Study of small sample based on Bootstrap variance homogeneity test

Yongjing Xiang [a], Zhongming Yu

Data science and information engineering institute Guizhou Minzu University, Guizhou 550025, China.

[a]2466300950@qq.com

Abstract

Variance analysis and regression analysis are widely used in mathematical statistics, when the data is analyzed, the error term is always required to satisfy the Gauss-Markov hypothesis. In which the homogeneity of variance test, whether in parameters, semi-parametric or non-parametric model in many statisticians have made their corresponding analysis, but when the sample size is very small, the test results are not very accurate, for this phenomenon we introduced the Bootstrap method, empirical analysis, in general, especially in the case of small sample testing Bootstrap method is superior to other general methods.

Keywords

Test of Homogeneity of Variance, small sample, Bootstrap.

1. Introduction

In the management, pharmaceutical and economic areas of the model analysis, when analyzing data often used two samples t test, significance test, analysis of variance and regression analysis. But the model must satisfy the hypothesis: The variance of the observed values is homogeneous. If it does not meet the conditions of its no statistical significance, will also cause some adverse consequences. So the study of homogeneity of variance is very meaningful and necessary.

At present, whether the parameters, nonparametric and semi-parametric model statisticians have made some research on it. For example, Bao Heping et al.(2006) for some problems with the Bartlett test, two more practical test methods are obtained under the same sample size; Sun Gao et al.(1999) for the problem of variance homogeneity test, some other expressions of test statistic are discussed to extend its application scope. Pan Xiaoping et al.(2002) mainly about the Levene test some of the introduction. They solved some of the problems that existed in the variance test and enriched the content of the field. On the common method of variance homogeneity test, there are two samples of the Mood test, Moses test and multi-sample Levene test. But in some studies the data is difficult to collect, and sometimes can only be collected several numbers, that is, the sample size is very small this situation, did not see the relevant research. If you use those classic test methods, then the calculation of its statistics will be very inaccurate, resulting in the results are not accurate, and finally can not make the right judgments, that use the above method to test may not be identified. In this paper, the Bootstrap method is used to test the variance homogeneity of the sample data, for the shortcoming of the small sample size, the use of Bootstrap method to solve, for different situations, tested on simulated data and actual data , the results are compared with the results of the Mood test, the Moses test and the Levene test, and the relevant conclusions are drawn, which is helpful to judge the overall distribution of the samples.

2. Classical variance test method

In general, under different conditions, there will be different variance homogeneity test method, if the two samples are from the normal distribution, then use the F test directly, If the two samples are not from the normal distribution, then use the Mood variance test, when not considering the position parameters, the use of Moses variance test. For multi-sample, the common Bartlett multi-sample variance homogeneity test is mainly used for the data of normal distribution, but the Levene test can be used for data of both normal and non-normal distribution. In general, we do not know the distribution of data, so this article does not discuss the method of normal distribution.

2.1 Mood variance test

The key idea of the Mood variance test is to replace the original data with rank, when the null hypothesis H_0 is established, the distribution of the two samples is the same, the mixed samples are independent and identical distribution (IID), then the variance of the rank of one of the samples in the mixed sample should be equal to the variance of the mixed sample.

Suppose:

1. The position parameters of the two data samples are equal, without loss of generality, assuming that the position parameter is zero.

2. The variance are σ_1^2 and σ_2^2 .

Test questions and principles:

Suppose that the two samples are $X_1, X_2, \cdots, X_m \sim F(\frac{x}{\sigma_1})$ and $Y_1, Y_2, \cdots, Y_n \sim F(\frac{x}{\sigma_2})$

Hypothesis test:

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

F is continuous, $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ are independent, Let R_i be the rank of Xi in the mixed sample,

When H_0 is established, $X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n$ is IID,

$$E(R_i) = \sum_{i=1}^{m+n} \frac{i}{m+n} = \frac{m+n+1}{2} \tag{1}$$

When H_0 is established, For X samples, Consider the Mood rank statistics:

$$M = \sum_{i=1}^{m} (R_i - \frac{m+n+1}{2})^2 \tag{2}$$

If the variance of X is too large, then the value of M should also be too large (Wang Xing,2013), At the moment, we think can to reject the null hypothesis.

