10.46299/ISG.2022.MONO.TECH.4.1.3

1.3 Probabilistic – statistical sequential analysis of the results of geodetic measurements

1.3.1 The essence of sequential analysis. Limit standards

When performing topographic surveying, when performing planning, construction - installation and other works, it is necessary to perform control - geodetic measurements of the quality of their performance. Until now, control - geodetic measurements have been established for various reasons, which did not have a very specific scientific approach.

The most correct will be the performance of control - geodetic measurements by the method of probabilistic - statistical sequential analysis, the basics of which were developed by Wald [21]. Sequential analysis has already been used to optimally determine the methods of angular measurements in triangulation [24]. But it should be noted that the estimated number of measurements in triangulation (polygonometry, trilateration, *GPS*) is known in advance, it is given in the instructions [27].

When using control - geodetic measurements, their number is unknown in advance, for this purpose the theory of the consecutive analysis with use of Markov random processes is finished.

The essence of consistent analysis.

For each number of measurements, we calculate the acceptance and rejection numbers in advance according to the formulas [21].

$$\left\{\begin{aligned} a_\nu &= \frac{\sigma_1^2\times\sigma_2^2}{\sigma_2^2-\sigma_1^2}\left[2\times ln\frac{\beta}{1-\alpha}+\left(\frac{\sigma_2}{\sigma_1}\right)^2\right] \\ r_\nu &= \frac{\sigma_1^2\times\sigma_2^2}{\sigma_2^2-\sigma_1^2}\left[2\times ln\frac{1-\beta}{\alpha}+\left(\frac{\sigma_2}{\sigma_1}\right)^2\right] \end{aligned}\right\} \qquad (1)$$

In these formulas: σ_1, σ_2 - standard standard deviations of measurements, respectively under favorable and unfavorable conditions, which will be the limit standards in the future; α - reject works when in fact they are of good quality; β - to accept works of good quality, when in fact they are subject to rejection; ν is the number of degrees of freedom, if the actual value of the quantity (measured) is monitored, then

$v = n$, where n is the number of measurements (if the actual value is unknown, then ν $= n - 1$).

After each measurement, the accumulated sum of squares of deviations $[\delta^2]$ of control measurements xi from their actual value $M(x)$ or from the arithmetic mean xcp is calculated, after which this sum is compared with the acceptance and rejection numbers, with the appropriate number of measurements.

If $[\delta_2]\ v \le a_v$, the measurements are completed and the work performed is considered to be of good quality. In the case when $a_v \le [\delta^2]_v \le r_v$, then the control work continues and if it turns out that $[\delta^2]_v \ge r_v$ - the work performed is missing. The deviation of δi, with an unknown real value of the quantity being measured, and the deviation of a known real value, denoted by Δ_i, are from the relations:

$$\begin{Bmatrix} \delta_i = x_i - x_{cp} \\ \Delta_i = x_i - M(x) \end{Bmatrix} \qquad i = 1, n. \tag{2}$$

Expressions (1) do not assume that the values of σ_1 and σ_2 are standards, since the standards under the appropriate conditions are constant values that are from the general population, so there can be no two standards, and it would be correct to call these values σ_1 and σ_2 (1) empirical standards, or root mean square errors, but more simply called marginal standards.

Wald [21] proposes to set the values of σ_1 and σ_2 in advance for the same measurements for all dimensions, in which case expression (1) will be linear, but the acceptance and rejection numbers are nonlinear, which is more in line with practical and theoretical requirements.

The difference between the defective and receiving numbers calculated by expressions (1), for any number of measurements will be a constant value, in fact, this phenomenon does not agree with theory and practice. It has long been known that with the increase in the number of measurements, their variance increases, which is confirmed in practice, so it is more logical to assume that the difference between receiving and missing numbers, with increasing number of measurements should increase

It is proposed to find the limit standards when constructing confidence intervals, for the standard, for each number of measurements, or the number of degrees of freedom. The instructions [27] do not specify a standard but a threshold from which you can jump to a standard. The confidence interval of the standard is built in the form of equations [22].

$$Z_1 \times m(x) \le \sigma(x) \le Z_2 \times m(x), \tag{3}$$

де Z_1 та Z_2 – coefficients determined by the formulas:

$$Z_1 = \frac{\sqrt{\nu}}{\chi_{2,\nu}}; \qquad Z_2 = \frac{\sqrt{\nu}}{\chi_{1,\nu}}; \tag{4}$$

$\chi_{1,\nu}$; $\chi_{2,\nu}$ – values selected from the distribution table χ2 (chi-square) depending on the confidence interval *p* and the number of degrees of freedom *v*; *m* (*x*) - root mean square measurement error; *σ* (*x*) is the standard deviation of the standard (standard).

We do not know the value of *m* (*x*), but as the number of measurements increases, it approaches the standard and we can assume that *m* (*x*) ≈ *σ* (*x*), then the confidence interval will look like:

$$Z_1 \times \sigma(x) \le \sigma(x) \le Z_2 \times \sigma(x) \tag{5}$$

Since the left-hand side of equation (5) is less than *σ* (*x*) and the right-hand side is larger, which will correspond to good and unfavorable conditions of observations, these extreme parts of equations will be equal to the limit standards, ie

$$\sigma_{1,\nu} = Z_1 \times \sigma(x); \quad \sigma_{2,\nu} = Z_2 \times \sigma(x) \tag{6}$$

The value of the standard *σ* (*x*) is set or calculated according to the limit tolerance specified by the instruction, substituting (4) in expression (6) and squaring the limit standards we obtain

$$\sigma_{1.\nu}^2 = \nu \times \sigma^2(x)/\chi_{2.\nu}^2 \ ; \quad \sigma_{2.\nu}^2 = \nu \times \sigma^2(x)/\chi_{1.\nu}^2. \tag{7}$$

Given (7), the equations of acceptance and defective numbers (1) will finally take the form [21]

$$a_\nu = \left(2ln^{\beta}\big/_{(1-\alpha)} + \nu ln^{\chi_{2,\nu}^2}\big/_{\chi_{1.\nu}^2}\right) \nu\sigma^2(x)\big/_{(\chi_{2.\nu}^2 - \chi_{1.\nu}^2)}$$

$$r_\nu = \left(2ln\,{}^{(1-\beta)}\!/_{\alpha} + \nu ln\,{}^{\chi^2_{2,\nu}}\!/_{\chi^2_{1.\nu}}\right)\,{}^{\nu\sigma^2(x)}\!/_{(\chi^2_{2.\nu}-\chi^2_{1.\nu})}.$$

(8)

Equation (8) will not be equations of lines, but will be equations of curves.

Given the values *α*, *β* of defective and receiving numbers, standard deviation *σ* (*x*), and the confidence interval p to find the values of $\chi^2_{1.\nu}$ and $\chi^2_{2.\nu}$ we can calculate the acceptance and rejection numbers for any how many control measurements. The values of $\chi^2_{1.\nu}$ and $\chi^2_{2.\nu}$ are chosen from the table χ2-distribution according to the confidence probabilities p_1 = 1-0.5 (1-*p*), p_2 = 0.5 (1-*p*) and depending on the number of degrees of freedom *ν*.

1.3.2 Clarification of the value of χ2 - distribution

To calculate the acceptance and rejection numbers according to formula (8), the number of significant digits of the values $\chi^2_{1.\nu}$ and $\chi^2_{2.\nu}$ must be at least six. However, even in the exact tables [28] the number of significant digits is three to four, which does not satisfy the required accuracy of determining the acceptance and rejection numbers, so it is necessary to calculate more accurate values $\chi^2_{1.\nu}$ and $\chi^2_{1.\nu}$, in addition, these values will be needed in determining the number of degrees of freedom or the number of measurements at which the process of sequential analysis will end with almost a probability close to unity, which is important when planning the process of control - geodetic measurements.

For the values of $\chi^2_{i.1} < 1$ and the number of degrees of freedom ν = 1 the probability is calculated by the formula [28]:

$$P(\chi^2_{i.1}) = 2 \times \left[1 - \phi\left(\sqrt{\chi^2_{i.1}}\right)\right]\ ,\ i = 1,2 \tag{9}$$

where $\phi\left(\sqrt{\chi^2_{i.1}}\right)$ –is a Laplace function of the random variable $\sqrt{\chi^2_{i.1}}$, the exact values

of which are in [28]. From formula (9) we find that

$$\phi\left(\sqrt{\chi^2_{i.1}}\right) = 1 - 0{,}5p(\chi^2_{i.1}). \tag{10}$$

For example, for a confidence probability p = 0.95, the probability p_1 = 1-0.5 (1-p) = 0.975, for which we find the value of $\chi^2_{1.1}$, then by inverse linear interpolation on the value of the function 0.1525 we find the argument $\sqrt{\chi^2_{1.1} = 0{,}03133835}$ and magnitude $\chi^2_{1.1} = 0.000982092$.

To find a more accurate value of $\chi^2_{2.1}$ and subsequent values, the following technique is proposed.

