# CRITERION OF THE CONTROL OF EXPERIMENTAL DATA

E.M. FARHADZADEH, A.Z. MURADALIYEV, Y.Z. FARZALIYEV
*Azerbaijan Scientific-Research and Design-Prospecting Institute of Energetic*
*Zardabi ave., 94*

The cases are frequent, when at check of the assumption about expediency of classification of the experimental data to the given attribute it appears, that the results application of the appropriate criteria are opposite. The method of an estimation of a mistake the second sort for the data on duration of condition power block electrical station and method of comparison of a number of criteria is resulted.

There is enough often results of passive or active experiment appear connected with a number of attributes which versions can be submitted by a discrete number of values. Display of these versions casually also does not depend on the experimenter. The importance of attributes and their versions at the best is clear at an intuitive level. Classification of the data on the set versions of attributes results in sharp decrease in number of the data of experiment and is connected to an opportunity of fallacies and wrong decisions. The method below is resulted, allowing to avoid the specified mistakes. We shall enter some concepts. We shall present the data of experiment as some set of results of measurement of concrete parameter ($P$) and we shall designate it through $\{P\}_M$, where $M$ - number of experiments. But the data with preset values of versions of attributes we shall present as sample and we shall designate through $\{P\}_m$, where $m$ - number of the data of sample

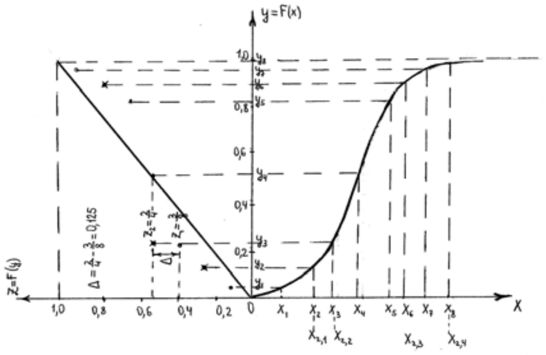

*Fig. 1.* A graphic illustration of consecutive calculation of statistics m for hypothesis $H_1$.

In [1] the new criterion of the control of imposing appearance of sample (if sample is representative an attribute of classification of the data insignificant) $\Delta_m$, describing the maximal value of a divergence of sample and a data set and its updating - factor of imposing appearance of sample $K_\Pi$, describing average quadratic value of a divergence of distributions and $\Delta_{cp}$, describing average value of divergences has been offered. Being analogues of Smirnov's nonparametric criterion $D_{m.n}$ [2], they had and the certain advantages among discrete values of sizes of divergences that reduced an interval between probabilities of adjacent values of the possible divergences, an including significance value $\alpha$.

Application of one of the criteria marked above demands development of methodology of their automated comparison, including development of methodology of an estimation of distributions of the criteria $\Delta_m$, $K_\Pi$, $\Delta_{cp}$ and $D_{m.n}$ provided that sample is not present. To distinguish distributions at hypotheses $H_0$ and $H_1$ we shall agree to represent them as $F^*(\Delta_m/H_0)$, $F^*(K_\Pi/H_0)$, $F^*(\Delta_{cp}/H_0)$ and $F^*(\Delta_{mn}/H_0)$, $F^*(K_\Pi/H_1)$, $F^*(\Delta_{cp}/H_1)$, $F^*(\Delta_{mn}/H_1)$.

The graphic illustration of consecutive calculation of statistics $\Delta_m$ for hypothesis $H_1$ is resulted on fig. 1. for $m$=4 and $M$=10

It is accepted, that sample is $\{X_{2,i}\}_m$, the random variables supplementing sample up to set is $\{X_{1,i}\}_n$, and set of random variables are $\{X_i\}_M$.

From fig. 1 it is evidently visible, that at calculation $\Delta_m$ for hypothesis $H_1$ biunique conformity between members of variations lines of random variables of set $\{X_i\}_M$ and members of variations lines $Y_j=F(X_i)_M$ is broken. For example, to the fifth member of lines $\{X_i\}_M$ there corresponds the seventh member of lines $\{Y_j\}_M$. Hence, *property of nonparametric (independence of criteria $\Delta_m$ , $K_\Pi$ , $\Delta_{cp}$ and $D_{m,n}$ from type of distributions $F_m(X)$ and $F_n(X)$) at check of hypothesis $H_1$ considered criteria lose).* Change of the law of distribution $F_m(x)$ and $F_n(x)$ is reflected in absolute size of each criterion. However, at any distributions $F_m(x)$ and $Fn(x)$ parity of each of $N$ realizations of statistic $\Delta_m$, $K_\Pi$ and $\Delta_{cp}$ and $D_{m.n}$ remains constant, namely $D_{m,n}>\Delta_m>K_\Pi>\Delta_{cp}$ Therefore change $F_m(x)$ and $F_n(x)$ will not influence result of comparison of these criteria.