2.2 Moses variance test

Principle and calculation process:

It is not necessary to assume that the mean is equal, Set $x_1, \cdots, x_m$ are random sample of the first distribution and the variance is σ_1^2 . $y_1, \cdots, y_n$ are random sample of the second distribution and the variance is σ_2^2 . Hypothesis test :

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

The Moses variance test statistic T is calculated as follows

1. Will $x_1, \cdots, x_m$ is randomly divided into m_1 group, Each group have k observations, denoted as $A_1, A_2, \cdots, A_{m_1}$;will $y_1, \cdots, y_n$ is randomly divided into m_2 group, Each group have k observations, denoted as $B_1, B_2, \cdots, B_{m_{21}}$;

2. obtaining the square sum of the sample deviations in each group, use the sample deviation instead of the variance , carry out the next calculation:

$$SSA_r = \sum_{x_i \in A_r} (x_i - \bar{x})^2, \quad r = 1,2,\cdots,m_1$$
$$SSB_s = \sum_{y_i \in B_s} (y_i - \bar{y})^2, \quad s = 1,2,\cdots,m_2 \tag{3}$$

3. Mix SSA_r and SSB_s, And solve the corresponding rank in the mixed sample
4. Calculate the rank sum of sample SSA_r, denoted as S, Construct the Moses statistic

$$T_M = \frac{S - m_1(m_1+1)}{2} \tag{4}$$

If the value of T_M is large, consider rejecting the null hypothesis. The actual test can be checked Mann-Whitney of W_α table(Wang Xing,2013). The shortcomings of Moses variance test that SSA_r and SSB_s only have less sample size , That is, the sample size of test much less, and also need to look up the table.

2.3 Levene test

Inspection principle and method:

1. Hypothesis test

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$

H_1: $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$, That is, the variance of each group is no equal

2. Calculate the test statistic L value

$$L = \frac{(N-k)\sum_{i=1}^{k} N_i(\overline{Z_{i\cdot}} - \overline{Z_{\cdot\cdot}})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\overline{Z_{ij}} - \overline{Z_{i\cdot}})^2} \tag{5}$$

Here the test statistic L to obey F distribution of the degrees of freedom $\upsilon 1 = k-1$ and $\upsilon 2 = N-k$, k is the number of sample groups, N_i is the sample size of the i-th sample, N is sample size of the total sample, $\overline{Z_{i\cdot}}$ is the mean of the i-th sample, $\overline{Z_{\cdot\cdot}}$ is the total mean of all data.

Zij can be defined as one of the following three forms:

$$Z_{ij} = \begin{cases} |x_{ij} - \bar{x}_{i\cdot}| \\ |x_{ij} - \tilde{x}_{i\cdot}| \\ |x_{ij} - \bar{x}'_{i\cdot}| \end{cases} \tag{6}$$

Where x_{ij} is the original data, $\bar{x}_{i\cdot}$ is the arithmetic mean of the i-th group sample of in the original data, $\tilde{x}_{i\cdot}$ is the median of the i-th group sample of in the original data, $\bar{x}'_{i\cdot}$ is the 10% adjusted mean of the i-th group sample of in the original data. 10% adjusted mean refers to the remove less than P_5 and greater than P_{95} data, Calculate the arithmetic mean between P_5 and P_{95}(Cheng Cong et al. , 2005). The three kinds of conversion methods make the Levene test have good robustness and control. China's common Bartlett multivariate variance homogeneity test is mainly used for data of the normal distribution, for non-normal data test results are not ideal, But the Levene test can be used for data of normal distribution, and can also be used for data of non-normal distribution or unidentified, the test results are ideal. But for small sample need do some consideration.

3. Bootstrap variance test method

3.1 Bootstrap introduction

Bootstrap is commonly used method of statistical inference, Such as estimates of variance and distribution of statistics. In the statistical field, more influential a re-sampling method is Jackknif.

Mainly for small sample, In the case of without changing the original sample, expand the sample size of the statistics, to overcome the limitations of the original sample generation range, in the case of consistent distribution characteristics, to reduce its similarity with the original sample, That is mainly self-exploration about the true distribution.

Bootstrap variance estimation is in the case of use the large number theorem, sampling with replication to get a new sample, use the new sample instead of the original sample, and calculation obtained the approximate variance or other need approximation value.