If it is necessary to determine the probability from a given value of χ^2, it can be determined in the table of the probability integral χ^2 [28], using Bessel's formula for quadratic interpolation:

$$P(\chi^2_\nu) = P\left(\chi^2_{0,\nu}\right) + U * \Delta P\left(\chi^2_{0,\nu}\right) - \frac{U(1-U)}{2} * \frac{\Delta P\left(\chi^2_{+1,\nu}\right) - \Delta P\left(\chi^2_{-1,\nu}\right)}{2} \qquad (11)$$

Where $\chi^2_{-1,\nu}$; $\chi^2_{0,\nu}$; $\chi^2_{+1,\nu}$ – equidistant tabular values of the argument, and $\chi^2_{0,\nu} \leq \chi^2_\nu < \chi^2_{+1,\nu}$.,

$$U = \frac{\chi^2_\nu - \chi^2_{0,\nu}}{\chi^2_{+1,\nu} - \chi^2_{0,\nu}}\,; \qquad (12)$$

$$\Delta P\left(\chi^2_{j,\nu}\right) = P\left(\chi^2_{j+1,\nu} - P\left(\chi^2_{j,\nu}\right)\right) \qquad (j = -1; 0; +1) \qquad (13)$$

The values $\Delta P\left(\chi^2_{j,\nu}\right)$ of the first differences are given in Table [28] next to the corresponding values $P\left(\chi^2_{j,\nu}\right)$ of the function, such differences will always be negative. But it is necessary to find not the probability, but a more accurate value of χ^2. Write from table [28] less accurate value for the corresponding pi and ν, then the table of probability integral and formula (11) find the probability value $P\left(\chi^2_{j,\nu}\right)$ to seven significant digits and determine a more accurate value of χ^2 by formula

$$\chi^2_{i,\nu} = \chi^2_{0,\nu} + \left[\frac{p_i - p\left(\chi^2_{i,\nu}\right)}{\Delta P\left(\chi^2_{0,\nu}\right)} + U\right]\left(\chi^2_{+1,\nu} - \chi^2_{0,\nu}\right) \qquad (i = 1, 2) \qquad (14)$$

To control the finding of the quantities $\chi^2_{i,\nu}$, it is necessary to recalculate the probability $P\left(\chi^2_{j,\nu}\right)$, according to the more accurate value of $\chi^2_{i,\nu}$. If the adjusted values are calculated to the sixth significant figure, the discrepancy between the calculated and theoretical values of probability will not exceed 10 $^{-6}$. The calculated values of $\chi^2_{1,\nu}$ and $\chi^2_{2,\nu}$ with six significant figures at p = 0.95 for ν = from 1 to 20, which are given in table. 2. The discrepancy between the probabilities in the control calculations does not exceed 10^{-6}, this indicates that the proposed formula (14) allows with a high degree of accuracy to determine the value of χ^2, using the table of probability integrals [28].

For example, it is necessary to determine the refined value of $\chi^2_{1,3}$ for the confidence interval p = 0.95, for p_1 = 0.975, p_2 = 0.025 and ν = 3 from the table of probability integrals [28] we find that $\chi^2_{1,3}$ = 0.216, from the same tables write all the necessary data, which are given in table. 1.

Table 1.

Data to determine the refined value$\chi^2_{1.3}$

for trustworthiness p = 0.95

χ^2	$\nu = 3$	
	P	$-\Delta*10^{-5}$
0.15	0.98523	764
0.20	0.97759	845
0.25	0.96914	911

Then $P(\chi_{0,3}{}^2)$ = 0.99759; χ^2_3 = 0.216; $\chi^2_{0.3}$ = 0.2; $\chi^2_{+1.3}$ = 0.25, according to these values we calculate the value of U by the formula (12)

$$U = \frac{0.216-0.2}{0.25-0.2}.$$

Table. 2.

Value χ^2 from the tables [28] and clarifications according to the formula (14)

ν	Value χ^2			
	Hald and Synkbaek		Clarification by formula (14)	
	$\chi^2_{1.\nu}$ p_1=0.975	$\chi^2_{2.\nu}$ p_2=0.025	$\chi^2_{1.\nu}$ p_1=0.975	$\chi^2_{2.\nu}$ p_2=0.025
1	0.000982	5.024	0.000982092	5.02410
2	0.0506	7.378	0.0506364	7.37746
3	0.216	9.348	0.215798	9.34852
4	0.484	11.143	0.484404	11.1430

Continuation of table 2

v	Value χ^2			
	Hald and Synkbaek		Clarification by formula (14)	
	$\chi^2_{1.v}$ p_1=0.975	$\chi^2_{2.v}$ p_2=0.025	$\chi^2_{1.v}$ p_1=0.975	$\chi^2_{2.v}$ p_2=0.025
5	0.831	12.832	0.831140	12.8323
6	1.237	14.449	1.23742	14.4490
7	1.690	16.013	1.68989	16.0132
8	2.180	17.535	2.17976	17.5342
9	2.700	19.023	2.70030	19.0224
10	3.247	20.483	3.24709	20.4828
11	3.816	21.920	3.81590	21.9211
12	4.404	23.336	4.40379	23.3369
13	5.009	24.736	5.00863	24.7353
14	5.629	26.119	5.62874	26.1192
15	6.262	27.488	6.26221	27.4887
16	6.908	28.845	6.90775	28.8460
17	7.564	30.191	7.56418	30.1919
18	8.231	31.526	8.23087	31.5266
19	8.907	32.852	8.90667	32.8517
20	9.591	34.170	9.59082	34.1705

We write out the differences $\Delta P\left(\chi^2_{0,3}\right) =$ -845*10^{-5}; $\Delta P\left(\chi^2_{+1,3}\right) =$ 11*10^{-5}; $\Delta P\left(\chi^2_{0,3}\right) =$ -0.764*10^{-5}, (Table 1), we find the product U(1-U) = 0.2176, then the probability according to Bessel's formula (11) will be $P(\chi_{1,3}{}^2)$ = 0.9749659, by formula (14) calculate the refined value$\chi^2_{1.3}$ = 0.215798.

For control, we calculate again according to these formulas U = 0.31596;

U (1- U) = 0.2161293; $P(\chi_{1,3}{}^2)$ = 0.9749996, the discrepancy between the probabilities is 0.9749996- 0.975 = -0.4*10^{-6}, which is insignificant.

The obtained values χ^2 even to the third digit after the comma are more accurate than in the probability integral table, the values obtained with a rounding error not exceeding $5*10^{-6}$, the interpolation error does not exceed $10*10^{-6}$.

Для знаходження значень χ^2, for any probability p, which is not in the table, you must first calculate the value χ^2 між сусідніми значеннями ймовірності p_1, p_2 and $\chi^2_{1.v}$ and $\chi^2_{2.v}$, which are given in the table by the formula of logarithmic interpolation

$$\chi^2_{i,v} = \chi^2_{M.v} + \left(\chi^2_{max.v} - \chi^2_{min.v}\right)\frac{\ln p_i - \ln p_{max}}{\ln p_{min} - \ln p_{max}}, \quad (i = 1, 2), \tag{15}$$

where $p_{\min}$ and $p_{\max}$ – less and more adjacent probability values; $\chi^2_{min.v}$ та $\chi^2_{max.v}$ – less and more adjacent values χ^2, which correspond $p_{\min}$ and $p_{\max}$.

Perform the calculation similarly to the previous case, ie the approximate value χ^2 знаходимо точні значення ймовірності $P(\chi^2)$ and χ^2 according to formulas (11) and (14) and perform control calculations.

For example, you need to calculate the refined value $\chi^2_{1.3}$ for confidence p = 0.995, for p_1 = 0.9975, p_2 = 0.0025 і v = 3 from the table of probability integrals [28] and by formula (15) we find $\chi^2_{1.3}$ =0.04205, auxiliary value according to the formula (12) U = 0.205, U*(1 - U) = 0.162975, then the probability found by the formula (11) $P(\chi_{1,3}^2)$ = 0.9977372, by formula (14) calculate the refined value $\chi^2_{1.3}$ = 0.0449066. To control, once again calculate the probability by formula (11), we obtain $P(\chi_{1,3}^2)$ = 0.997504, the difference between the probabilities is 0.997504 – 0.9975 =$4*10^{-6}$.

As a result, using formulas (8) and table 2, you can easily calculate the acceptance and rejection numbers a_v and r_v with the required accuracy.

1.3.3 Sequential analysis of control measurements in topographic and geodetic production

Let's consider some typical examples of quality control of topographic and geodetic works that are performed.

Calculate by formulas (8) the values a_v and r_v of the acceptance and rejection numbers at different α and β for the confidence interval p = 0.95 and the unit standard σ = 1, which are shown in table 3.

Table 3.

