Application of Homogeneity Tests: Problems and Solution

B. Yu. Lemeshko[*], S. B. Lemeshko[*], I. V. Veretel'nikova[*], A. Yu. Novikova[*]

[*] *Department of Theoretical and Applied Informatics,*
Novosibirsk State Technical University,
20, Prospekt K. Marksa, Novosibirsk, 630073, Russia

Abstract. In this paper, we consider the application of k-sampling criteria for homogeneity with the possibility of calculating the achieved significance level p-value. The approximating models for the limiting distributions of the k-sample Anderson–Darling test statistics have been obtained. A comparative analysis of the homogeneity tests in terms of power has been carried out. A lot of parametric and nonparametric tests for testing the homogeneity of mathematical expectations are considered. Distributions of classical test statistics for homogeneity of variance have been investigated including the case, when the standard assumption of the normality is violated. A comparative analysis of the power of many parametric and nonparametric criteria, including the case of the violation of normality assumption. The obtained results enable us to draw correct statistical conclusions using the considered tests, including their application in the case of the violation of standard assumptions.

Keywords: homogeneity tests, homogeneity of of means, homogeneity of variances, power of tests.

1. Introduction

The criteria for testing the hypothesis of homogeneity are divided into three sets: criteria for testing the hypothesis about the homogeneity of the laws of the k samples being compared, on the homogeneity of means (on the equality of average), on the homogeneity of variances (on the equality of variances).

When checking the homogeneity of laws, the hypothesis that all samples extracted from the same general population is verified

$$H_0 : F_1(x) = F_2(x) = \cdots = F_k(x)$$

for any x. When testing the hypothesis of homogeneity of means, the hypothesis being tested has the form $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, and when the dispersions are homogeneous – $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$.

To test the homogeneity of laws, nonparametric tests can be used [1] of Smirnov, Lehmann–Rosenblatt, Anderson–Darling tests, Zhang test statistics Z_K, Z_A, Z_C. To test the hypothesis of homogeneity of means, one can use [1] a number of parametric tests (z-test for comparing two averages with known variances; for unknown, but equal variances (t-Student's test);

for unknown and unequal variances; k-sampling F-test) and nonparametric tests (Wilcoxon test, $\tilde{z}$-Mann – Whitney test, Kruskal – Wallis H test, Van der Waerden test, Fisher – Yates – Terry – Hoeffding test, k-sampling Van der Waerden test).

To test the hypothesis of homogeneity of variances, we have a set of parametric (Bartlett, Cochran, Fisher, Hartley, Levene, Neyman – Pearson, O'Brien, Link, Newman, Bliss – Cochran – Tukey, Cadwell – Leslie – Brown, Overall – Woodward, Z-variance and modified Overall – Woodward Z-variance tests) and nonparametric (Ansari – Bradley, Mood, Siegel – Tukey, Capon, Klotz, Fligner – Killeen tests) tests [1].

The application of parametric tests is due to the assumption that the analyzed particles belong to a partial parametric law. The non-parametric tests on the homogeneity of the mean or characteristic scattering changes do not imply the assumption of normality. However, it is required to fulfill an equally strong assumption about the belonging of the analyzed samples of one type of law (which, as a rule, is not mentioned).

2. Problems of application

Several general problems are associated with the application of the tests. First, the presence of a set of tests for testing the same hypothesis makes the researcher choose the most preferable criterion, which has a higher power with respect to the alternatives in question. A comparative analysis of the power of the groups of tests can be made only on the basis of numerical methods of statistical modeling, which requires appropriate software.

Secondly, in the real world of applications, the standard assumption that samples belong to the normal law, as a rule, is not satisfied or difficult to test due to insufficient sample sizes. Due to the violation of this assumption, the distribution of statistics of parametric tests corresponding to the validity of the hypothesis being tested H_0, can differ from the available classical results. This is especially critical for parametric tests of dispersion homogeneity.

