [mm]	$(x_i - \bar{x})^2$ [mm^2]	[m]	$x_i - \bar{x}$ [mm]	$(x_i - \bar{x})^2$ [mm^2]
1	83.693	–0.5	0.25	83.687	1.3	1.69
2	83.694	0.5	0.25	83.685	–0.7	0.49
3	83.693	–0.5	0.25	83.685	–0.7	0.49
4	83.693	–0.5	0.25	83.686	0.3	0.09
5	83.693	–0.5	0.25	83.686	0.3	0.09
6	83.693	–0.5	0.25	83.686	0.3	0.09
7	83.693	–0.5	0.25	83.686	0.3	0.09
8	83.694	0.5	0.25	83.685	–0.7	0.49
9	83.694	0.5	0.25	83.685	–0.7	0.49
10	83.694	0.5	0.25	83.687	1.3	1.69
11	83.694	0.5	0.25	83.685	–0.7	0.49
12	83.694	0.5	0.25	83.685	–0.7	0.49
13	83.693	–0.5	0.25	83.685	–0.7	0.49

Table 4. cont.

No.	Series I			Series II		
	distance; Topcon GTS-220			distance; Leica TS 02		
	[m]	$x_i - \bar{x}$ [mm]	$(x_i - \bar{x})^2$ [mm²]	[m]	$x_i - \bar{x}$ [mm]	$(x_i - \bar{x})^2$ [mm²]
14	83.693	–0.5	0.25	83.685	–0.7	0.49
15	83.694	0.5	0.25	83.686	0.3	0.09
16	83.694	0.5	0.25	83.686	0.3	0.09
17	83.694	0.5	0.25	83.686	0.3	0.09
18	83.694	0.5	0.25	83.686	0.3	0.09
19	83.694	0.5	0.25	83.686	0.3	0.09
20	83.693	–0.5	0.25	83.685	–0.7	0.49
21	83.693	–0.5	0.25	83.685	–0.7	0.49
22	83.693	–0.5	0.25	83.686	0.3	0.09
23	83.693	–0.5	0.25	83.685	–0.7	0.49
24	83.693	–0.5	0.25	83.685	–0.7	0.49
25	83.693	–0.5	0.25	83.685	–0.7	0.49
26	83.693	–0.5	0.25	83.686	0.3	0.09
27	83.693	–0.5	0.25	83.686	0.3	0.09
28	83.694	0.5	0.25	83.687	1.3	1.69
29	83.694	0.5	0.25	83.687	1.3	1.69
30	83.694	0.5	0.25	83.686	0.3	0.09
$\bar{x}$	83.6935	–1.0	7.500	83.6857	0.0	14.300

Mean error for individual observations is:

- for measurements by Topcon GTS-220 (series I): $\bar{m}_I = 0.5085$ mm
- for measurements by Leica TS 02 (series II): $\bar{m}_{II} = 0.7022$ mm

$$\bar{x}_{II} - \bar{x}_I = -0.0078 \text{ m},$$

$$\alpha = 0.05.$$

For both series, we construct the confidence intervals for mean value:

$$Q_{X_I} = \frac{\bar{m}_I}{\sqrt{n_I}} = 0.09 \text{ mm}; \qquad Q_{X_{II}} = \frac{\bar{m}_{II}}{\sqrt{n_{II}}} = 0.13 \text{ mm}.$$

Student's distribution tables provide the value $t_{0.05/29} = 2.045$.

The confidence interval of the mean for series 1 shall be:

$$P(\bar{x}_I - t_{\alpha_I/n-1} \cdot Q_{X_I} \le \alpha \le \bar{x}_I + t_{\alpha_I/n-1} \cdot Q_{X_I}) = 1 - \alpha \tag{4}$$

$$P(83.6935 - 2.045 \cdot 0.00009 \le \alpha \le 83.6935 + 2.045 \cdot 0.00009) = 0.95,$$

$$P(83.6933 \le \alpha \le 83.6937) = 0.95.$$

The confidence interval of the mean for series 2 shall be:

$$P(\overline{x}_{II} - t_{\alpha_{II}/n-1} \cdot Q_{X_{II}} \leq \alpha \leq \overline{x}_{II} + t_{\alpha_{II}/n-1} \cdot Q_{X_{II}}) = 1 - \alpha \quad (5)$$

$$P(83.6857 - 2.045 \cdot 0.00013 \leq \alpha \leq 83.6857 + 2.045 \cdot 0.00013) = 0.95\,,$$

$$P(83.6854 \leq \alpha \leq 83.6860) = 0.95\,.$$

These intervals do not overlap, which implied the need to reject H_0 at level $\alpha = 0.05$, i.e. that distance measurements made by the test instrument involve systematic errors.

5. Conclusions

1. The study of angle measurements made by the electronic total station Leica TS 02 showed, using the Kolmogorov–Smirnov test of identity, an absence of systematic errors.
2. The conducted additional test results from the same measurements using the Shapiro–Wilk test of conformity confirmed the conclusion from the Smirnov–Kolmogorov test about the absence of systematic errors.
3. Used to test the accuracy of the leveling measurements, the checksum T-rank identity test lead to the conclusion of the absence of systematic errors, and thus on the correctness of operation of the tested Leica Sprinter 150 M.
4. Accuracy studies of distance measurements made by Leica Total Station TS 02, carried out a according to a statistical test of confidence intervals, showed that the tested instrument should be rectified before performing measurements of distance.
5. Due to the fact that the Shapiro–Wilk test statistics W_d adopted a value very close to the critical area of the test $W_d = 0.926$ and the critical area is the interval [0, 920]), we can conclude that the Shapiro–Wilk test is "stronger" than the Smirnov–Kolmogorov test, which in turn leads to the conclusion that it is important to properly select statistical tests to research.
6. The research has led to the general conclusion that in addition to conformity testing of land surveying equipment one can also apply statistical tests of identity.

References

[1] Banasik P., Cichociński P., Czaja J., Góral W., Kozioł K., Krzyżek R., Kudrys J., Ligas M., Skorupa B.: *Podstawy geomatyki*. Wydawnictwa AGH, Kraków 2011.

[2] Ćwikła T.: *Badanie poprawności pomiarów kątowych wykonywanych tachimetrem Leica TS 02, z pomocą testów statystycznych* [unpublished engineering work prepared under the supervision of W. Krupiński]. UR, Kraków 2012.

[3] Gawlik J., Stachnik K.: *Statystyczne sposoby kontroli poprawności działania instrumentów i metod pomiarów geodezyjnych* [unpublished engineering work prepared under the supervision of W. Krupiński]. WSIE, Rzeszów 2009.
[4] Greń J.: *Modele i zadania statystyki matematycznej*. PWN, Warszawa 1970.
[5] Kaczmarek Z.: *Metody statystyczne hydrologii i meteorologii*. WKiŁ, Warszawa 1970.
[6] Ney B.: *Kryteria zgodności rozkładów empirycznych z modelami*. ZN PAN, Geodezja 7, Kraków 1970.
[7] Piasek Z.: *Geodezja budowlana dla inżynierii środowiska*. ZN PK, Kraków 2000.
[8] Puchalski T.: *Statystyka*. PWN, Warszawa 1969.