4th International DAAAM Conference
"INDUSTRIAL ENGINEERING – INNOVATION AS COMPETITIVE EDGE FOR SME"
29 - 30th April 2004, Tallinn, Estonia

THE IMPORTANCE OF ORDER CORRELATION

Tatjana Barashkova
Tatjana.Barashkova@mail.ee
Tallinn University of Technology
Department of Mechatronics
Ehitajate tee 5, 19086 Tallinn, Estonia

ABSTRACT

It is known, that the correlation between two estimations of entrance sizes $\overline{\overline{U}}_0$ *and* $\overline{\overline{U}}_x$ *is characterized by the appreciated correlation factor [1]*

$$r(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\frac{s(\overline{\overline{U}}_0,\overline{\overline{U}}_x)}{s(\overline{\overline{U}}_0)\cdot s(\overline{\overline{U}}_x)} \qquad (1)$$

In practice the use of the formula (1), as a rule, is considered as an estimation of linear factor of the Pearson correlation. It is known, that the Pearson rank correlation coefficient is determined by the formula [2]

$$r_p(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\frac{\sum\limits_{i=1}^{n}\overline{U}_{0i}\overline{U}_{xi}-\overline{\overline{U}}_0\cdot\overline{\overline{U}}_x\cdot n}{\left\{\left[\sum\limits_{i=1}^{n}\overline{U}^2_{0i}-(\overline{\overline{U}}_0)^2\cdot n\right]\cdot\left[\sum\limits_{i=1}^{n}\overline{U}^2_{xi}-(\overline{\overline{U}}_x)^2\cdot n\right]\right\}^{1/2}} \qquad (2)$$

If, for example, monotonous dependence between entrance sizes $\overline{\overline{U}}_0$ *and* $\overline{\overline{U}}_x$ *is observed, then the factor of correlation is necessary to estimate in a different way. In the present article the Pearson and Spearman rank correlation coefficients are compared.*

Keywords: Statistic. Rank correlation coefficient

1. INTRODUCTION

As an example the results of checking two resistances of measures with nominal resistances to a constant current 100 mΩ were used. Ten repeated measurements of the decrease of voltage on resistance measures were conducted during five days. Under the equation (3) the Spearman rank correlation factors were determined [3]

$$r_s=1-\frac{6\sum\limits_{i=1}^{N}(n_{0i}-n_{xi})^2}{N(N^2-1)} \qquad (3)$$

where

n_{0i} – the rank of entrance size $\overline{U}_{0i}$

n_{xi} – the rank of entrance size $\overline{U}_{xi}$

N – quantity of pair supervisions $(\overline{U}_0,\overline{U}_x)$

In order to be convinced of the statistical importance of correlation factor the following comparison were conducted:

- the check of the importance of the Pearson and Spearman rank correlation factors
- the analysis of the emissions.

2. THE VARIANT OF SYSTEM AT AN ESTIMATION OF THE CORRELATION

It is know that estimation of any output sizes is impossible without the certification of its narrowness of correlation connection. Therefore in practice the use of the formula (1) is considered as an estimation of correlation factor by the method of A or B.

If the estimation of mutual uncertainty and covariance is determined using the above-stated formula, it is possible to receive the generalized estimation of correlation factor by the result of two measurements.

The article presents the variant of system at an estimation of correlation factor of the results of measurements, which take into account covariance evaluated as statistically as by the method of B. Apart from that in the present article the statistical estimation covariance is determined by the criterion of uniformity of several correlation factors, and the Pearson and Spearman rank correlation coefficients are compared.

Apart from that the possibility of the application of the method into practice of the best estimation covariance is shown.

2.1 Repeated direct measurement

It statistical information (of Type A) corresponds to actually indicated values. In the given example as dependent values the average of the several average decrease of voltage on resistance measure $\overline{\overline{U}}_x$ and on standard $\overline{\overline{U}}_0$ were considered. Mutual uncertainty, correlation of quantities $\overline{\overline{U}}_x$ and $\overline{\overline{U}}_0$ is determined by the formula

$$s(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}(\overline{U}_{0i}-\overline{\overline{U}}_0)(\overline{U}_{xi}-\overline{\overline{U}}_x) \qquad (4)$$

where

n – quantity of the dependent pairs of input data $\overline{U}_0,\overline{U}_x$.