Thirdly, with limited sample sizes, the distributions of statisticians of parametric tests often differ significantly from the known asymptotic distributions of these statistics.

Fourth, even in the case of the standard assumption, the possibility of correct application of a number of parametric tests is limited by the fact that statistics distributions are not known, and there are only tables of critical values of statistics for a certain number of sample volumes. Therefore, one can not estimate the level of significance achieved p_{value}. For Zhang's homogeneity tests with statisticians Z_K, Z_A, Z_C there are no tables of critical values, and statistical distributions for justice H_0 depend on the volumes of the samples being compared and the number of samples.

Fifth, distributions of normalized statistics of nonparametric tests are discrete and for small sample sizes differ significantly from the asymptotic standard normal law.

Sixth, the non-parametric criteria for homogeneity of means or tests for the homogeneity of variances (equality of scale parameters) does not imply the assumption of normality. However, an equally strong assumption is required that the laws of the samples analyzed are homogeneous. This is especially important for nonparametric tests for the homogeneity of variances.

Seventh, parametric tests have advantage in power over nonparametric analogs, including under conditions of violation of the standard assumption (in the case of laws that differ from normal ones). This is most noticeable in the case of tests for checking the homogeneity of variances. This fact testifies to the use of parametric tests for the homogeneity of variances under conditions of violation of the standard assumption of normality. To implement this possibility and ensure the correctness of the conclusion about the results of testing the hypothesis, one only needs to be able to find (estimate) the distribution of statistics $G(S \mid H_0)$. The applied tests under non-standard conditions (in the real conditions of the application). This is quite feasible with the use of statistical modeling methods.

3. Tests for the homogeneity of laws

Problems that should be taken into account when using the tests of homogeneity of laws include the essential discreteness of the distribution of statistics of the Smirnov test and the essential difference between this distribution and the Kolmogorov limit distribution.

A factor hampering the application of the Zhang homogeneity test with statisticians Z_K, Z_A, Z_C, is the dependence of the distributions of statistics on the volumes of the samples being compared and the number of samples and the discreteness of the distribution of statistics Z_K. Therefore, in order to estimate p_{value}, it is preferable to use the real distributions of statistics found using statistical modeling.

Distributions of statistics Lehmann–Rosenblatt, Anderson–Darling—Pettitt quickly converge to the limit distributions. For k-sampling Variant of the Anderson-Darling criterion, approximate models of limit distributions of statistics (for $k = 2 \div 11$) are constructed and presented in [1].

The power of the tests by statistical modeling methods was investigated with respect to three types of alternatives: a shift, a change in scale, and relative to a situation where a pair of samples belonged to close but different laws (normal and logistic). The obtained power estimates allow us to draw the following conclusions.

Concerning competing hypotheses corresponding to a change in the shift parameter, two-sample Smirnov (Sm), Lehmann–Rosenblatt (LR), Anderson–Darling–Pettitt (AD) tests and Zhang's test with statisticians Z_K, Z_A, Z_C descending power in the following order ("$\succ$" is preference):

$$AD \succ LR \succ Z_C \succ Z_A \succ Sm \succ Z_K,$$

Against competing hypotheses corresponding to a change in the scale parameter, the criteria are arranged in a different order:

$$Z_A \succ Z_C \succ Z_K \succ AD \succ LR \succ Sm.$$

In a situation where if a competing hypothesis, one sample belongs to a normal distribution, and the second logistic criteria ordered by the power as follows:

$$Z_K \succ Z_A \succ Z_C \succ AD \succ Sm \succ LR.$$

In the case of the k samples in similar situations the same order of preference is saved for k-sampling options test for Anderson-Darling and Zhang.