Acceptance a_v and rejection r_v numbers at different values α and β

v	P = 0.95; α=0.05; β=0.05		P = 0.95; α=0.10; β=0.05		P = 0.95; α=0.10; β=0.10	
	a_v	r_v	a_v	r_v	a_v	r_v
1	0.52780	2.87251	0.54933	2.59653	0.82531	2.57500
2	1.11212	4.32709	1.14164	3.94867	1.52005	3.91915
3	1.77943	5.64827	1.81495	5.19289	2.27033	5.15737
4	2.49702	6.91703	2.53760	6.39677	3.05786	6.35619
5	3.24790	8.15483	3.29296	7.57726	3.87052	7.53221
6	4.02224	9.37107	4.07135	8.74149	4.70093	8.69238
7	4.81498	10.57093	4.86783	9.89344	5.54533	9.84059
8	5.62214	11.75860	5.67848	11.03632	6.40077	10.97997
9	6.44112	12.93537	6.50075	12.17096	7.26515	12.11134
10	7.26942	14.10276	7.33216	13.29845	8.13647	13.23571
11	8.10614	15.26183	8.17184	14.41957	9.01409	14.35387
12	8.95065	16.41551	9.01919	15.53686	9.89783	15.46833
13	9.80141	17.56302	9.87267	16.64944	10.78624	16.57818
14	10.65732	18.70441	10.73121	17.75724	11.67838	17.68335
15	11.51844	19.84136	11.59485	18.86171	12.57450	18.78530
16	12.38409	20.97385	12.46296	19.96279	13.47401	19.83393
17	13.25402	22.10254	13.33526	21.06103	14.37677	20.97979
18	14.12762	23.22799	14.21118	22.15684	15.28233	22.07328
19	15.00474	24.35020	15.09054	23.25020	16.19054	23.16440
20	15.88484	25.46817	15.97283	24.34017	17.10083	24.25218

This table is convenient to use when calculating the acceptance and rejection numbers for different standards, it is necessary to find the product of tabular values per square of the standard other than 1. In addition, table 3 will be used to calculate transient probabilities, finding the maximum number of control measurements random process.

Example. Controlled measurements are performed on the built-up area and compared with the distances determined on the topographic plan, according to [27] the discrepancy should not exceed 1 mm between the points of objects or contours of the situation and the nearest points of the shooting base. When shooting in Scale 1: 500 - 1mm on the plan corresponds to 0.5 m in the field.

To perform control measurements and their sequential analysis, it is necessary to make a table of acceptance and rejection numbers in advance, which can be calculated using the numbers in table. 3 by producting them by the square of the standard.

The value of the standard will be found by the limit value Δ = 1 mm on the plan or, Δ = 50 cm on the ground, the standard of measurement for the confidence interval $p = 0.95$ will be

$$\sigma(x) = \frac{\Delta}{2} = \frac{50}{2} = 25 \text{ см}. \quad (16)$$

We set the confidence probability $p = 0.95$ and the values $\alpha = \beta = 0.05$. Calculate the acceptance and rejection numbers, using the first two columns of numbers in table. 3, which are multiplied by $\sigma_2(x) = 625$, which are given in table. 4.

Table 4.

Acceptance and rejection numbers for the standard calculated by (16)

v	1	2		3	4	5	6	7	8	9	10	11	12
a_v	330	595		1112	1561	2030	2514	3009	3514	4026	4543	5066	5594
r_v	1795	2704		3530	4323	5097	5857	5607	7349	8085	8814	9539	10260

In this example, the control measurements are considered true, so the number of measurements n will be equal to the number of degrees of freedom v.

Perform the first control measurement to the contours of the situation. The difference between the distance measured by the tape measure and the distance obtained on the plan in accordance with the scale turned out 40 cm, raise it to the square and compare with the acceptance and rejection numbers for $\nu = 1$, table. 4, which shows that the square of the difference $\delta_1^2 = 1600$ is in the interval between the receiving and rejection numbers, ie $a_1 = 330 < \delta_1^2 = 1600 < r_1 = 1795$, from this we conclude that the control measurements should be continued. After performing the second control measurement, the difference between it and the measured on the plan was 25 cm, the square of which $\delta_2^2 = 625$.

Then we add the squares of the deviations (or the differences of the first and second) of the control measurements, we get the sum $[\delta^2]_2 = 1600 + 625 = 2225$. In the future the sum of the squares of the deviations will increase, so it will be accumulated. The results of the accumulated differences, their squares and accumulated amounts are shown in table 5.

Table 5.

The results of the obtained differences, their squares and accumulated sums

	1	2	3	4	5	6	7	8	9
	40	25	15	10	15	30	15	0	5
	1600	625	225	100	225	900	225	0	25
	1600	2225	2500	2600	2825	3725	3950	3950	3975

After the second measurement we compare the sum of squares of deviations with the acceptance and rejection numbers for $\nu = 2$, table. 4, we obtain the inequality $a_2 = 695 < [\delta^2]_2 = 2225 < r_2 = 2704$, and again we conclude that the control measurements must be continued.

In the third control measurement we get the difference of 15 cm and the accumulated sum $[\delta^2]_3 = \delta_3^2 + \delta_2^2 = 225 + 2225 = 2500$, and also compare it with the acceptance and rejection numbers for $\nu = 3$, and so on we continue these definitions, therefore, after the ninth measurement, we find that the accumulated sum of squares of deviations is less than the acceptance number for ν = 9, ie $[\delta^2]_9 = \delta_9^2\ [\delta^2]_8 = 25 + 3950$

= 3975 < a_9 = 4026 table. 4, as a result of which we stop control measurements and consider the performed topographic survey to be of good quality.

The instruction [27] states that the number of marginal discrepancies should not exceed 10% of the total number of control measurements.

In this example, nine control measurements were performed, but in none of them the difference reached the limit value of 50 cm, if we had to perform the tenth measurement, where the difference would be equal to the limit and assume that the eighth dimension difference is also 50 cm, then the work had to be rejected, ie $[\delta^2]_{10} = \delta_8^2 + \delta_9^2 + \delta_{10}^2 + [\delta^2]_7 = 3950 + 2500 + 25 + 2500 = 8975 > r_{10} = 8814$, in this example, two marginal differences are allowed, and we conclude that that these works should be rejected. Thus, at first glance, it seems that everything agrees with the requirements of the instructions, ie one limit is allowed for ten measurements, then the work given in the example will not be rejected, and control must be continued. If you look closely at the differences in this example, it is noticeable that they are mostly far from the limit values, but in practice this is rare. Usually, when performing surveys, the differences tend to go to the limit values, so the standard of calculations by formula (16) will not be satisfied, its value is slightly underestimated, ie overstated.

The following method of determining the standard by the limit value is proposed. According to the instruction [27], the number of marginal differences should not exceed 10%, ie one difference per ten control measurements is allowed, therefore, it is necessary to choose such a confidence interval p and its coefficient t_p so that the product of this coefficient is $\sigma(x)$ value less than the limit

$$t_p * \sigma(x) = \Delta - \omega, \tag{17}$$

where ω is a small range of measurement errors or differences, the number of which does not exceed 10% of the total. Expressions (17) will satisfy the following values of $t_p = 1.5$, $\sigma(x) = 30$ cm, $\omega = 5$ cm (accuracy of determining the distance on the scale in Sccale 1: 500). The value of $t_p = 1.5$ of the normal distribution corresponds to *a* confidence interval $p = 0.8664$ (or significance level $q = 0.1336$), since $t_p * \sigma(x) = 1.5 * 30 = 45$ cm, then at a given level of error or difference exceeding 45 cm, as well as errors of 50 cm and more will be 13.36%, the marginal errors themselves will not exceed 10%, which meets the requirements of the instructions [27].

The example shows that the control of distances measured from the plan and measured in kind deserves attention, so for quality control we take the standard $\sigma(x)$ = 30 cm

For example, to control the survey of the terrain with a section height of 0.5 m and a slope of up to 3°, it is necessary to determine the standard of measurements from the calculations of the acceptance and rejection numbers. The permissible differences between the control pickets and the marks found on the plan by interpolation in accordance with [27] should not exceed 25 cm, as in the previous example, the number of differences reaching the limit values should be less than 10% of the total number of control pickets.

Based on this condition, as in the previous example and in accordance with (17) we find $t_p * \sigma(x) = 25 - 1$, $\omega = 1$ cm, as the marks in Scale 1: 500 are determined to within one centimeter, for $t_p = 1.5$, standard $\sigma(x) = 24 / 1.5 = 16$ cm.

The number of errors exceeding 24 cm will be 13.36% of the total number of control measurements, acceptance and rejection numbers are calculated in the same way as in the previous example. In cases where there are no additional restrictions, and the limit value of the differences is specified, we use the formula to determine the standard

$$\sigma(x) = \frac{\Delta}{t_p}, \tag{18}$$

where t_p - is the argument of the normal distribution function, which is selected from the tables of this distribution on the confidence interval p.

Consecutive analysis of control - geodetic measurements gives the advantage that all the received information on quality is used for decision-making, instead of only its part, as it is specified in the instruction - judgments about quality are made only on the maximum differences. Thus, this analysis regulates the instructional tolerances and it is possible to more flexibly approach the quality control of topographic survey, but for this you need to choose the right value of the standard σ (x), the confidence probability p, and set the probabilities α and β.

