Aspects of statistical quality control of products: choice of criteria for uniformity of dispersions

A.V. Faddeenkov\textsuperscript{1}, L.A. Landman\textsuperscript{1}

\textsuperscript{1}Novosibirsk State Technical University, 20, Karla Marksa ave., Novosibirsk, 630073, Russia
E-mail: faddeenkov@corp.nstu.ru

Abstract. Analysis and data processing methods allowing to identify nature of the obtained data and their interrelation quickly and accurately are widely demanded now. To identify the relationship between input and output variables, a researcher faces regression analysis tasks, in particular, the estimation of unknown parameters of regression models. In this case application of classical methods is limited by a number of assumptions, in particular, distribution heterogeneity on the measurement region. Using of various stable, nonparametric and adaptive estimation methods in the presence of heterogeneity in the random errors distribution has not been sufficiently studied on the measurement region, therefore the isolation of homogeneous sub-regions in the general heterogeneity measurement region could help to solve the problem of parameter estimation. The criteria for testing homogeneity hypotheses are very often used in various applications of the problems of statistical analysis. In this case we can talk about testing distribution homogeneity hypotheses corresponding to the samples being analyzed or about mean homogeneity or about variances homogeneity. The aim of the study is to investigate the power of parametric and nonparametric criteria under various competing hypotheses.

1. Introduction
Nowadays, the methods of data analysis and processing are widely demanded to accurately identify the relationship between variables. To identify the relationship between input and output variables, the researcher is faced with the tasks of regression analysis, in particular, with the estimation of unknown parameters of the regression models. At the same time, the use of classical methods is limited by the assumption of the homogeneity of the distributions of random errors. One type of heterogeneity is the heterogeneity of dispersions. Among the many criteria for testing the uniformity of dispersions, it is necessary to choose the most powerful.

2. Problem statement
The main task is to test the statistical hypothesis of the constancy of the dispersion of two samples:

$$H_0 : \sigma_1^2 = \sigma_2^2.$$  
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

In the literature, among the parametric criteria for checking the homogeneity of dispersions, Bartlett, Cochran, Hartley, Levene criteria and ANOVA criterion are noted. Non-parametric criteria include criteria Ansari-Bradley, Siegel-Tukey, Mood, Klotz. These criteria differ significantly in their statistical properties. The main interest is the comparative power of these criteria with different amounts of baseline data, different alternatives to the main hypothesis.
The evaluation of the power of the criteria was carried out using the Monte Carlo numerical statistical simulation.

For each of the criteria, the power estimate was calculated according to the following algorithm:
Step 1. Set a critical value for a given level of significance \( \alpha \).
Step 2. Set \( m = 0 \).
Step 3. Generate a sample of \( Y \) subject to the validity of the hypothesis \( H_0 \).
Step 4. Calculate the criterion statistics value.
Step 5. If statistics falls into a critical area with a given level of significance \( \alpha \), then \( m = m+1 \).
Step 6. Repeat steps 3-5 \( n \) times.

At the end of this algorithm, the criterion power estimate is calculated as \( m/n \).

The number of independent samples modeled in the Monte Carlo method was \( 10^6 \). This number of repetitions made it possible to achieve an accuracy of the obtained power estimates of at least 0.001.

3. Results of computational experiments
A comparative analysis of the power of the criteria considered the following competing hypotheses:
\[
H_0: \sigma_2 = \sigma_1,
H_1: \sigma_2 = 1.1\sigma_1,
H_2: \sigma_2 = 1.2\sigma_1,
H_3: \sigma_2 = 1.5\sigma_1.
\]

Studies power were carried out on samples of size \( N = 20, 30, 50, 80, 100, 200, 300, 400, 500, 1000 \) observations. The level of significance was chosen equal to \( \alpha = 0.01 \). The studies of power were carried out with the normal, logistic and the Laplace distribution laws, since these distributions are symmetrical relative to the mean, which is consistent with the notion of an error in the regression model. Next, we consider the results obtained for the normal distribution law.

As can be seen from table 1, with a small change in the dispersion (by 10\%), the parametric criteria of Hartley, Bartlett, Cochran, Levene demonstrate a higher power than non-parametric ones. Among non-parametric criteria, the best power indicators are in the Ansari-Bradley criterion. In the case when the sample contains more than 200 observations, quite good results are shown by the Klotz and Mood criteria.

