Economic Modelling 29 (2012) 610–613

Contents lists available at SciVerse ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

# Testing population variance in case of one sample and the difference of variances in case of two samples: Example of wage and pension data sets in Serbia

Vesna Cojbasic Rajic, Jelena Kocovic, Dragan Loncar *, Tatjana Rakonjac Antic

*University of Belgrade, Faculty of Economics, Serbia*

ARTICLE INFO

*Article history:*
Accepted 1 January 2012

*Keywords:*
Bootstrap
Edgeworth expansion
Testing hypotheses
Variance
Data set

ABSTRACT

Testing one population variance and the difference in variances of two populations based on the ordinary $t$-statistics combined with the bootstrap method are suggested in this article. Suggested techniques are combined with Hall's transformation approach. Application of presented methods in domain of real economic data set is described and analyzed. We compare the outputs of suggested methods and traditional methods for considered data set. The results show that these introduced methods have small advantages in comparisons with traditional methods especially for small samples.



## 1. Introduction

The statistical method used in testing population variance is based upon the $\chi^2$-statistic. If we are interested in testing the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against some alternative $H_1$ existing method is based on the test statistic which is of the form (Pearson, 1900):

$$\chi^2 = (n-1)S^2/\sigma_0^2, \tag{1}$$

where $\sigma_0^2$ is the hypothesized population variance and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a sample variance. Statistic (1) has $\chi^2$ distribution with $n-1$ degrees of freedom, where $n$ is sample size.

If we want to test that two populations have the same variance, existing method is based on the $F$ statistic which is of the form (Snedecor, 1934):

$$F = S_{X_1}^2 / S_{X_2}^2, \tag{2}$$

where $S_{X_1}^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}\left(X_{1j} - \bar{X}_1\right)^2$ and $S_{X_2}^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}\left(X_{2j} - \bar{X}_2\right)^2$ are corresponding sample variances and $n_1$ and $n_2$ are sample sizes. Statistic (2) has $F$ distribution with $n_1-1$ and $n_2-1$ degrees of freedom. These single-sample chi-square test and $F$-test are known to be highly sensitive to non-normality (for example, see: Box, 1953; Lehman, 1986; and Markowski and Markowski, 1990).

In this paper we examine new test statistics for the testing unknown population variance and the difference in variances of two populations, based on the ordinary $t$-statistics and resampling methods (for details about resampling methods see for example, Efron and Tibshirani, 1993 and Davison and Hinkley, 1997). We determine the $p$-value of all suggested tests and of traditional tests.

## 2. Testing population variance in case of one sample

Let $X_1, \ldots, X_n$ be i.i.d. random variables from normal distribution with mean $\mu$ and variance $\sigma^2$. Suppose that we want to test a population variance, i.e. we want to test $H_0 : \sigma^2 = \sigma_0^2$ against some alternative $H_1$. Because of the fact that the distribution of the standardized variable $Z = \frac{\frac{(n-1)S^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} = \frac{S^2 - \sigma^2}{\sqrt{\mathrm{var}(S^2)}}$ converges to standardize normal distribution as $n$ increases to infinity (see Cojbasic and Tomovic, 2007) we shall consider the statistic:

$$T = \frac{S^2 - \sigma_0^2}{\sqrt{\widehat{\mathrm{var}}\left(S^2\right)}}, \tag{3}$$

where $\widehat{\mathrm{var}}\left(S^2\right)$ is a consistent estimator of the variance of $S^2$.

We can consider bootstrap test based on statistic (3). We generate $B$ bootstrap samples, and in each bootstrap sample we compute the value of statistic:

$$T^{*b} = \frac{S^{2^*} - S^2}{\sqrt{\widehat{\mathrm{var}}\left(S^{2^*}\right)}}, \; b = 1, \ldots, B, \tag{4}$$

where $S^{2^*}$ is a bootstrap replication of the statistic $S^2$.

* Corresponding author at: University of Belgrade, Faculty of Economics, Kamenicka 6, 11000 Belgrade, Serbia. Tel.: +381 113021075, +381 652405020(mobile).
*E-mail address:* loncar@ekof.bg.ac.rs (D. Loncar).