1.3.4 Determination of the maximum number of control measurements in geodesy by the method of Markov random process

Random value $[\delta^2]$ - the sum of squares of deviations of the results of control - geodetic measurements from their actual values, or from their arithmetic mean, which is compared with the receiving a_v and defective r_v numbers, is nothing but χ^2 for $\sigma(x)$ = 1, respectively in this case, the values (measurements) are subject to χ^2 - distribution.

Random variable

$$\chi^2 = \frac{(n-1)m^2(x)}{\sigma^2(x)} \tag{19}$$

also ordered χ^2 – distribution since

$$m^2(x) = \frac{[\delta^2]}{n-1}, \tag{20}$$

substitute (20) into (19), obtain

$$\chi^2 = \frac{[\delta^2]}{\sigma^2(x)}, \tag{21}$$

If $\sigma(x)$ = 1, then $[\delta^2] = \chi^2$, it is important to find the probability that a random variable $[\delta^2]$ falls within the intervals

$$0 \div a_v;\ a_v \div r_v;\ r_v \div +\infty, \tag{22}$$

such probabilities will be found assuming that a_v and r_v subordinate also χ^2 distribution at $\sigma(x)$ = 1.

The process of sequential analysis of control measurements can end with one of two decisions: either the measurements will be accepted as benign if $[\delta^2]_v < a_v$ or they will be rejected if $[\delta^2]_v > r_v$ and this process can be considered as a heterogeneous Markov random process with discrete states, for this it is necessary to correctly compose a model of the Markov process [29], which is the main link of the Markov process.

It is necessary to find the probabilities of random value $[\delta^2]_v$ in different intervals of the corresponding measurement, ie in the interval of measurements $0 \div a_v$, in the interval of continuation of measurements $a_v \div r_v$ and in the interval of rejection of measurements $r_v \div +\infty$, and these probabilities must be determined very accurately , as the probability of completion of the analysis process should be almost equal to one. For this purpose, more accurate values of χ^2 (six significant digits, Table 2) are determined, by which we find a_v and r_v, according to which at $\sigma(x) = 1$ we calculate the probabilities of random values $[\delta^2]$ ν in the receiving interval, the interval of continuation of control measurements or defective interval (22).

Consider examples of finding the probability of a random variable $[\delta^2]_v$ in intervals (22), set the probability $p = 0.95$ and $\alpha = \beta = 0.05$ and the number of degrees of freedom $v = 1$. Determine the probability of falling into the interval of measurements (Table 3) $0 \div a_v$; $0 \div 0.52780$, such a probability will be found in the following way. The probability that the random variable $[\delta^2]_v = \chi^2$ will exceed zero is equal to one, ie $P([\delta^2]_v > 0) = 1$, and the probability that $[\delta^2]_v$ will exceed $a_1 = 0.52780$ is calculated by Bessel's formula (11) :

$$P([\delta^2] > a_1) = 10^{-5}[47950 + 0.556(-2118) - 0.061716(-1974 + 2284)] = 0.46753,$$

according to the formula (12) $U = 0.556$, $U(1\text{-}U)/4 = 0.061716$.

For values $[\delta^2]_v < 1$ perform control over formula (10), then

$$\phi(\sqrt{a_1}) = 1 - 0{,}5 * 0.46753 = 0.7662335;\ \sqrt{a_1} = 0.726498,$$

$a_1 = 0.52780$, $\chi_1^2 = 0$, $p_1 = 1.00000$, $\chi_2^2 = 0.52780$, $p_2 = 0.46753$, then the probability of falling into the interval $0 \div a_v; 0 \div 0.52780$ will be $P_{12} = 1\text{-}0.46753 = 0.53247$.

The probability of a random variable $[\delta^2]_v$ in the interval of continuation of measurements of table. 3, $a_1 \div r_1; 0.52780 \div 2.87251$, is calculated as the difference of the probabilities of the ends of the interval, which are determined by the Bessel formula, but for the left end of the interval the probability is already obtained, it is necessary to calculate the probability for the right end of the intervaly

$$P([\delta^2] > r_1) = 10^{-5}[9426 - 0.36255 * 1100 - 0.23107(-962 + 1260)/4] = 0.09010,$$

U = 0.36255, U*(1-U) = 0.231107, received the following data: χ_1^2 =0.52780, p_1 = 0.46753, χ_2^2 = 2.87251, p_2 = 0.09010, accordingly, the probability of falling into the interval $0 \div a_\nu; 0 \div 0.52780$ will be equal to P_{11} = 0.46753 – 0.09010 = 0.37743.

The probability of hitting a random variable $[\delta^2]_\nu$ in the interval of missing measurements $r_\nu \div +\infty$; $2.87251 \div +\infty$ will be defined as the probability that a random variable $[\delta^2]_\nu$ exceed r_1, that is $P_{13} = P([\delta^2]_\nu > r_1)$ =0.09010. The sum of all three probabilities must be equal to one, ie $P_{11} + P_{12} + P_{13} = 1$, or 0.37743 + 0.53247 +0.09010 = 1.00000.

In fig. 1 shows in the general case of probability $P([\delta^2]_\nu > 0) = 1$, $P([\delta^2]_\nu > a_\nu)$ and $P([\delta^2]_\nu > r_\nu)$, with the appropriate number of degrees of freedom ν.

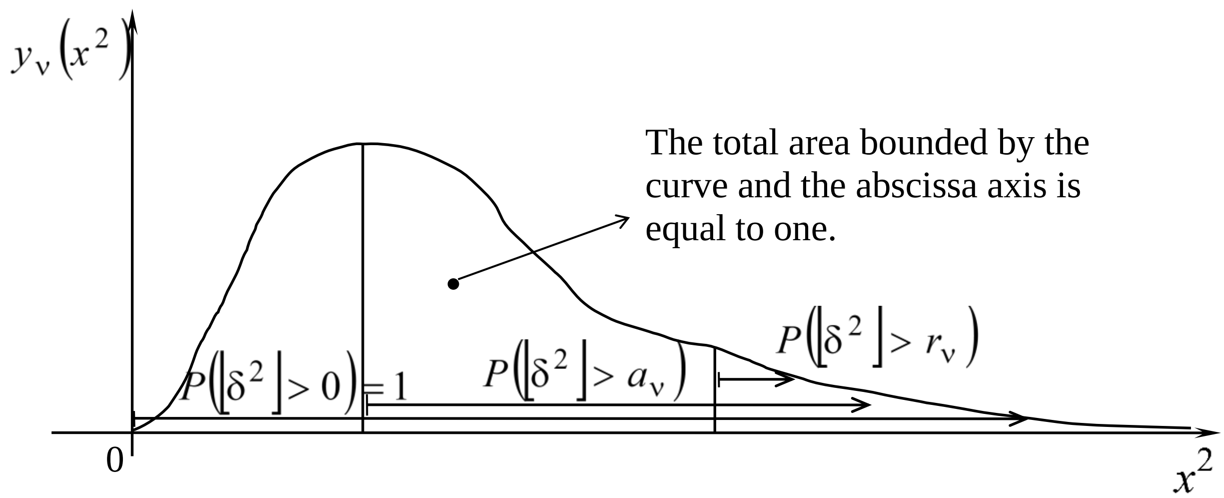

Fig. 1. Probability density χ^2 – distribution with specified missing and acceptable intervals

These probabilities are equal to the areas under the probability density curve χ^2 – distribution to the right of the corresponding ordinates, as shown by the arrows going to infinity, the sum of all three areas bounded by the abscissa intervals $0 \div a_\nu$; $a_\nu \div r_\nu$; $r_\nu \div +\infty$, equal to the total area bounded at the top by a common curve, below the abscissa χ^2, ie equal to one.

The probabilities of the random variable $[\delta^2]_\nu$ in the specified intervals, with the corresponding number of degrees of freedom ν, are determined by the proposed equations:

$$P_{12}^{(\nu)} = 1 - P([\delta^2]_\nu > a_\nu),$$

$$P_{11}^{(\nu)} = P([\delta^2]_\nu > a_\nu) - P([\delta^2]_\nu > r_\nu)$$

$$P_{13}^{(\nu)} = P([\delta^2]_\nu > r_\nu). \tag{23}$$

The probability of confidence is $p = 0.95$, the probability $\alpha = \beta = 0.05$; $\alpha = 0.10$ $\beta = 0.05$ and $\alpha = \beta = 0.10$, calculated probabilities of random values $[\delta^2]_\nu$ in these three considered intervals, for different degrees of freedom $\nu = 14$, $\nu = 12$ and $\nu = 10$, such probabilities are calculated as in previously considered examples, by formula (11) and by equations (23) are given in table. 6.

Table 6.

