Advances in Social Science, Education and Humanities Research, volume 646
Proceedings of the IX International Scientific and Practical Conference "Current Problems of Social and Labour Relations" (ISPC-CPSLR 2021)

Classification of the hypothesis testing reliability criteria in socio-economic research

Elena Galkina [1]

[1] *The Federal State Budgetary Educational Establishment of Higher Education «The Orel State University of Economics and Trade», Orel, Russia*
*Email: gev578@mail.ru

ABSTRACT
The article presents a methodology for assessing the acceptability of a hypothesis tested by a researcher based on statistical criteria applicable to the study results. The generally accepted criteria of reliability and power obtained by studying combinations of quantities of results recognised as true or false have been expanded by including the completeness criterion in the methodology. This criterion is obtained by supplementing the classification of research results with an area of undefined values. In addition to the general indicators of the reliability of observations (reliability, power and completeness), intermediate indicators of the data structure are proposed to clarify the results structure of the tests carried out. The universality of the model considered in work allows it to be applied in various socio-economic and other studies. The calculations results based on the model can be evaluated and used for decision-making, depending on the area and purpose of a particular study.

Keywords: *Reliability, Research, Criteria, Power.*

1. INTRODUCTION

In socio-economic, sociological and other areas of research, it is often necessary to confirm the validity of the hypothesis put forward by the researcher. Separate scientific and practical works, which have specialised areas, are devoted to the evaluating criteria methods. In particular, methods of forecasting and evaluation of complex systems are covered in the work of Isaev O.V., Isaeva K.V., Tolstykh O.V. [1], methods of evaluating efficiency in the energy-saving management system - in the work of Semenov V.N. [2], methods of forming criteria for the suitability for the use of information systems - in the study of Batenkina O.V. [3]. The evaluation of technical systems is touched upon in the works of Dudakov D.S. [4], Sulyan G.S. and Kazanchyan M.S. [5]. The works of Shenderey P.E., Shenderey E.E. [6], Ilyukhina T.A., Sergeeva V.V. [7], Ivashchenko E.N., Badanova A.A. are devoted to the issues of constructing criteria for assessing knowledge [8]. The importance of identifying information criteria for the audit of hypotheses is recognised not only for socio-economic matters but also for other areas, for example, chemical and technological systems [9].

2. RESEARCH METHODOLOGY

This study aims to develop a methodology for evaluating a hypothesis test by clarifying the classification of statistical criteria for such an assessment. Quantitative, well-formalised methods will allow the use of computer technologies for processing test results, which is especially important in the era of the digital economy. Specialists of information panels development of risk management, such as Bunting R. F Jr, Siegal D. emphasise: "Identifying the most powerful and informative measures, designing the most appropriate dashboards, and incorporating visual best practices are crucial steps required for evaluating the effectiveness and value of an enterprise risk management program» [10]. Domestic experts Shenderey P.E. and Shenderey E.E. also note the trend of automation of social research: "In recent years, the *machine method of evaluating* various sociological parameters has been increasingly used in sociology" [6].

Table 1. Data structure for evaluating the reliability and power of the criterion

		Assumed values (hypothesis)		
		True	False	Total estimated values
Test (control) results	True	Reliability of the criterion (hypothesis)	Inconsistency (theoretically False values that are actually True)	True(row)-True (column) plus True-False
	False	Mismatch (theoretically True values that are actually False)	The power of the criterion (hypothesis)	False-True plus False-False
	Total test results	True-True plus False-True	True-False plus False-False	True-True plus True-False plus False-True plus False-False

The study stages include presenting generally accepted criteria for testing the hypothesis, the addition of these criteria and thus the expansion of the classification model, and the approbation of the model on empirical data.

The methods used include general scientific methods (analysis and synthesis, abstraction and concretisation, modelling) and special methods (grouping, arithmetic).

Model numerical values were used as analytical data.

3. RESEARCH RESULTS

The basis for checking the reliability of a hypothesis made by the researcher is percentage criteria of compliance level of the assumptions about the phenomenon nature (the hypothesis provisions a priori) and test results (the hypothesis conditions a posteriori). Such percentage criteria are reliability and power. In general, the reliability of hypothesis testing (criterion reliability) is the percentage of true (positive) values recognised as true by the test results. The hypothesis testing power (criterion power) is the percentage of false (negative) values recognised as false by the test results. True or false values are understood as values, respectively, confirming or not confirming the hypothesis (or having the opposite content for the hypothesis). At the same time, such a method of assessing reliability assumes that the researcher has information about the deemed true values, which is usually expressed in the form of some statistical distribution of the values of the studied quantity, for which the hypothesis is checked (distribution). In this case, the test results are compared with the statistical distribution. However, the

Table 2. Data structure for evaluating the reliability and power of the criterion (model values)

		Assumed values (hypothesis)		
		True	False	Total estimated values
Test (control) results	True	70%	4%	74%
	False	6%	20%	26%
	Total test results	76%	24%	100%

results of subsequent studies, for example, control and audit measures on socio-economic issues, can also be used as assumed true values. Then the test results are compared with the control results. The data structure for evaluating the reliability and power of the criterion is presented in Table 1. In the cells of the final row and column of Table 1, the names of the cells whose values are summed up when calculating the total are indicated.