3.2 Variance homogeneity test of two samples

Set $x_1,\cdots,x_{n1}$ are random samples of the first distribution, and overall variance is σ_1^2 . Set $y_1,\cdots,y_{n2}$ are random samples of the second distribution, and overall variance is σ_2^2 .

hypothetical test:

H0: The variance of the two distribution is equal. that is $\sigma_1^2=\sigma_2^2$

H1: The variance of the two distribution is no equal. that is $\sigma_1^2\neq\sigma_2^2$

Statistical analysis:

When the number of B is large, the large number theorem shows that the sample obeys the normal distribution, so that the data can be effectively analyzed with z statistic. Inspection steps:

1. Samp、ling with replacement about the two samples B times and calculate their variance;
2. Calculate the mean of the variance from (1), and the total variance;
3. Based on the mean and variance obtained in (2), calculated

$$Z=\frac{|\mathrm{Mx}-\mathrm{My}|}{\sqrt{Varxy}} \tag{7}$$

Where Mx and My are the mean values calculated by (2), *Varxy* is the total variance;

4. According to (3) calculated z value check the normal distribution table or directly calculate its p value.

3.3 Variance homogeneity test of multi - sample

Taking three samples as an example, set sample data is $x_{11},\cdots,x_{n1}$, $x_{12},\cdots,x_{n2}$ and $x_{13},\cdots,x_{n3}$,and the overall variance is σ_1^2 ,σ_2^2 and σ_3^2 .

hypothetical test:

H0: $\sigma_1^2=\sigma_2^2=\sigma_3^2$

H1: $\sigma_1^2\neq\sigma_2^2\neq\sigma_3^2$, That is, the variance of each group is no equal

Statistical analysis:

When the null hypothesis *H0* is established , inter-group variance and intra-group variance of mixed samples will not have very difference.

Inspection steps:

1. Mix the data together and sampling without replacement have the same sample size with original data, calculate the statistics W, repeat B times:

$$W=\frac{MS_{inter-group}}{MS_{intra-group}}=\frac{(N-k)\times\sum_{i=1}^{k}N_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2}{(k-1)\times\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\bar{x}_{ij}-\bar{x}_{i\cdot})^2} \tag{8}$$

Where $MS_{\text{inter-group}}$ is inter-group variance of mixed samples, $MS_{\text{intra-group}}$ is intra-group variance of mixed samples, N is sample size of the total sample, N_i is the sample size of the i-th sample, $\bar{x}_{i.}$ is the mean of the i-th sample, $\bar{x}_{..}$ is the total mean of all data. $\bar{x}_{ij} = |x_{ij} - \bar{x}_{i.}|$ x_{ij} is the original data.

2. When B is large enough, the distribution of the statistics W can be obtained, it is possible to further derive the probability of rejecting the null hypothesis.

4. Empirical analysis

Through the above theoretical analysis, it is not clear enough to understand the difference between these methods. Therefore, in order to deepen the understanding of different methods, in the following we use the data of two samples and multiple samples for experimental comparison.

4.1 Variance homogeneity test of two samples

1. The simulation data are tested and compared. Assume that two samples are from a normal distribution with a mean of zero. The variance of the first column test data is 1 and 1.5, each sample size is 15.that is $x_i \sim N(0,1), y_j \sim N(0,1.5)$ i=1,2,…,15,j=1,2,…,15; the variance of the second column test data is 1 and 1.6, samples size is 10 and 12,that is $x_i \sim N(0,1), y_j \sim N(0,1.6)$ i=1,2,…,10,j=1,2,…,12; the variance of the third column test data is 1 and 1.7, samples size is 9, that is $x_i \sim N(0,1), y_j \sim N(0,1.7)$ i=1,2,…,9, j=1,2,…,9. Use R programming, For above these types of data, the two samples Bootstrap variance test method presented in this paper are used to test and compared with the results of Mood variance test and Moses variance test. The results are shown in Table 1.