| Table 1. Power estimation criteria for correct hypothesis \( H_1 \) at various \( N \) for the normal distribution law |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Criterion       | \( N=20 \) | \( N=40 \) | \( N=60 \) | \( N=80 \) | \( N=100 \) | \( N=200 \) | \( N=300 \) | \( N=400 \) | \( N=500 \) | \( N=1000 \) |
| ANOVA           | 0.05  | 0.06  | 0.07  | 0.08  | 0.10  | 0.14  | 0.21  | 0.27  | 0.33  | 0.52  |
| Ansari-Bradley  | 0.06  | 0.07  | 0.10  | 0.11  | 0.11  | 0.19  | 0.28  | 0.31  | 0.34  | 0.65  |
| Klotz           | 0.02  | 0.02  | 0.04  | 0.05  | 0.05  | 0.14  | 0.22  | 0.29  | 0.34  | 0.73  |
| Mood            | 0.02  | 0.02  | 0.02  | 0.02  | 0.02  | 0.18  | 0.22  | 0.32  | 0.36  | 0.70  |
| Siegel-Tukey    | 0.03  | 0.04  | 0.04  | 0.05  | 0.05  | 0.19  | 0.19  | 0.28  | 0.32  | 0.45  |
| Bartlett        | 0.08  | 0.10  | 0.12  | 0.13  | 0.16  | 0.23  | 0.39  | 0.49  | 0.56  | 0.84  |
| Cochran         | 0.07  | 0.10  | 0.12  | 0.13  | 0.16  | 0.24  | 0.40  | 0.48  | 0.54  | 0.84  |
| Hartley         | 0.07  | 0.10  | 0.12  | 0.13  | 0.16  | 0.25  | 0.40  | 0.49  | 0.55  | 0.84  |
| Levene          | 0.06  | 0.09  | 0.11  | 0.11  | 0.14  | 0.21  | 0.36  | 0.40  | 0.48  | 0.80  |

If the sample belongs to the logistic distribution or the Laplace distribution, the power indicators of the Klotz criterion are worse, but the Siegel-Tukey criterion demonstrates good performance.

If the dispersion changes by 30\% (see Table 2), the parametric criteria of Hartley, Bartlett, Cochran and Levene also demonstrate higher power than non-parametric ones. Among non-parametric criteria, the best power indicators are in the Ansari-Bradley criterion. In the case when the sample contains more
than 200 observations, good results show the criteria of Klotz, Siegel-Tukey and Mood.
With laws different from the normal, the Hartley, Bartlett, Cochran, and Leuven criteria show the greatest power, although, in general, the power decreases.

Table 2. Power estimation criteria for correct hypothesis $H_2$ at various N for the normal distribution law

| Criterion  | N=20 | N=40 | N=60 | N=80 | N=100 | N=200 | N=300 | N=400 | N=500 | N=1000 |
|------------|------|------|------|------|-------|-------|-------|-------|-------|--------|
| ANOVA      | 0.08 | 0.15 | 0.23 | 0.36 | 0.45  | 0.70  | 0.86  | 0.95  | 0.99  | 1.00   |
| Ansari-Bradley | 0.12 | 0.24 | 0.36 | 0.42 | 0.49  | 0.80  | 0.94  | 0.98  | 0.99  | 1.00   |
| Klotz      | 0.04 | 0.14 | 0.29 | 0.36 | 0.47  | 0.85  | 0.98  | 1.00  | 1.00  | 1.00   |
| Mood       | 0.05 | 0.11 | 0.20 | 0.23 | 0.32  | 0.87  | 0.98  | 0.99  | 1.00  | 1.00   |
| Siegel-Tukey | 0.07 | 0.14 | 0.23 | 0.27 | 0.33  | 0.80  | 0.94  | 0.98  | 0.98  | 1.00   |
| Bartlett   | 0.21 | 0.44 | 0.54 | 0.60 | 0.73  | 0.94  | 1.00  | 1.00  | 1.00  | 1.00   |
| Cochran    | 0.21 | 0.44 | 0.54 | 0.60 | 0.73  | 0.94  | 1.00  | 1.00  | 1.00  | 1.00   |
| Hartley    | 0.21 | 0.44 | 0.54 | 0.60 | 0.74  | 0.94  | 1.00  | 1.00  | 1.00  | 1.00   |
| Levene     | 0.15 | 0.37 | 0.48 | 0.51 | 0.65  | 0.89  | 1.00  | 1.00  | 1.00  | 1.00   |

As can be seen from table 3, when the variance changes by 50%, the parametric criteria of Hartley, Bartlett, Cochran and Leuven also demonstrate higher power than non-parametric ones. Among the non-parametric criteria, the best indicators of power are the Ansari-Bradley and Klotz criteria. In the case when we sample contains more than 200 observations, all other criteria also show good power indicators.

Table 3. Power estimation criteria for correct hypothesis $H_3$ at various N for the normal distribution law

| Criterion  | N=20 | N=40 | N=60 | N=80 | N=100 | N=200 | N=300 | N=400 | N=500 | N=1000 |
|------------|------|------|------|------|-------|-------|-------|-------|-------|--------|
| ANOVA      | 0.05 | 0.13 | 0.21 | 0.33 | 0.43  | 0.67  | 0.86  | 0.95  | 0.98  | 1.00   |
| Ansari-Bradley | 0.17 | 0.35 | 0.48 | 0.53 | 0.62  | 0.94  | 1.00  | 1.00  | 1.00  | 1.00   |
| Klotz      | 0.06 | 0.21 | 0.35 | 0.43 | 0.55  | 0.94  | 0.99  | 1.00  | 1.00  | 1.00   |
| Mood       | 0.14 | 0.29 | 0.43 | 0.48 | 0.60  | 0.97  | 1.00  | 1.00  | 1.00  | 1.00   |
| Siegel-Tukey | 0.17 | 0.33 | 0.44 | 0.48 | 0.59  | 0.95  | 1.00  | 1.00  | 1.00  | 1.00   |
| Bartlett   | 0.21 | 0.39 | 0.55 | 0.63 | 0.70  | 0.94  | 0.99  | 1.00  | 1.00  | 1.00   |
| Cochran    | 0.21 | 0.40 | 0.54 | 0.63 | 0.70  | 0.94  | 0.99  | 1.00  | 1.00  | 1.00   |
| Hartley    | 0.21 | 0.39 | 0.55 | 0.64 | 0.70  | 0.94  | 0.99  | 1.00  | 1.00  | 1.00   |
| Levene     | 0.21 | 0.43 | 0.55 | 0.71 | 0.77  | 0.98  | 1.00  | 1.00  | 1.00  | 1.00   |