Table 3. Data structure for evaluating the reliability, power and completeness of the criterion

		Assumed values (hypothesis)			
		True	Not defined	False	Total estimated values
Test (control) results	True	Reliability of the criterion (hypothesis)	*Uncertainty revealed as True*	Inconsistency (theoretically False values that are actually True)	True-True plus True-ND* plus True-False
	Not defined	*True, not confirmed due to lack of data*	*Confirmed uncertainty (incompleteness)*	*False, not confirmed due to lack of data*	ND-True plus ND-ND plus ND-False
	False	Mismatch (theoretically True values that are actually False)	*Uncertainty revealed as False*	The power of the criterion (hypothesis)	False-True plus False-ND plus False-False
	Total test results	True-True plus ND-True plus False-True	True-ND plus ND-ND plus False-ND	True-False Plus ND-False plus False-False	True-True plus True-ND plus True-False plus ND-True plus ND-ND plus ND-False plus False-True plus False-ND plus False-False

*ND – not defined.

Table 2 shows an illustration of the structure of Table 1 with model values.

According to the model example, the reliability of the hypothesis is 70%, even though the true values are 4% more. The power of the hypothesis is 20%, even though there are 6% more false values. The cumulative reliability of the research results is calculated as the sum of reliability and power 70%+20% and is equal to 90%. Incorrectly predicted results are 100%-90% or 10%.

At the same time, in practical work, especially in socio-economic research, a situation may arise when the results of a study on any observation are missing or unreliable, for example, the respondent refused to answer the question. In this case, in addition to true and false values, the third type of value arises – "indefinite" or "absent"). Therefore, it is necessary to expand the classification of criteria for the reliability of the hypothesis, adding a zone of uncertain values to it. Such an extended classification is presented in table 3. The presence or absence of uncertain values (the "grey zone" in Table 3) characterises the degree of incompleteness or completeness of the criterion (hypothesis). The presence of uncertain values at the hypothesis level means that the hypothesis itself assumes some level of inaccuracy in data classification. Uncertain values at the results level in the absence of such values in the hypothesis means an incomplete collection of information during the study. The completeness of the study is achieved in the absence of values not defined during the study.

Table 4 presents a model calculation of the data structure for assessing the criterion's reliability, power, and completeness. Such data can be, for example, the results of past and current sociological research of self-assessment of the level of material security of households (where True is the recognition of material security, uncertainty is the absence of an answer or the option "I find it difficult to answer", False is the recognition of insufficient material security).

The model data in Table 4 show that the reliability of the hypothesis according to the test results is 65%, with a total share of true values of 74%. The power of the hypothesis is 9% with a total share of false values of 20%. The incompleteness of the hypothesis classification (uncertainty) is 2%, with a total number of actually undefined values of 6%. The final accuracy of the hypothesis predictions (cumulative reliability) is equal to 65%+9%+2%=76%. Incorrectly predicted results are 100%-76% or 24%.

In addition to the generalised indicators of the reliability of observations (reliability, power and completeness), intermediate indicators of the data structure can be calculated to clarify the picture of the test results obtained. Such indicators characterise the ratio of other values of cells in Table 4 that were not used in early calculations. The following coefficients can be selected as intermediate indicators of the data structure (the calculation procedure based on the model data in Table 4 is shown in parentheses):

- the confirmation coefficient if the hypothesis is True (74%/71%=1.04 with rounding);

- the coefficient of confirmation of the hypothesis falsity (20%/16% =1.25);

- the coefficient of confirmation of the hypothesis uncertainty (6%/13%=0.46 with rounding).

Thus, according to the model data, the true values turned out to be 4% more than expected, false ones - 25% more than expected, undefined values - 54% less than expected (that is, the results of the study turned out to be

Table 4. Data structure for evaluation of reliability, power and completeness of the criterion (model values)

		Assumed values (hypothesis)			
		True	Not defined	False	Total estimated values
Test (control) results	True	65%	*5%*	4%	74%
	Not defined	*1%*	*2%*	3%	6%
	False	5%	*6%*	9%	20%
	Total test results	71%	13%	16%	100%

more accurate than the hypothesis). The results of calculations are interpreted and used depending on the objectives of a particular study.

If it is necessary to detail the results analysis further, the ratio indicators of individual cells in Table 4 can be calculated in pairs. Such indicators will demonstrate how many times one value (numerator) differs from another (denominator).

4. CONCLUSIONS

Thus, as a result of the study, a universal model for evaluating hypotheses was obtained, applicable in various socio-economic and other studies. The author expanded the classification of criteria for the reliability of the hypothesis, supplemented by criteria of uncertain values. The calculations carried out using the model are quantifiable measurable and can be used to justify management decisions.