Probabilities of finding a random value in intervals

ν	$P_{11}^{(\nu)}$	$P_{12}^{(\nu)}$	$P_{13}^{(\nu)}$
$p = 0.95$, $\alpha = 0.05$, $\beta = 0.05$			
1	0.37743	0.53247	0.09010
2	0.45854	0.42654	0.11492
3	0.48938	0.38058	0.13004
4	0.50483	0.35483	0.14034
5	0.51398	0.33818	0.14791
6	0.51991	0.32633	0.15376
7	0.54206	0.31748	0.15846
8	0.52718	0.31052	0.16230
9	0.52952	0.30492	0.16556
10	0.53143	0.30021	0.16836
11	0.53294	0.29624	0.17082
12	0.53420	0.29286	0.17294
13	0.53526	0.28991	0.17483
14	0.53616	0.28728	0.17656
15	0.53695	0.28495	0.17810

Continuation Table 6

v	$P_{11}^{(v)}$	$P_{12}^{(v)}$	$P_{13}^{(v)}$
p = 0.95, α = 0.10, β = 0.05			
1	0.35150	0.54141	0.10709
2	0.42620	0.43494	0.13886
3	0.45348	0.38832	0.15820
4	0.46651	0.36208	0.17141
5	0.47380	0.34508	0.18112
6	0.47837	0.33298	0.18865
7	0.48139	0.32392	0.19469
8	0.48353	0.31682	0.19965
9	0.48509	0.31106	0.20385
10	0.48630	0.30624	0.20746
11	0.48716	0.30217	0.21067
12	0.48791	0.29871	0.21338
p = 0.95, α = 0.10, β = 0.10			
1	0.25507	0.63637	0.10856
2	0.32674	0.53234	0.14092
3	0.35760	0.48177	0.16063
4	0.37411	0.45181	0.17408
5	0.38424	0.43179	0.18397
6	0.39107	0.41730	0.19163
7	0.39594	0.40628	0.19778
8	0.39959	0.39757	0.20284
9	0.40243	0.39046	0.20711
10	0.40472	0.38449	0.21079

The probabilities P12 and P13 are transient, and P11 is the delay probability, for each number of degrees of freedom such probabilities are calculated (Fig. 2).

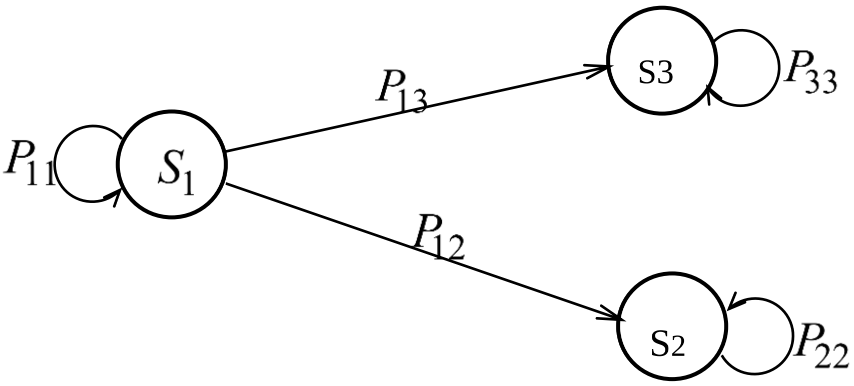

Fig. 2. State graph of a random Markov process for determining the maximum number of control measurements

To determine the maximum number of measurements at which the analysis process will end, make a graph of states, which is shown in Fig. 2, circular arrows indicate the probabilities of delay in the respective states, and the directions of transition of the analysis process from state to state

From fig. 2 shows that the measurement analysis process is in one of the states:

S_1 - measurement must be continued;

S_2 - measurements are accepted as benign;

S_3 - measurements are missing.

There will be three delays for this state graph: P_{11}, P_{22}, P_{33}, for this state graph these probabilities are equal to

$$\begin{gathered} P_{11} = 1 - (P_{12} + P_{13}); \\ P_{22} = 1 \\ P_{33} = 1 \end{gathered} \tag{24}$$

For each number of degrees of freedom, a square matrix of transient transition probabilities and delay probabilities is formed, and the delays probabilities are on the main diagonal of the matrix, ie

$$\left|P_{ij}^{(\nu)}\right| = \begin{vmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{vmatrix} \tag{25}$$

index (ν) indicates the number of degrees of freedom, when the number of degrees of freedom ν = 1 matrix of transient probabilities (Table 6) will take the form

$$\left[P_{ij}^{(1)}\right] = \begin{bmatrix} 0.37743 & 0.53247 & 0.09010 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

At the initial moment, when one control measurement is performed, the analysis process is in state S_1, and control measurements must be continued. The probabilities of states or the probability that the analysis process will be in the appropriate states before the measurements at $v = 0$ will be equal to $P_1(0) = 1$, $P_2(0) = 0$, $P_3(0) = 0$. The probabilities of states at stage $v = 1$ and the following are the product of the matrix - the row of the previous stage to the square matrix of transition probabilities of the current stage, ie

$$|P_1(\nu), P_2(\nu), P_3(\nu),| = |P_1(\nu-1), P_2(\nu-1), P_3(\nu-1),| * \begin{vmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{vmatrix} \tag{26}$$

For the number of degrees of freedom $v = 1$ the probabilities of the states will be equal $P_1(1)$=1*0.37743 + 0*0+ 0*0 = 0.37743, $P_2(1)$ = 1*0.53247 + 0*1+ 0*0 = 0.53247, $P_3(1)$ = 1*0.09010 + 0*0 + 0*1 = 0.09010. $P_1(1) + P_2(1) + P_3(1)$ = 1.00000.

With the number of degrees of freedom $v = 2$ and the matrix of transient probabilities (Table 6) will take the form

$$\left[P_{ij}^{(2)}\right] = \begin{bmatrix} 0.45854 & 0.42654 & 0.11492 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

By formula (26) we find the probabilities of all three states and their sum:

$P_1(2)$ = 0.37743*0.45854 = 0.17307, $P_2(2)$ = 0.37743*0.42654 + 0.53247 = 0.69346, $P_3(2)$ =0.37743*0.11492 + 0.09010 = 0.13347, $P_1(2) + P_2(2) + P_3(2)$ = 1.00000.

Subsequent calculations for the next number of degrees of freedom were performed similarly. For $v = 15$ and $v = 16$ the probabilities of the states will be equal accordingly:

$P_1(15)$=0.00004, $P_2(15)$=0.81849 $P_3(15)$=0.18147, $P_1(15)$+ $P_2(15)$+ $P_3(15)$=1.00000

$P_1(16)=0.00002$, $P_2(16)=0.81850$ $P_3(16)=0.18148$, $P_1(16)+P_2(16)+P_3(16)=1.00000$

Thus, the Markov random process established that the process of sequential analysis of control measurements will end for the number of degrees of freedom $\nu = 17$ with probabilities of states equal to $P_1(17)=0.00001$, $P_2(17)=0.81851$ $P_3(17)=0.18148$, $P_1(17)+P_2(17)+P_3(17)=1.00000$, with the accuracy of calculating the probabilities of states 10^{-5} by making one of two decisions: measurements are taken as benign with a probability of 0.81851 or measurements are missing with a probability of 0.18148, the probability of continuing control measurements after $\nu = 17$ will be small and 0.00001 and will continue to decline, and analysis will end for data $p = 0.95$, $\alpha = \beta = 0.05$. In general, sequential analysis can end with a much smaller degree of freedom. It should be noted that the probability of rejecting work 0.18148 is quite strict, which means that such data should be taken to calculate the acceptance and rejection numbers to control the quality of very responsible work.

From the calculation of the maximum number of control measurements we can conclude that the choice of probabilities p, α, β is correct, also such calculation is important for planning control - geodetic measurements.

Calculated probabilities of states for $p = 0.95$, $\alpha = 0.10$, $\beta=0.05$, when the number $\nu = 12$, is equal to $P_1(12)=0.00009$, $P_2(12)=0.79738$ $P_3(12)=0.20254$, $P_1(12)+P_2(12)+P_3(12)=1.00000$. The maximum number of control measurements will be 15, with the accuracy of the calculation of states 10^{-5}, it is seen that the probability of rejection is increased compared to the first case.

But the probabilities of states for $\nu = 10$ are equal $P_1(10)=0.00004$, $P_2(10)=0.83346$, $P_3(10)=0.16650$, $P_1(12)+P_2(12)+P_3(12)=1.00000$, this means that one of the six (1: 0.16650 = 6) surveyed areas will be rejected, but this does not mean that the entire survey will be rejected, as the most vulnerable areas are selected for control.

In this case, the probability of rejection is reduced compared to the first and second case, and the maximum number $\nu = 12$.

From these three calculations of the maximum number of control measurements it can be concluded that to control the quality of less responsible types of work, the confidence probability p should be reduced and the probabilities α, β increased by small values (0.10), while the probability of rejection will decrease.

5.On the number of measurements of line lengths by light rangefinders in polygonometry

How many measurements of line lengths must be made depending on the class (or category) of polygonometry, the accuracy of the device and meteorological conditions of the environment. To do this, use the mathematical apparatus of probably statistical sequential analysis.

It is known that under favorable weather conditions (when it is cloudy and cool) the convergence of measurement results with light rangefinders is good, and under adverse weather conditions in summer, when there are cool sunny days, the convergence of measurement results is poor. We can conclude that in the first case you can limit yourself to a few dimensions, and the second 10 ÷ 15 dimensions, and sometimes reject them and move to another time (day).

