V.N. Kuzmin, S.N. Lapach «Stochastic Analysis of Short Samples Using Interval Functions» INTERNATIONAL CONFERENCE Stochastic Analysis and Random Dynamical Systems, June 14-20, 2009 Lviv, Ukraine. Abstract. P.139–140.

Предлагается прямое вычисление моды и определение наличия ассиметрия для малых выборок используя интервальную функцию

Кл. сл.
Малые выборки, интервальная функция, мода

Short Samples, Interval Functions, mode definition

Kuzmin V.N., Lapach S.N.

Stochastic analysis of small samples by interval function

Interval function in this study is a correspondence of intervals values between proximate elements of ordered sample from the interval mean value [1]. Let us assume that we have ordered sample of random variables

$$z1 < z2 < z3 < \ldots < zn \qquad (1)$$

Then we form two statistics

$$D_i = z_{i+1} - z_i \quad R_i = (z_{i+1} - z_i)/2 \qquad (2)$$

Their quantity is equal to K=n–1. We generate them in XOY system of coordinates, where $D_i \equiv y_i$, and $R_i \equiv x_i$. Let us approximate these values by parabola of the second order, using the method of least squares.

$$Y = a + bx + cx^2 \qquad (3)$$

Interval function, arranged in such a manner, affords the following:

1. direct mode calculation by formula Mo=–b/2c.
2. testing symmetric (asymmetric) by statistical significance analysis of b-coefficient.

For a range of problems for statistical data analysis direct and correct mode calculation by small samples can be more important than mean or median definition. In particular, it is related to the range of econometric and reliability analysis problems. Mode definition for small samples by histograming for samples with size of 9 – 25 is incorrect.

Investigation of properties of the above-mentioned interval function was carried out by simulation modelling. Were generated normally distributed N(0,1) random samples in size of 9, 15, 25, 49, 99. The quantity of samples for each size – 150. Each sample was ranged; median Me, mode Mo, mean, minimal and maximum values were calculated for each sample, also coefficients for interval function approximation a, b, c; Student’s *t*-test $t=b/s_b$ for definition of b-coefficient significance.

Although approximation parabola for normal distribution should be convex down (coefficient c>0), but for a range of samples in size of 9 (39,3%) and 15 (26,7%) it is convex up. It is another evidence of complexity of small-size samples analysis.

Upon samples size growth for normal distribution mode Mo and b-coefficient should tend to zero. The results of simulation modelling, given in Table 1, substantiate this assumption.

Table.1 Mode and b-coefficient means

n	9	15	25	49	99
Mo	-0,02662	-0,0097	-0,0336	0,025519	0,007231
b	-0,03107	-0,02198	0,000867	-0,00412	-0,00068

At that, b-coefficient converges to zero faster than mode. It follows from the experimental results that the application of interval function for mode definition can be recommended starting from the sample in size away from 49, for testing symmetric – away from 15.

Conclusions

1. The application of interval function for small samples statistical analysis affords reliable definition of symmetry properties.
2. In case of total absence of a priori information about distribution law, the interval function is presented by the single method of direct mode definition.
3. The method is not universal and should be applied in combination with other methods of analysis. But in many cases application of interval functions affords adjustment of small samples parameters.

References:

1. Kuzmin V.N. “The interval function as a tool for statistical analysis of a small samples” Materials of XI M. Kravchuk scientific conference, May 18-20, 2006, Kyiv. p.718.