On fig. 2 experimental dependences of distributions of the statistics $\Delta_m$ are resulted, allow us estimate character of change of distributions $F(\Delta_m)=\beta(\Delta_m)$ depending on change $m$ and $n$. As follows from fig. 2 with increase in number of random variables of sample ($m$) average value $\Delta_m$ and an average quadratic deviation $\sigma(\Delta_m)$ will decrease (we compare distributions 1 and 3 fig. 2).

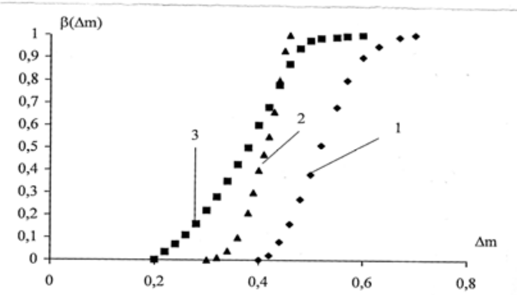


*Fig. 2.* Distribution of the greatest disorder of distributions $F_m(x)$ and $F_M(x)$ at hypothesis $H_1(\delta$=0,5) 1 - $m$ =10; $n$ =200; 2 - $m$ =10; $n$ =20; 3 - $m$=50; $n$=200

Distributions of random variables samples with identical m, taken of sets with differing number of random variables (for example $M_1$>>$M_2$), will have various $(\Delta_m)_{cp}$ and $\sigma(\Delta_m)$. Than M is more, the $(\Delta_m)_{cp}$ more and $\sigma(\Delta_m)$ less. The increase $\sigma(\Delta_m)$ at reduction of $M$ is caused by growing influence of casual character of distribution $F_M(\tau)$.

**Method of comparison of the criteria.** As it has been marked above, the preference is given to theories of check of statistical hypotheses to criterion for which at the fixed mistake of the first sort, the mistake of the second sort is minimal. Hence, to compare criteria it is necessary to compare dependencies $\beta(S) = f[\alpha(S)]$, where $S$ - one of considered ($\Delta_m$, $K_П$, $\Delta_{cp}$ and $D_{m,n}$) statistic.

Thus it is necessary to take into account:

1. Discrete character of sizes of the greatest deviation $S$. Addition $\alpha$ and $\beta$ also is allowable only on condition $\alpha$ and $\beta$ that are determined for same value $S$ with $i$=1, $r_\Sigma$ where $r_\Sigma$ - number of identical digitizations of distributions $F^*(S/H_0)$ and $F^*(S/H_1)$.

The size $r_\Sigma$ is defined:

- number distributions $F^*(S/H_0)$ and $F^*(S/H_1)$ (accordingly $r_1$ and $r_2$);

- a parity of boundary values of a confidential interval of average values of statistics $S$. The overlapping of these intervals is more, the more and $r_\Sigma$ ;

The certain difficulties at construction $\beta(S) = f[\alpha(S)]$ arise both at small, and at great values $m$ and $M$. At small values $m$ and $M$, sizes $r_1$ and $r_2$ also are small, $r_\Sigma << r_i$ and $r_\Sigma << r_2$. At great value $m$ and $M$, despite of essential increase $r_1$ and $r_2$, the size $r_\Sigma$ remains enough small though distinction between separate realizations $S$ does not exceed 1%. The increase $r_\Sigma$ without decrease in accuracy can be achieved by a method of a rounding off of values of realizations $S$ up to the second significant figure.

2. Features of statistical modeling. One of the basic difficulties at modeling estimations of optimum values of criterion checking to representativenesses of the sample) CRS $\Delta^*_{m,onm}$ is decrease in influence of casual character modeled samples on result of calculation. Fluctuations of numerical values $\Delta^*_{m,onm}$ cause the certain probability of the erroneous decision which, in particular, is more, than it is less number of realizations of sample and depends on number of iterations a little. Overcoming of this difficulty has been achieved as application of some known methods, as a method of the general random numbers together with a method of supplementing random variables [3], and new approaches, in particular:

- applications of criterion of Kolmogorov for the control of conformity programmed random numbers modeled sample to the uniform law "Sorting" samples at a significance value (0,3-0,4) not only reduces a mistake of the second sort, but also carries out protective functions from imperfection of program realizations samples and failures of the COMPUTER.