The Pearson rank correlation coefficient is determined by the formula

$$r_P(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\frac{s(\overline{\overline{U}}_0,\overline{\overline{U}}_x)}{s(\overline{\overline{U}}_0)\cdot s(\overline{\overline{U}}_x)} \qquad (5)$$

where

$s(\overline{\overline{U}}_0), s(\overline{\overline{U}}_x)$ – the best estimates of the standard uncertainty $u(\overline{\overline{U}}_0), u(\overline{\overline{U}}_x)$.

2.2 Nonstatistical information (of Type B)

If the correlation between input quantities $\overline{\overline{U}}_0$ *u* $\overline{\overline{U}}_x$ cannot be neglected and there is no opportunity to determine factor of correlation experimentally, use this method. By the formula (9), (6) and (5) the evaluated correlation factor $r_P(\overline{\overline{U}}_0,\overline{\overline{U}}_x)$, connected with two estimations of input quantities $\overline{\overline{U}}_0$ *u* $\overline{\overline{U}}_x$, is determined

$$u_B^2(\overline{\overline{U}}_0)=\sum_{i=1}^{K}\left(\frac{\partial F}{\partial q_i}\right)^2\cdot u^2(q_i) \qquad (6)$$

where

$u_B^2(\overline{\overline{U}}_0)$ – estimate of dispersion of the input value $\overline{\overline{U}}_0$, dependent from random quantities q_i.

The tested dependence is described by function

$$\overline{\overline{U}}_0=F(q_1,q_2,\ldots,q_i,\ldots,q_K) \qquad (7)$$

Similarly

$$\overline{\overline{U}}_x=G(q_1,q_2,\ldots,q_i,\ldots,q_K) \qquad (8)$$

$$u_B(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\sum_{i=1}^{K}\frac{\partial F}{\partial q_i}\cdot\frac{\partial G}{\partial q_i}\cdot u^2(q_i) \qquad (9)$$

In practice the use of the formula (5), as a rule, is considered as an estimation of correlation factor by the method of A or B.

2.3 The generalized estimation of correlation factor between two measurands

The way of the estimation of correlation factor of the results of measurements, which take into account covariance evaluated as statistically as by the method of B. If the estimation of total uncertainty $u(\overline{\overline{U}}_0)$,$\cdot u(\overline{\overline{U}}_x)$ and $u(\overline{\overline{U}}_0,\overline{\overline{U}}_x)$ covariance is determined by the formula (10), using the above-stated formula (5), it is possible to receive the generalized estimation of correlation factor between two values.

$$u(\overline{\overline{U}}_{0,x})=\sqrt{s^2(\overline{\overline{U}}_{0,x})+u_B^2(\overline{\overline{U}}_{0,x})}$$
$$u(\overline{\overline{U}}_0,\overline{\overline{U}}_x)=\sqrt{s^2(\overline{\overline{U}}_0,\overline{\overline{U}}_x)+u_B^2(\overline{\overline{U}}_0,\overline{\overline{U}}_x)} \qquad (10)$$

The statistical estimation covariance $s(\overline{\overline{U}}_0,\overline{\overline{U}}_x)$ is determined by the criterion of uniformity of several correlation factors [3]. As an example the results of checking two resistances of measures with nominal resistances to a constant current 100 mΩ were used. Ten repeated measurements of the decrease of voltage on resistance measures were made during five days. Under the equation (1) the experimental correlation factors were determined [2]

$$r_i=\frac{s_i(U_0,U_x)}{\sqrt{s_i(U_0,U_0)s_i(U_x,U_x)}} \qquad (11)$$

As a result of the specified actions $i=\overline{1\ldots5}$ correlations factors were received, according to $n_1,n_2,\ldots,n_i$ samples of volumes. The quantity of pairs of dependent entrance sizes U_0, U_x in each sample can be various, but in this case all samples of one volume, $n_i=10$.