Table 1 Comparison of variance homogeneity test of two normal distribution samples

Testing method	p-value	p-value	p-value
Mood variance test	0.1565	0.1077	0.2513
Moses variance test			
k=2	0.3829	0.08225	0.1143
k=3	0.8413	0.2286	0.4
k=4	0.4	0.2	0.3333
Bootstrap variance test			
B=100	0.04425067	0.04242941	0.06892186
B=1000	0.04853543	0.04976903	0.06699058
B=10000	0.04652941	0.04995908	0.06936944

As can be seen from the above table, when $\alpha = 0.05$ or 0.10 , For the two groups of data with significant differences in variance, using the Mood variance test show that results have not difference, Moses variance test even if regardless of k value, the results also have not difference. And when k takes different value, it corresponds to the p-value changes are relatively large, that is, Moses variance test is very instability, And there is no uniform standard for k to be selected, so it may lead to different people obtain the opposite conclusion. But if use the Bootstrap variance test can detect the difference between the data, And look up that the p-value does not intense change as B times increases, indicating that it is relatively stable and also can true to explanation character of the sample.

2. The actual data are tested and compared. Examined that sample of in the nonparametric statistics book written by Wang Xing, The first sample is about the concentration of uric acid in the blood of the stroke patient and healthy adult. The data is shown in Table 2:

Table 2. The concentration of uric acid in the blood of patients and normal people

Patient(x)	8.2	10.7	7.5	14.6	6.3	9.2	11.9	5.6	12.8	5.2	4.9	13.5
Normal people (y)	4.7	6.3	5.2	6.8	5.6	4.2	6.0	7.4	8.1	6.5		

is shown in Table 3:

Table 3. A and B soil organic matter content

A	8.8 8.2 5.6 4.9 8.9 4.2 3.6 7.1 5.5 8.6 6.3 3.9
B	13.0 14.5 16.5 22.8 20.7 19.6 18.4 21.3 24.2 19.6 11.7 18.9 14.6 19.8 14.5

For the above two types of data, using the two-sample Bootstrap method given in this paper, whether there is significant difference in test variance, that is, whether there is homogeneity of variance. And the results were compared with the Mood variance test and the Moses variance test. The results are shown in Table 4:

Table 4. Comparison of homogeneity test results of variance in two samples

Testing method	p-value	p-value
Mood variance test	0.004138	0.1211
Moses variance test		
k=2	0.05468	0.1014
k=3	0.05714	0.06349
k=4	0.2	0.2
Bootstrap variance test		
B=100	0.0310546	0.03906919
B=1000	0.03266432	0.0410023
B=10000	0.03275313	0.0404224

From the above results can know, the first group is about the concentration of uric acid in the blood of the stroke patient and healthy adult, under the significance level of 0.05, Mood variance test p-value = 0.004138 <0.05, The null hypothesis is rejected, It is considered that there is significant difference in the concentration of uric acid in the blood of the stroke patient and healthy adult. Under the significance level of 0.1, when k = 2 or 3 the Moses variance test results have significant difference, but k = 4 can not test the difference between them, hence can not to make the exact conclusion. But if use the Bootstrap variance test to test it, under the significance level of 0.05, the p-value is less than 0.05 and rejects the null hypothesis, It is considered that there is significant difference in the concentration of uric acid in the blood of the stroke patient and healthy adult, and the fluctuation of the p-value is not very large with the change of times B, that is, it has a very good robustness. Similarly, for the second group the A and B two soil organic matter content sampling results, the Mood variance test, under the significance level of 0.1, p-value=0.1211>0.1 It means that there is not enough information to reject the null hypothesis, that is, to accept the null hypothesis, Believes that soil organic matter content A and B have not significant difference, the Moses variance test ,under the significance level of 0.1, only when k = 3, A and B soil organic matter content have significant difference, and when k is other values, we can not determine the difference between them. Thus the Moses variance test showed instability. However, if we use the Bootstrap variance test to test the data, under the significance level of 0.05, we can test the A and B soil organic matter content is significantly different. And with the change of times B, the fluctuation of the p-value is not great, That is, it has a very good robustness and can be a good explanation character of the sample.