Particularly unfavorable conditions occur when in clear weather part of the line we measure falls in the solar zone, and part in the shadow, or passes over the water surface, it is obvious that this disadvantage is caused by differences in temperature and humidity of these zones. Another type of adverse conditions is when a shaky object (tree branch) falls into the line of measurement. In such cases, the discrepancy of measurements will be significant, respectively, and their accuracy will be lower compared to the accuracy in favorable conditions with the same number of measurements in both cases, there are more and less favorable conditions of measurement, but in nature there are many others (intermediate) conditions that are almost impossible to take into account, and in each case the scope of measurement results will be different with an equal number of measurements.

To determine the number of measurements it is necessary to determine the acceptance and rejection numbers, for the calculation of which it is necessary to establish the values of the standards σ (x) of linear measurements in polygonometry.

Denote the denominator of the relative error of the stroke [] T, the denominator of the average relative error after 2 * T, the denominator of the average relative error of linear measurements is determined from the expression

$$T_l = T * \sqrt{2}. \quad (27)$$

These values are given in table. 7

Table 7.

Line lengths and denominators of relative errors in polygonometry

Indicators	Polygonometry		
	4 class	1 category	2 category
Line length: the largest, m the smallest, m	 3000 250	 800 120	 500 80
Relative error of course 1:T	1: 25000	1: 10000	1: 5000
The denominator of the average relative error 2*T	50000	20000	10000
The denominator of the average relative error of linear measurements T_l	70710.678	28284.271	14142.136
Standard of a separate measurement result, at n =4, mm	7.071	8.485	11.314

The standard root mean square deviation of the arithmetic mean of linear measurements is determined by the minimum side length of the expression on(1 category)

$$\sigma\left(x_{cp}\right) = \frac{l_{min}}{T_l}. \qquad (28)$$

Since it is not known in advance at what number the measurements will end, it is impossible to determine the exact value of the standard of a single measurement σ (x). However, sequential analysis must begin after at least three measurements, so for the calculations we take the number of measurements $n = 4$. The standard of a single measurement result will be

$$\sigma(x) = \sigma\left(x_{cp}\right) * \sqrt{n}\,, \qquad (29)$$

at n equal to five, six, and so on, its value will increase. Therefore, for the calculation standard, for four measurements, the accuracy of the linear measurements will be slightly inflated compared to if the measurements ended at five, six, and so on, and vice versa, the accuracy will be slightly lower for three measurements. Given (28) and the accepted average number of endings of measurements $n = 4$, the standard of a separate measurement result

$$\sigma(x) = \sqrt{n} * \frac{l_{min}}{T_l}\,. \qquad (30)$$

The values calculated by formula (30) are given in table. 7.

To determine the constant light rangefinder on a field comparator with a length of 120 *m*, set the relative error of 1: 120000 and the average number of measurements $n = 16$, then the standard of a single measurement will be

$$\sigma(x) = \sqrt{n} * \frac{l_{min}}{T_l} = 4 * \frac{120000}{120000}\,.$$

With a confidence interval $p = 0.95$, probability values $\alpha = \beta = 0.05$ and for the obtained values of standards, in accordance with formulas (8), calculated acceptance and rejection numbers, which are given in table. 8.

Table 8.

Acceptance and rejection numbers for measuring line lengths

v	Comparison of the light rangefinder		Polygonometry					
			4 class		1 category		2 category	
	a_v	r_v	a_v	r_v	a_v	r_v	a_v	r_v
1	8	46	26	144	38	207	68	368
2	18	69	56	216	80	312	142	554
3	28	90	89	282	128	407	228	723
4	40	111	125	346	180	498	320	885
5	52	130	162	408	234	578	416	1044
6	64	150	201	469	290	675	515	1199
7	77	169	241	529	347	761	616	1353
8	90	188	281	588	405	847	720	1505
9	103	207	322	647	464	931	824	1656
10	116	226	363	705	523	1015	930	1805
11	130	244	405	763	584	1099	1038	1954
12	143	263	448	821	644	1182	1146	2101
13	157	281	490	878	706	1264	1255	2248
14	171	299	533	935	767	1347	1364	2394
15	184	317	576	992	829	1429	1474	2540
16	198	336	619	1049	892	1510	1585	2685
17	212	354	663	1105	954	1591	1697	2829
18	226	372	706	1161	1017	1672	1808	2973
19	240	390	750	1218	1080	1753	1921	3117
20	254	407	794	1273	1144	1834	2033	3260

Sequential analysis of measurements according to the above method should be used when comparing light rangefinders. It is assumed that the actual length of the measured value is unknown, as the constant correction of the light meter is unknown, which must be determined, so the analysis should start with the second measurement and use the receiving and rejection numbers of the first measurement. so on.

For example, the comparison of the ST - 5 light rangefinder is performed, the first ten results of measurements in meters are obtained: 120.006, 120.004, 120.005, 120.012, 120.004, 120.011, 120.008, 120.003, 120.006, 120.000. The arithmetic mean $X(10)_{cp}$ = 120.0059m, the sum of the squares of the deviations $[\delta^2]_{10} = 119$, compare it with the acceptance and rejection numbers at $v = 9$ (Table 8), we obtain the inequality $a_9 = 103 < 119 < r_9 = 207$, from this we conclude that the measurement should be continued. Five more measurements were performed 120.002, 120.010, 120.006, 120.008, 120.004. According to fifteen measurement results, the average value of $X(15)_{cp}$ = 120.0059m, the sum of the squares of the deviations $[\delta^2]_{15} = 159 < a_{14} = 171$, so we finish the measurement and consider them of good quality. If it turns out that $[\delta^2]_{15} > r_{14} = 299$, then the measurement should be rejected and choose more favorable conditions for comparison.

When analyzing linear measurements, it is desirable that the display of the rangefinder shows not only the measurement result, but also the average value, the accumulated sum of squares of deviations, and the error of the arithmetic mean. Of course, it would be better if the calculator's memory could be stored and rejected numbers and that the analysis of measurements was automatically performed with a conclusion about their quality.

In linear measurements in polygonometry, sequential analysis is used in a simple form, but one that provides the required accuracy.

According to the calculated acceptance and rejection numbers Table (8), we find the value of the scale of measurements from the expressions:

$$\begin{Bmatrix} R_a = f_n * m_a \\ R_r = f_n * m_r \end{Bmatrix} \tag{31}$$

where f_n is the coefficient selected from the tables depending on the number of measurements [28]; m_a and m_r are the root mean square errors of the acceptance and rejection numbers, which are determined by the formulas:

$$\left.\begin{aligned} m_a &= \sqrt{\frac{a_v}{n-1.45}} \\ m_r &= \sqrt{\frac{r_v}{n-1.45}} \end{aligned}\right\}. \tag{32}$$

In these formulas, the number of degrees of freedom $v = n$ - 1.45. It is known that with the normal distribution of measurement errors and their small number, use the numbers of degrees of freedom not (n - 1), but (n -1.45), if the true value of the measured value is unknown [].

Based on the table. 8, the values of the coefficient fn, formulas (30) and (31), the calculated values of the scope are given in table. 9.

Table 9.

Coefficients of the range of measurement of line lengths

n	f_n	Polygonometry					
		4 class		1 category		2 category	
		R_a	R_r	R_a	R_r	R_a	R_r
3	1.69	10	20	12	24	16	32
4	2.06	12	22	15	26	19	35
5	2.33	14	23	17	28	22	37
6	2.53	15	24	18	29	24	38
7	2.70	16	25	20	30	26	40
8	2.85	17	26	21	31	28	41
9	2.97	18	26	22	31	29	42
10	3.08	19	27	23	32	30	43
11	3.17	20	27	23	32	31	44
12	3.26	20	28	24	33	32	44
13	3.34	21	28	25	33	33	45
14	3.41	21	28	26	34	34	46
15	3.48	22	29	26	34	35	46
16	3.53	22	29	27	35	36	47
17	3.59	23	29	27	35	37	47

Continuation of Table 9

n	f_n	Polygonometry					
		4 class		1 category		2 category	
		R_a	R_r	R_a	R_r	R_a	R_r
18	3.64	23	30	28	36	37	48
19	3.69	24	30	28	36	38	48
20	3.74	24	30	28	36	38	48

This table is used for linear measurements in polygonometry with a rangefinder ST - 5.

The swing efficiency at the number of measurements n = 2 is equal to 1.00, and for n = 5; 10; 15; 20 and 25, it will be 0.96, respectively; 0.85; 0.77; 0.70 and 0.65, which shows that the highest efficiency will be at n < 10, and then it decreases.

Measurements almost end at n < 10, occasionally when the number of measurements reaches 15.

Consecutive analysis of measurements on their scope is performed when the number of measurements is not less than two, but it is better to do it with three or more measurements.