It is necessary to note, that the further increase in a significance value (more 0,3÷0,4) conducts to essential reduction of number of discrete values $\Delta_m$ and absence of identical values $\Delta_m$ for distributions $[1-F^*(\Delta_m/H_0)]$ and $F^*(\Delta_m/H_1)$.

- elimination of influence of the random variables supplementing sample up to a data set;

- alignment of number of discrete values of distributions $[1 - F^*(\Delta_m / H_o)]$ and $F^*(\Delta_m / H_1)$ by assignment by absent discrete value zero frequency. Application of these methods has allowed to provide high stability of the decision. The control of objectivity of recommended methods was provided with application of a method of the decision of a «return problem».

On fig. 3. experimental dependencies $\beta(S) = f[\alpha(S)]$, for the values marked above $m$, $M$ and $\delta$ for criteria $\Delta_m$, $K_П$, $\Delta_{cp}$ and $D_{m,n}$. Similar dependencies are received and for lines of other values $m_1$, $M$ and $\delta$. The analysis of these data has allowed concluding:

- the least value $\beta$ at 0<$\alpha$<1 takes place for criterion $\Delta_m$;

- values $\beta$ at 0<$\alpha$ <1 for criteria $K_П$ and $\Delta_{cp}$ also are practically equal and is essentially higher, than at criterion $\Delta_m$. In other words, distinction of distributions $F_m(P)$ and $F_m(P)$ also is defined not so much by average or average quadratic value of deviations, how many the greatest divergence.

- the indirect method of an estimation of representativenesses of sample (Smirnov's criterion) concedes to a direct method recommended authors, i.s. $\beta(D_m)>\beta(\Delta_m)$ at $0< \alpha <1$.

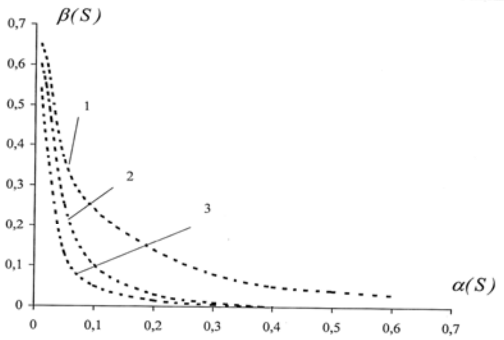


*Fig. 3.* Curve changes of dependence $\beta(S)=f[\alpha(S)]$ at $m$ =10; $M$=30 and $\delta$=0,5 for criteria $(S)$: 1- $K_П$ и $\Delta_{cp}$; 2 – $D_{m,n}$ ; 3 - $\Delta_m$


[1] *B.V. Gnedenko, J.K. Beljaev, A.D. Solovyov.* Mathematician's methods in the theory of reliability. M; "Science", 1965, with 524.

[2] *E.M. Farhadzadeh, A.Z. Muradaliyev, J.Z. Farzaliyev.* Estimation of accuracy of parameters of individual reliability of power units. Ekoenerji №2, 2006.

[3] *I.A. Ryabinin.* Bases of the theory and calculation of reliability of ship power systems. L., "Shipbuilding", 1971.

[4] *J. Kleynen*. Statistical methods in imitating modeling. M.: "Statistics", 1978, 224 with.


**Э.М. Фархадзаде, А.З. Мурадалиев, Ю.З. Фарзалиев**

## КРИТЕРИЙ КОНТРОЛЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

При проверке предположения о целесообразности классификации экспериментальных данных по заданному признаку результаты применения соответствующих критериев часто оказываются противоположными. Приводится метод оценки ошибки второго рода и метод сопоставления критериев.

**E.M. Fərhadzadə, A.Z. Muradəliyev, Y.Z. Fərzəliyev**

## TƏCRÜBİ VERİLƏNLƏRİN YOXLANMASI KRİTERİYALARI

Verilən əlamətlərə ğörə təcrübi verilənlərin təsnifatlarının məqsədəuyğunluğu haqqında olan gümanın yoxlanması zamanı uyğun kriteriyaların tətbiqi nəticələrə çox zaman əks olur. Bunun üçün aşağıda ikinci növ səhvlərin qiymətləndirilməsi və kriteriyalarının müqayisəsi üsulları göstərilir.

*Received: 08.07.06*