4. Test for the homogeneity of variances

To test the hypothesis of homogeneity of variances, we propose a set of parametric (Bartlett (B), Cochran (C), Fisher (F), Hartley (H), Levene (L), Neyman–Pearson (NP), O'Brien (OB), Newman (N), Overall– Woodward Z-variance (Z) modified Overall–Woodward Z-variance tests (MZ), Bliss– Cochran–Tukey (BCT), Cadwell–Leslie–Brown (KLB), Link (Lk)) and nonparametric (Ansari– Bradley (AB), Mood (M), Siegel–Tukey (ST), Capon (Ca), Klotz (Kl), Fligner – Killeen (FK)) tests [1].

The violation of the assumption of normality leads to significant changes in the distributions $G(S \mid H_0)$ statistics of parametric test, which excludes the possibility of using classical results concerning distributions of statistics. To a lesser extent, this remark concerns the O'Brien, Overall–Woodward Z-variance tests and Levene test.

In the case of the analysis of two samples and the fulfillment of the standard assumption of normality, the general picture is as follows:

$$(F \sim B \sim C \sim H \sim NP \sim Z) \succ OB \succ MZ \succ (Kl \sim Ca) \succ L \succ$$
$$\succ FK \succ M \succ N \succ (AB \sim ST) \succ (BCT \sim KLB \sim Lk).$$

If the standard assumption is violated and two samples are assigned to laws with lighter tails than the normal law, the above order of preference is retained.

For symmetric laws with heavier tails compared with the normal law, the criteria are ordered as follows:

$$FK \succ Kl \succ M \succ L \succ (AB \sim ST) \succ OB \succ MZ \succ (F \sim B \sim C \sim H \sim$$
$$\sim NP \sim Z) \succ N \succ (BCT \sim KLB).$$

If a larger number of samples are analyzed, then, under the standard assumption, or in the case of sampling of laws with lighter tails compared to normal law, the criteria for decreasing power are arranged in the following order:

$$C \succ OB \succ Z \succ MZ \succ (B \sim NP) \succ H \succ L \succ FK \succ BCT \succ KLB.$$

It should be noted that for symmetric laws with lighter tails, Levene test is inferior in power to the Fligner–Killeen test.

With symmetric laws with heavier tails, the situation changes:

$$FK \succ L \succ OB \succ MZ \succ (B \sim NP) \succ Z \succ H \succ C \succ KLB \succ BCT.$$

5. Application of tests in "non-standard" conditions

Deciding on the results of hypothesis testing H_0 on the basis of the achieved level of significance p_{value} is always more justified than by comparing the obtained value of the statistics S^* with the given critical values extracted from the corresponding table.

In the presence of appropriate software in a situation where the distribution of the statistics of the criterion used to test a certain hypothesis is unknown at the time of the start of the test (for various reasons), this distribution can be investigated using statistical modeling techniques in an interactive mode (directly in the process of analysis). The empirical distribution of statistics can be constructed for those volumes n_i, which correspond to the analyzed samples, under conditions of specific assumptions, possibly differing from the standard ones. Realized interactive procedure for distributing statistics makes it possible to apply various test for homogeneity, including under conditions of violation of standard assumptions.

6. Conclusions

In the work, many tests of homogeneity (homogeneity of laws, homogeneity of means, homogeneity of variances) were tested. The corresponding groups of tests are ordered by power. The opportunity of research of distribution of statistics in an interactive mode is realized. This makes it possible to apply the tests in the conditions of violation of standard assumptions, making it possible to calculate the estimates p_{value} which increases the informativeness of statistical conclusions. An obligatory preliminary condition for the application of test in non-standard conditions is the identification of the type of distribution law that best describes the analyzed samples [2].

Acknowledgments

References

1. *Lemeshko B. Yu.* Tests for homogeneity. Guide on the application. — M. : INFRA-M, 2017. — DOI: 10.12737/22368.
2. *Lemeshko B. Yu., Lemeshko S. B., Postovalov S. N., and Chimitova E. V.* Statistical Data Analysis, Simulation and Study of Probability Regularities. Computer Approach: NSTU Monographs. — NSTU Publishing House, Novosibirsk, 2011.