For example, the results of measurements in polygonometry 1 - discharge in meters: 150.211, 150.208, 150.200. The scope is: $R = x_{\max} - x_{\min}$ = 150.211 -150.200 = 0.011m, or 11 mm. Using the table. 9, we find that R = 11mm <Ra_3 = 12mm, respectively, we can conclude that the measurements should be stopped and considered benign.

For example, there are three measurement results in meters: 171.337, 171.343, 171.351, span R = 14 mm, which is in the range $Ra_3 = 12 < R = 14 < Rr_3 = 24$, respectively, you need to continue the measurement, and perform another measurement - 171.336 m , respectively, for four measurements, the range R = 15 mm, which will be equal to the receiving range Ra_4 = 15 mm, which means that we stop the measurements and consider them benign.

For example, three measurements were performed in meters: 180.211, 180.217, 180.232 as $Ra_3 = 12 < R = 14 < Rr_3 = 24$, the measurement is continued, as a result of the fourth measurement 180,240 m are obtained. The span $R = 29 > Rr_4 = 26$ mm, respectively the obtained results must be rejected and the reason found out, it is possible that an object that is oscillating got into the line of the measured line.

For example, three measurements were made in meters: 201.451, 201.454, 201.478 since $R = 27 > Rr_3 = 24$ mm, such results must be rejected. Given that the first two results differ by only 3 mm, exclude the third measurement and continue the measurement, we obtain 201.467, 201.457, 201.459. The scope of the five measurements is $R = 16 < Ra_5 = 17$ mm, so we stop the measurements and consider them of good quality.

Thus, the sequential analysis is performed quite simply and quickly using the table. 9 and at the same time it is quite reasonable to stop the measurement, with a certain number of measurements, which eliminates the uncertainty regarding the number of measurements.

Using the table. 9 it is possible to perform the analysis of the measurements made by light rangefinders of various accuracy. In this case, for rangefinders with greater accuracy, the measurements will end with fewer, and for rangefinders with less accuracy with more measurements. In general, the number of measurements will be regulated in the process of their sequential analysis.

n accordance with the instruction [25], the root mean square error of measuring distances with a light rangefinder CT - 5 is determined by the formula

$$m_d = (10 + 5 * \ 10^{-6} * D)mm \tag{33}$$

і для електронного тахеометра за формулою

$$m_d = (5 + 3 * \ 10^{-6} * D)mm, \tag{34}$$

where D - is the measured distance in meters.

6.One of the methods of finding the correlation coefficient in geodetic measurements

A formula for determining the correlation coefficients is proposed, according to which with a small number of measurements the values of the coefficients will be larger compared to the values calculated by known formulas, ie these values will be less offset relative to the true values of correlation coefficients.

In geodesy in general are limited to a small number of measurements, in photogrammetry at mutual orientation of pictures of a stereo pair in an automatic mode at identification of points on the left and right pictures it is necessary to calculate precisely values of correlation coefficients.

Known [] with a small number of measurements in the case of dependent series of measurements, the correlation coefficient calculated by a known formula

$$r_0 = [\sum_{i=1}^{n} \delta(x_i)\delta(y_i)] /[(n-1)\sigma(x)\sigma(y)], \tag{35}$$

will be underestimated. In this formula, the deviations $\delta(x_i)$ and $\delta(y_i)$ are from the expressions

$$\delta(x_i) = x_i - x_c; \ \delta(y_i) = y_i - y_c \quad , \tag{36}$$

where x_i, y_i - measurements in the series X and Y; x_c, y_c - arithmetic mean values for these series of measurements.

Empirical standards or root mean square errors of the results of measurements on the series X and Y are determined by known formulas:

$$\sigma(x) = \sqrt{\sum_{i=1}^{n} \delta(x_i)/(n-1)};$$

$$\sigma(y) = \sqrt{\sum_{i=1}^{n} \delta(y_i)/(n-1)}, \tag{37}$$

where n is the number of measurements in a row.

There is a formula by which you can find more accurate values of the correlation coefficient:

$$r' = r_0 \ \{1 + (1 - r_0^2)/[2 * (n-3)]\}, \tag{38}$$

but this formula, as it turns out when the number of measurements n> 5 gives a slightly underestimated value of the correlation coefficient.

First, let's focus on the equations of direct regression for two dependent series of measurements X and Y, such as:

$$y = b_1 + k_1 * x; \quad x = b_2 + k_2 * y. \tag{39}$$

In these equations, the values of b_1 and b_1 are represented by expressions

$$b_1 = y_c - k_1 * x_c,$$

$$b_2 = x_c - k_2 * y_c. \tag{40}$$

Regression coefficients or angular coefficients k_1 and k_2 can be found by formulas

$$k_1 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sum_{i=1}^{n} \delta^2(x_i)\,;$$

$$k_2 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sum_{i=1}^{n} \delta^2(y_i)\,. \tag{41}$$

Formula (35) for determining the correlation coefficient can be represented taking into account (41) as follows:

$$|r_0| = \sqrt{k_1 * k_2}. \tag{42}$$

If the regression equations (39) coincide, ie the angle between these lines $\varphi = 0$, the correlation coefficient will be equal to 1, $r_o = 1$. With increasing angle, the correlation relationship decreases and when $\varphi = \pi / 2$, this relationship is zero , ie $r_o = 0$. Thus, the correlation coefficient will depend on the angle between the direct regression equations (39), such a dependence is proposed to be represented as

$$r = 1 - 2 * \varphi/\pi, \tag{43}$$

where φ is the numerical value of the angle in radians (degres).

To find the angle φ we use the formula of the angle between two lines, for this we find the second equation (39) with respect to y, then this system will take the form

$$y = b_1 + k_1 * x$$

$$y = -b_2/k_2 + (1/k_2) * x. \tag{44}$$

Then the angle between the two lines will be determined as

$$\tan\varphi = (1/k_2 - k_1)/(1 + k_1/k_2) = (1 - k_1 * k_2)/(k_1 + k_2) \tag{45}$$

Given (45) and (43), the formula of the correlation coefficient will look like

$$r = 1 - 2/\pi * \tan^{-1}[(1 - k_1 * k_2)/(k_1 + k_2)]. \tag{46}$$

The resulting formula is used to calculate the correlation coefficient with a positive correlation. In expression (41) the coefficients k_1 and k_2 can only be with the same signs, since in their expressions the same numerator and denominators are always positive values. The product of the coefficients k_1 and k_2 in accordance with (42), and the fact that $| r_o | \leq 1$ will

$$0 \leq k_1 * k_2 \leq 1. \quad (47)$$

Therefore, the numerator in (45) will be a positive value, ie

$$1 - k_1 * k_2 \geq 0. \quad (48)$$

Accordingly, the sign of the fraction in (45) will depend on the sign of the denominator, which in turn depends on the sign of the coefficients k_1 and k_2.

In the case of a negative correlation, when an increase in one value corresponds to a decrease in another, the signs of the coefficients k_1 and k_2 will be negative, then the angle φ will be negative, and the formula of the correlation coefficient will take the form

$$r = -1 - 2/\pi * \tan^{-1}[(1 - k_1 * k_2)/(k_1 + k_2)]. \quad (49)$$

Consider examples of calculating correlation coefficients, first with a positive relationship. At the point of triangulation, the directions were measured by twelve techniques according to the method of circular techniques, deviations from the average values of the directions are presented in table. 10.

Table 10.

Deviation of the measured directions from the average values

Receptions	$\delta(x_i)$	$\delta(y_i)$	Receptions	$\delta(x_i)$	$\delta(y_i)$
1	-2.18	+0.57	7	-0.48	-1.13
2	-2.38	-2.03	8	-0.98	-1.03
3	+0.62	+1.17	9	+2.32	+0.77
4	-0.38	+0.07	10	+0.42	-0.23
5	+0.82	+2.07	11	+2.32	+1.97
6	+0.02	-1.63	12	-0.08	-0.53

According to table1 $\sum \delta(x_i) * \delta(y_i) = 13.807$; $\sum \delta^2(x_i) = 23.757$;

$\sum \delta^2(y_i) = 19.907$; $k_1 = 0.5812$; $k_2 = 0.6936$.

The correlation coefficient is calculated by the formula (45) or (42) $r_0 = 0.635$, according to the formula (38) $r' = 0.656$, according to the formula (46) r = 0.721.

This example shows that the correlation coefficient calculated by formula (38) is greater by 0.021, and by formula (35) is greater by 0.086 compared to its value calculated by formulas (35) or (42).

Consider a negative correlation. For two series of measurements at n = 5, the following values were obtained $\sum \delta(x_i) * \delta(y_i) = -0.41$; $\sum \delta^2(x_i) = 0.72$;

$\sum \delta^2(y_i) = 0.84$; $k_1 = -0.5694$; $k_2 = -0.4881$, the correlation coefficient in this case is calculated by the formula

$$r_0 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sqrt{\sum_{i=1}^{n} \delta^2(x_i) * \sum_{i=1}^{n} \delta^2(y_i)} = -0.527. \qquad (50)$$

According to formula (38) the coefficient r'= - 0.622, and according to formula (49) r = -0.619, then for the number of measurements n = 5 the correlation coefficients, which are calculated by formulas (38) and (46) are almost the same, but necessary Note that in the study of correlation, the number of measurements is always more than five, so it is proposed to calculate the correlation coefficient by formula (46), the value of which will be plausible in comparison with the values found by formulas (35), (42), (50) and by formula (38) when the number of measurements is more than five.