With laws other than normal, the power of the Klotz criterion is worse, but the Siegel-Tukey criterion demonstrates good performance. For example, we present the calculation results for the hypothesis $H$ and the logistic distribution (see Table 4).
Table 4. Power estimation criteria for correct hypothesis \( H_3 \) at various N for the logistic distribution law

| Criterion            | N=20 | N=40 | N=60 | N=80 | N=100 | N=200 | N=300 | N=400 | N=500 | N=1000 |
|----------------------|------|------|------|------|-------|-------|-------|-------|-------|--------|
| ANOVA                | 0.01 | 0.05 | 0.1  | 0.19 | 0.29  | 0.62  | 0.88  | 0.96  | 0.99  | 1.00   |
| Ansari-Bradley       | 0.08 | 0.21 | 0.35 | 0.41 | 0.53  | 0.99  | 1.00  | 1.00  | 1.00  | 1.00   |
| Klotz                | 0.01 | 0.07 | 0.19 | 0.28 | 0.43  | 0.93  | 0.99  | 1.00  | 1.00  | 1.00   |
| Mood                 | 0.05 | 0.18 | 0.33 | 0.41 | 0.56  | 0.98  | 0.99  | 1.00  | 1.00  | 1.00   |
| Siegel-Tukey         | 0.06 | 0.17 | 0.30 | 0.37 | 0.49  | 0.94  | 1.00  | 1.00  | 1.00  | 1.00   |
| Bartlett             | 0.15 | 0.29 | 0.44 | 0.61 | 0.64  | 0.97  | 1.00  | 1.00  | 1.00  | 1.00   |
| Cochran              | 0.15 | 0.29 | 0.48 | 0.60 | 0.64  | 0.97  | 1.00  | 1.00  | 1.00  | 1.00   |
| Hartley              | 0.13 | 0.29 | 0.48 | 0.61 | 0.64  | 0.97  | 1.00  | 1.00  | 1.00  | 1.00   |
| Levene               | 0.11 | 0.32 | 0.53 | 0.64 | 0.61  | 0.99  | 1.00  | 1.00  | 1.00  | 1.00   |

4. Conclusion
The results of the study show that the best performance indicators demonstrate the parametric criteria of Bartlett, Cochran, Hartley and Levene. Among non-parametric criteria with small sample sizes (less than 100 observations), the Ansari-Bradley criterion should be used. Among non-parametric criteria with sample sizes of more than 200 observations, the best power indicators for the Ansari-Bradley and Mood criteria, and when the variance changes by 30% or more, the Klotz criterion can be added to them, and if the variance changes by 50%, all the considered criteria show good performance power. It is also worth noting that the power of the criteria is higher when the samples belong to the normal distribution law.

5. Acknowledgments
This research has been supported by the Ministry of Education and Science of the Russian Federation as part of the state task (project No 2.7996.2017/8.9).

References
[1] Timofeev V.S., Faddeenkov A.V., Shchekoldin V.Y. 2015 Econometrics (Novosibirsk: NSTU).
[2] Bartlett M.S. 1937 Properties of sufficiency of statistical tests Proc. Roy. Soc., Series A, 31 268-282.
[3] Cochran W.G. 1941 The distribution of the largest of a set of estimated variances as a fraction of their total Ann. of eugenics 11 47-52.
[4] Ansari A. R., Bradley R. A. 1960 Rank-tests for dispersions AMS 31 1174-1189.
[5] Siegel S., Tukey J. W. 1960 A nonparametric sum of ranks procedure for relative spread in unpaired samples JASA 55 429-445.
[6] Mood A. 1954 On the asymptotic efficiency of certain nonparametric tests AMS 25 514-522.
[7] Sandvik L., Olsson B. 1982 A nearly distribution-free test for comparing dispersion in paired samples Biometrika 69 484-485.
[8] Hartley H.O. The maximum 1950 F-ratio as a short-cut test of heterogeneity of variance Biometrika 37 308-312.
[9] Levene H. 1960 Robust tests for equality of variances. In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (Stanford University Press) 278-292.
[10] Brown M.B., Forsythe A.B. Robust Tests for Equality of Variances 1974 J. Amer. Statist. Assoc 69 364-367.
[11] Klotz J. 1962 Nonparametric tests for scale AMS 33 498-512.