The estimations of selective correlation factors were determined with the help of Fisher transformation

$$r_i'=\operatorname{arcth} r_i(U_0,U_x)=\frac{1}{2}\ln\frac{1+r_i(U_0,U_x)}{1-r_i(U_0,U_x)} \qquad (12)$$

As the best estimation of correlation factor the statistics was used [3]

$$r'=\frac{\sum_{i=1}^{5}(n_i-3)\operatorname{arcth} r_i}{\sum_{i=1}^{5}(n_i-3)} \qquad (13)$$

The criterion of uniformity of several correlation factors is applied, as a rule, at deficiency of initial data. If, for example, monotonous dependence between entrance sizes $\overline{\overline{U}}_0$ and $\overline{\overline{U}}_x$ is observed, then the factor of correlation is necessary to estimate in a different way. In such case the Spearman rank correlation factors were determined.

2.4 The Spearman rank correlation factors definition

For example, in the third series of the experimental data monotonous dependence between entrance sizes $\overline{U}_0$ and $\overline{U}_x$ is observed. Under the equation (14) the Spearman rank correlation factors were determined [3]

$$r_s(\overline{U}_0,\overline{U}_x)=1-\frac{6\sum_{i=1}^{N}(n_{0i}-n_{xi})^2}{N(N^2-1)}=1-\frac{6\sum_{i=1}^{N}d_i^2}{N(N^2-1)} \qquad (14)$$

where

n_{0i} – the rank of entrance size U_{0i}

n_{xi} – the rank of entrance size U_{xi}

N – quantity of pair supervisions (U_0, U_x)

Table 1. Complete result of calculation of r_s

i	U_{0i}, mV	U_{xi}, mV	n_{0i}	n_{xi}	$d_i^2 = (n_{0i} - n_{xi})^2$
1	9.98877	9.984635	1	1.5	0.25
2	9.988795	9.984635	3	1.5	2.25
3	9.9888	9.98464	5	3.5	2.25
4	9.98881	9.984645	7	6	1
5	9.988795	9.98465	3	8.5	30.25
6	9.988815	9.98464	8.5	3.5	25
7	9.988795	9.984645	3	6	9
8	9.988805	9.984655	6	10	16
9	9.988815	9.984645	8.5	6	6.25
10	9.988825	9.98465	10	8.5	2.25

$\overline{U}_0 = 9.988803$ mV ; $\overline{U}_x = 9.984644$ mV ; $\sum_{i=1}^{N=10} d_i^2 = 94.5$

Estimated upper limit of conceivable deviations of rank factors of the Spearman correlation:

$$t_{lim} = r_s \sqrt{\frac{N-2}{1-r_s^2}} = 1.34 \qquad (15)$$

Value of $t_p(\nu)$ from Student`s distribution for probability β and degree of freedom $\nu = N-2$:

$$t_p(\nu) = t[N-2; \beta] = t[8; 68.27\,\%] = 1.07 \qquad (16)$$

If $t_{\lim} > t_p(\nu)$, it is assumed that Spearman rank correlation factors are statistically significant.

3. CONCLUSION

- The variant of system at an estimation of the correlation is universal way of definition of correlation coefficients
- In the article is presented the way of the estimation of correlation factor of the results of measurements, which take into account covariance evaluated as statistically as by the method of B.
- Apart from that the possibility of the application of the method into practice of the best estimation covariance is shown.
- In the article is presented the way of the estimation of correlation factor if, for example, the deficiency of initial data is observed.

4. BIBLIOGRAPHY

[1] Guide to the Expression of Uncertainty in Measurement. First edition. Genève: ISO, 1993

[2] Румшинский Л. З. Математическая обработка результатов эксперимента. – М.: Наука, 1971

[3] Айвазян С. А. и др. Прикладная статистика. Исследование зависимостей. Справочник. – М.: Финансы и статистика, 1985

The work is carried out the support of Estonian Ministry of Education Science Grant No 14250s03 and Estonian Science Grant No 5621