Let us investigate the proposed new formula for calculating the correlation coefficient. First, we show that the number of measurements n in the series in which the correlation coefficients, which are calculated by formulas (38) and (46), is approximately the same and varies around five. To do this, formula (46) is presented in another form. From formula (42) it is seen that the correlation coefficient is equal to the geometric mean of the coefficients k_1 and k_2. If (k_1 + k_2): 2 is the arithmetic mean, $k_1 = k_2 = r_0$, which follows from (42), then

$$k_1 + k_2 = 2 * r_0. \qquad (51)$$

Given (51), expression (46) will take the form

$$r = 1 - 2/\pi * \tan^{-1}[(1 - r_0^2)/(2 * r_0)] . \qquad (52)$$

For the example given earlier with a positive correlation r_0 = 0.635, we substitute this value in (4), we obtain r = 0.720. According to the formula (46) r = 0.721, it is seen that these obtained values are practically equal, and the difference between the coefficients k_1 - k_2 = 0.69 - 0.58 = 0.11 is quite significant.

To find the number of measurements n at which the correlation coefficients calculated by formulas (38) and (52) will be equal, for this we equate the right-hand sides of equations (38) and (52), we obtain

$$r_0 * \left[1 + \frac{1-r_0^2}{2*(n-3)}\right] = 1 - \frac{2}{\pi} * \tan^{-1}\frac{1-r_0^2}{2*r_0} \tag{53}$$

From expression (53) we will find

$$n = \frac{r_0\left(1-r_0^2\right)}{2*\left(1-r_0-1-\frac{2}{\pi}*\tan^{-1}\frac{1-r_0^2}{2*r_0}\right)} + 3 \tag{54}$$

Given different values of the correlation coefficient r_0, we determine the value of n by expression (54), they are given in table. 11.

Table. 11

Number of measurements

r_0	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
n	4.8	4.8	4.9	4.9	5.0	5.1	5.2	5.3	5.4	5.6	5.7

From the table. Table 11 shows that the equality of correlation coefficients found by formulas (38) and (46) or (52) corresponds to the number of measurements of about five.

To compare the values of the coefficients obtained by different formulas, it is first necessary to obtain formulas by which it is possible to find the coefficients k_1 and k_2 depending on the given values of ρ - the theoretical value of the correlation coefficient.

For equilibrium measurements in two rows x and y, the maximum allowable ratio of empirical variances is determined by the inequality [30]

$$\frac{\sigma^2(x)}{\sigma^2(y)} \le Fq,\ \left(\sigma(x) \ge \sigma(y)\right) \tag{55}$$

Where F_q is the variance ratio, which is selected from the table F - distribution depending on the level of significance q and the number of degrees of freedom n - 1. Given formulas (37) and (41) and (55), the variance ratio will be presented as

$$\frac{\sigma^2(x)}{\sigma^2(y)} = \frac{\sum_{i=1}^{n}\delta^2(x_i)}{\sum_{i=1}^{n}\delta^2(y_i)} = \frac{k_2}{k_1} = Fq \tag{56}$$

Consider the case when the ratio of the coefficients k2 and k1 will be maximum, then the values of the correlation coefficients, which are determined by formulas (18) and (46) will have the largest difference. We compose a system of equations from (22) and (42), from the solution of which we find the coefficients k1 and k2

$$k_2/k_1 = Fq,\ k_1 * k_2 = r_0^2\ , \quad (57)$$

where

$$k_1 = \frac{r_0}{\sqrt{Fq}}\ ;\ k_2 = r_0 * \sqrt{Fq} \quad .$$

(58)

For the significance level q = 0.10 and the number of degrees of freedom n - 1 = 6 - 1 = 5, we find in the table the variance ratio F_q = 3.5, with such initial data the difference in the values of the correlation coefficient found by formulas (52) and (46) will be the largest, ie the worst case is considered.

Given different values of the theoretical correlation coefficient ρ, by formulas (58) we calculate the values of k_1 and k_2, by formula (42) the coefficient r_0, by formula (38) r' by formulas (52) and (46) r, all these values are given in table 12.

Table 12

According to the values of the theoretical correlation coefficient ρ, we will calculate the values using formulas (58) k_1 and k_2, according to formula (42) coefficient r_0, according to the formula (38) r', according to the formulas (52) and (46) r

ρ	Coefficients		Value of correlation coefficients			
	k_1	k_2	за (42) r_0	за (38) r'	за (52) r	за (46) r
0	0	0	0	0	0	0
0.1	0.05345	0.18708	0.100	0.116	0.127	0.152
0.2	0.10690	0.37417	0.200	0.232	0.251	0.296
0.3	0.16036	0.56125	0.300	0.346	0.371	0.427
0.4	0.21381	0.74833	0.400	0.456	0.484	0.543
0.5	0.26726	0.95541	0.500	0.562	0.590	0.645
0.6	0.32071	1.12250	0.600	0.664	0.688	0.734

Continuation of table 12

0.7	0.37417	1.30958	0.700	0.760	0.778	0.813
0.8	0.42762	1.49666	0.800	0.848	0.859	0.882
0.9	0.48107	1.68375	0.900	0.928	0.933	0.944
1.0	0.53452	1.87083	1.000	1.000	1.000	1.000

The correlation coefficients calculated by formula (52) were found for $k_1 = k_1$, but the values of r found in (52) are greater than in formula (38). The largest values of the correlation coefficients are reached when finding by formula (46).

Let's give an example, the measured values of displacements of points are taken from work [33], which are given in table. 4, according to which the correlation coefficients, coefficients of the regression model (linear model) were calculated, as well as the analysis of the relationships between independent values was carried out $S(h)$, $S(x)$, $S(y)$.

Table 13.

Measured values of point displacements

№ points	$S(h)$, mm	$S(x)$, mm	$S(y)$, mm	№ points	$S(h)$, mm	$S(x)$, mm	$S(y)$, mm	№ points	$S(h)$, mm	$S(x)$, mm	$S(y)$, mm
1	-5,0	-12,0	5,0	8	-4,0	-10,0	-3,0	15	-11,0	-8,0	12,0
2	-5,5	-10,0	7,0	9	-2,0	-8,0	7,0	16	-13,0	-8,0	9,0
3	-6,0	-4,0	15,0	10	-2,5	-5,0	-1,0	17	-15,0	-7,0	-1,0
4	-7,0	-3,0	10,0	11	-4,0	-1,0	0,0	18	-13,5	-4,0	0,0
5	-5,0	-7,0	9,0	12	-5,0	-5,0	-4,0	19	-10,0	1,0	4,0
6	-6,5	5,0	8,0	13	-7,0	0,0	5,0	20	-9,0	2,0	7,0
7	-3,0	6,0	-1,0	14	-9,0	2,0	9,0	21	-7,0	5,0	8,0

According to the measured values given in the table. 13, linear regression equations are defined

$$\begin{cases} y_i = 5.116392 + 0.034426 \cdot x_i \\ x_i = -3.56080 + 0.035971 \cdot y_i \end{cases} \quad (59)$$

The second equation (59) is rephrased according to equation (44), and we obtain

$$\begin{cases} y_i = 5.116392 + 0.034426 \cdot x_i \\ y_i = 98.993606 + x_i/0.35971 \end{cases} \tag{60}$$

According to the formula (45), we will find the angle between the lines (60), it will be

$$\varphi = 85.^o\,9682159 = 85^o 58' 05.'' 58$$

Correlation coefficients between these values $r_{h,x} = 0.012274$, $r_{h,y} = -0.14465$, $r_{x,y} = 0.03519$, that is, they indicate that there is a linear relationship between the values $S(h)$, $S(x)$, $S(y)$ does not have, and this is indicated by the fact that the angle between the regression equations is quite close to $\pi/2$ [31]

Conclusions. Thus, as can be seen from table. 12 the estimate of the correlation coefficient starting from $n = 6$, which is performed by formula (38), will be shifted, so a more accurate estimate can be obtained by formula (46), formula (52) is not taken into account because it is obtained for the ideal case when $k_1 = k_2 = r_0$, this situation is rare. The values of r found by formula (52) will be less than those found by formula (46). In essence, the coefficients obtained by formula (46) will be maximum at the appropriate level of significance and the number of degrees of freedom, obtained on the basis that the level of significance q> 0.25 and a large number of measurements, $Fq \to 1$.

Calculating more accurate values of correlation coefficients (normalized correlation moments) is important in digital photogrammetry in the mutual orientation and geodetic reference of a stereo pair of images, as well as in the study of systematic errors in a series of measurements. Underestimated values of correlation coefficients will affect the equations of the correlation function, and ultimately will give a distorted view of the magnitude of systematic errors.