4.2 Variance homogeneity test of multi - sample

1. The simulation data are tested and compared. Suppose that the three sets of sample data are from the normal distribution, the mean is 0, the variance of the first column data is 0.1,0.05 and 0.12, the sample size is 5. the variance of the second column data is 0.1,0.25 and 0.35, and the sample size is 10. the variance of the third column data is 0.1, 0.05 and 0.12, and the sample size is 5, 10, and 12. Using the multi-sample Bootstrap variance test method presented in this paper to test the data, Using R

programming to get the results, and then compare the results with the results of the Levene test. The results are shown in Table 5:

Table 5. Comparison of Variance Homogeneity Test Results of multi-sample Normal Distribution

Testing method	p-value	p-value	p-value
Levene test	0.4878	0.0639	0.1927
Multi-sample Bootstrap variance test			
B=100	0.01	0.02	0.01
B=1000	0.022	0.023	0.007
B=10000	0.0218	0.0232	0.0067

The results of the above table show that when the significance level is 0.05, for the samples with significant difference in variance, using the Levene test method can not be recognition, but using the Bootstrap test method can test the difference between them, and with change of times B the p-value is basically stable, you can make a better judgment on the sample.

2. The actual data are tested and compared. Test the actual data is Example 7.4.3 come from Tang Yincai R language and statistical analysis book, mainly for three different athletes A, B, C under the same conditions for the shooting competition, according to hit the number of rings to determine the stability of the athletes. The data are shown in Table 6:

Table 6. The number of rings of the athlete shooting

A	B	C
8	8	10
7	7	10
9	9	9
10	6	6
9	8	8
6	9	3
5	10	5
8	7	6
10	8	7
5	9	4

The other is different treatment for the A, B, C three groups of animals, detection of the index data in the blood, analyze the variance homogeneity of the three sets of data, The data are shown in Table 7:

Table 7. A index data in the blood of three groups of animals

A	B	C
6	22	40
10	28	48
10	30	54
12	36	58
15	46	78
19	48	88

The above two sets of data, using the multi-sample Bootstrap method presented in this paper to test, Test variance whether have significant difference, and compare the results with the classic Levene test results, The results are shown in Table 8:

Table 8. Comparison of variance homogeneity test results of multi-sample data

Testing method	p-value	p-value
Levene test	0.08259	0.1014
Multi-sample Bootstrap variance test		
B=100	0.02	0.11
B=1000	0.021	0.059
B=10000	0.0267	0.0646

The results of the above table shows, The first set of data under significant level of 0.05, Levene test p-value =0.08259>0.05, therefore there is not enough information to reject the null hypothesis, it is believed that the stability of the three athletes are the same. But for the Bootstrap test regardless of the value of B , will to reject the null hypothesis, that is, the stability of the three athletes are different. Similarly, for the results of the second set of data, under the significance level of 0.1, the Levene test p-value=0.1014 >0.1 indicates that there is insufficient information to reject the null hypothesis, there is no significant difference of the three groups of animals, When the B value is small, the Bootstrap test value and Levene test value is very closer, can not reject the null hypothesis, But when the value of B is large, can reject the null hypothesis, there is significant differences of the three groups of animals, So when the sample size is very small, We should try to take a large B value, in order to arrive at effective results. But in general when the sample size is very small, Bootstrap test methods are better than other test methods.

5. Conclusion

Through the above experiment, whether it is using the computer simulation data or the example of book, two or many samples data, the results show that the Bootstrap variance test results are superior to the classical variance test method and the Bootstrap method is more stable when the sample size is small.

Acknowledgements

Fund project: Guizhou Minzu University for Research Fund funded projects (NO: 16yjsxm046).

References

[1] Cheng Zhong, Fan Hua. Levene variance homogeneity test Chinese health statistics[J] , 2005, 22 (6):408-408.
[2] Dai Shaowu, Gao Huaming. Research on Small Sample Data Analysis Method Based on Self - help Method[J]. Journal of Naval Aeronautical Engineering Institute, 2009, 24(1):27-30.
[3] Pan Xiaoping, Ni Xongzan. A Robust Homogeneity Test Method[J]. Modern Preventive Medicine, 2002, 29(6):774-776.
[4] Tang Yincai. R Language and Statistical Analysis [M], Higher Education Press, 2009.
[5] Wang Xing, Zhe Tingjin. Non-parametric statistics[M], Tsinghua University Press, 2014.
[6] HyoIl Park, Seung Man Hong. A permutation test for multivariate data with grouped components[J].Journal of Applied Statistics, 2010,37(5):767-778.