REFERENCES

[1] O.V. Isaev, K.V. Isayeva, O.V. Tolstykh, Methods of forecasting and evaluating the effectiveness of complex systems in conditions of various threats, criteria and evaluation parameters of the models under study [Metody prognozirovaniya i ocenki effektivnosti slozhnyh sistem v usloviyah mnogoobraziya ugroz, kriteriev i ocenochnyh parametrov issleduemyh modelej], Bulletin of the Voronezh Institute of the Federal Penitentiary Service of Russia [Vestnik Voronezhskogo instituta FSIN Rossii] 3 (2020) 53-59.

[2] V.N. Semenov, Methods of substantiation of factors, indicators and criteria for evaluating efficiency in the energy-saving management system of life support of municipal enterprises and the population [Metody obosnovaniya faktorov, pokazatelej i kriteriev ocenki effektivnosti v sisteme upravleniya energosberezheniem zhizneobespecheniya municipal'nyh predpriyatij i naseleniya], Bulletin of the Volgograd State University of Architecture and Civil Engineering. Series: Construction and Architecture [Vestnik Volgogradskogo gosudarstvennogo arhitekturno-stroitel'nogo universiteta. Seriya: Stroitel'stvo i arhitektura] 18(37) (2010) 104-109.

[3] O.V. Batenkina, Methods for assessing user satisfaction as one of the criteria for the usability of information systems [Metody ocenki udovletvorennosti pol'zovatelej kak odnogo iz kriteriev yuzabiliti informacionnyh sistem], In the collection: Visual culture: design, advertising, information technology, Proceedings of the XV International Scientific and Practical Conference [V sbornike: Vizual'naya kul'tura: dizajn, reklama, informacionnye tekhnologii. sbornik trudov XV mezhdunarodnoj nauchno-prakticheskoj konferencii], 2016, pp. 105-107.

[4] D.S. Dudakov, Problems of determining criteria for the quality of the urban environment and their assessment methods [Problemy opredeleniya kriteriev kachestva gorodskoj sredy i metody ih ocenki], In the collection: Science, Education and experimental design, The works of MARHI. Materials of the international scientific and practical conference, Collection of articles [V sbornike: Nauka, obrazovanie i eksperimental'noe proektirovanie, Trudy MARHI, Materialy mezhdunarodnoj nauchno-prakticheskoj konferencii. Sbornik statej], 2016, pp. 43-45.

[5] G.S. Sulyan, M.S. Kazanchyan, Methods of construction and analysis of criteria for evaluating the effectiveness of technical projects [Metody postroeniya i analiz kriteriev ocenki effektivnosti tekhnicheskih proektov], Scientific Notes of the Russian State Social University [Uchenye zapiski Rossijskogo gosudarstvennogo social'nogo universiteta] 7-2(70) (2009) 164-169.

[6] P.E. Shenderey, E.E. Shenderey, Methods of folding hierarchical criteria for assessing knowledge [Metody svertyvaniya ierarhicheskih kriteriev ocenki znanij], Secondary Vocational Education [Srednee professional'noe obrazovanie] 7 (2007) 45-49.

[7] T.A. Ilyukhina, V.V. Sergeeva, Methods of providing criteria for assessing the progress in physics of engineering students [Metody obespecheniya kriteriev ocenki uspevaemosti po fizike studentov inzhenernyh special'nostej], In the collection: Innovative technologies in improving the quality of education, Materials of the international scientific and practical conference, In 2 parts [V sbornike: Innovacionnye tekhnologii v sovershenstvovanii kachestva obrazovaniya. Materialy mezhdunarodnoj nauchno-prakticheskoj konferencii. V 2-h chastyah], 2017, pp. 106-109.

[8] E.N. Ivashchenko, A.A. Badanov, Methods for evaluating the criteria for the effectiveness of teaching computer science to cadets in military universities of the internal troops of the Ministry of Internal Affairs of Russia [Metody ocenki kriteriev effektivnosti obucheniya informatike kursantov v voennyh vuzah vnutrennih vojsk MVD Rossii], In the collection: Directions and prospects for the development of education in military institutes of the internal forces of the Ministry of Internal Affairs of Russia, Collection of scientific articles of the VII

International Scientific and Practical Conference: In 2 parts. Under the general ed. S.A. Kutsenko [V sbornike: Napravleniya i perspektivy razvitiya obrazovaniya v voennyh institutah vnutrennih vojsk MVD Rossii. Sbornik nauchnyh statej VII Mezhdunarodnoj nauchno-prakticheskoj konferencii: V 2 chastyah. Pod obshch. red. S.A. Kucenko], 2016, pp. 177-183.

[9] V.V. Kafarov, V.P. Meshalkin, G.I. Manko, Informational criterion of audit of the distribution hypotheses for reliability characteristics of chemicotechnological systems, Doklady Akademii nauk SSSR 249(4) (1979) 923-927.

[10] R.F.Jr. Bunting, D. Siegal, Developing risk management dashboards using risk and quality measures: A visual best practices approach, Journal of healthcare risk management: the journal of the American Society for Healthcare Risk Management 2(37) (2017) 8-28. DOI: https://doi.org/10.1002/jhrm.21287