0264-9993/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2012.01.002

If Cramer's condition holds (see Hall, 1992b) and if $E(X_j^6)<\infty$, $T$ admits a first-order Edgeworth expansion:

$$P(T\leq x)=\Phi(x)+\frac{1}{\sqrt{n}}q(x)\phi(x)+O\left(n^{-1}\right), \tag{5}$$

where $q(x)=\frac{M'_3}{6}\left(2x^2+1\right)$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density function, respectively. $M'_3=E\left((1/n)\sum_{i=1}^{n}X'^3_i\right)$, where random variables $X'_i$, $i=1,2,\ldots,n$, are defined in the following way:

$$X'_i=\frac{(X_i-\overline{X})^2-((n-1)/n)\sigma_0^2}{\sqrt{V_1}}, \text{for } i=1,2,\ldots,n, \tag{6}$$

and $V_1=E\left((X_i-\overline{X})^2-((n-1)/n)\sigma_0{}^2\right)^2$ (for the proof see Cojbasic and Tomovic, 2007).

In order to eliminate the effect of the skewness, we can use expansion (5) and got this transformed test statistic:

$$g_1(T)=T+\frac{1}{3}\frac{1}{\sqrt{n}}\widehat{M}'_3T^2+\frac{1}{27}\frac{1}{n}\widehat{M}'^2_3T^3+\frac{1}{6}\frac{1}{\sqrt{n}}\widehat{M}'_3, \tag{7}$$

where $\widehat{M'_3}$ is a moment estimator of $M'_3$ (for details see Cojbasic and Loncar, 2011). Idea for this transformation follows from Hall's transformation on standardized sample mean statistic (see Hall, 1992a).

## 3. Testing population variance in case of two samples

Let $X_{11},X_{12},\ldots,X_{1n_1}$ and $X_{21},X_{22},\ldots,X_{2n_2}$ be i.i.d. from two normal distributions with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively. Suppose that we want to test $H_0:\sigma_1^2=\sigma_2^2$ (against some alternative $H_1$) i.e. that two populations have the same variance. Because of the fact that the distributions of standardized variables $Z_1=\frac{S^2_{X_1}-\sigma_1^2}{\sqrt{\mathrm{var}\left(S^2_{X_1}\right)}}$ and $Z_2=\frac{S^2_{X_2}-\sigma_2^2}{\sqrt{\mathrm{var}\left(S^2_{X_2}\right)}}$, converge to standardize normal distribution for enough large $n_1$ and $n_2$ (see Cojbasic and Tomovic, 2007) we consider the statistic:

$$T=\frac{S^2_{X_1}-S^2_{X_2}-\left(\sigma_1^2-\sigma_2^2\right)}{\sqrt{\widehat{\mathrm{var}}\left(S^2_{X_1}\right)+\widehat{\mathrm{var}}\left(S^2_{X_2}\right)}}. \tag{8}$$

We can consider bootstrap test based on statistic (8). We generate $B$ bootstrap samples, and in each bootstrap sample we compute the value of statistic:

$$T^{*b}=\frac{S^{2^*}_{X_1}-S^{2^*}_{X_2}}{\sqrt{\widehat{\mathrm{var}}\left(S^{2^*}_{X_1}\right)+\widehat{\mathrm{var}}\left(S^{2^*}_{X_2}\right)}}, b=1,\ldots,B, \tag{9}$$

where $S^{2^*}_{Xi}$ $(i=1,2)$ is a bootstrap replication of the statistic $S_{X_i}{}^2$.

If Cramer's condition holds (see Hall, 1992b) and if $E(X_{ij}^6)<\infty$, $T$ admits a first-order Edgeworth expansion:

$$P(T\leq x)=\Phi(x)+\frac{1}{\sqrt{N}}q(x)\phi(x)+O\left(\frac{1}{N^{-\min(1,r+1/2)}}\right), \tag{10}$$

where $q(x)=\frac{A}{6}\left(2x^2+1\right)$ and $r\geq 0$. $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density function, respectively. $A=\frac{\frac{V_1^{3/2}M'_{31}}{\lambda^2}-\frac{V_2^{3/2}M'_{32}}{(1-\lambda)^2}}{\left(\frac{V_1}{\lambda}+\frac{V_2}{1-\lambda}\right)^{3/2}}$, where $M'_{31}=E\left(\frac{1}{n_1}\sum_{j=1}^{n_1}X'^3_{1j}\right)$ and $M'_{32}=E\left(\frac{1}{n_2}\sum_{j=1}^{n_2}X'^3_{2j}\right)$. Random variables $X'_{ij}$ $(i=1,2$ and $j=1,2,\ldots,n_i)$ are defined in the following way:

$$X'_{ij}=\frac{\left(X_{ij}-\overline{X}_i\right)^2-\frac{n_i-1}{n_i}\sigma_i^2}{\sqrt{V_i}}, \tag{11}$$

where $V_i=E\left(\left(X_{ij}-\overline{X}_i\right)^2-\frac{n_i-1}{n_i}\sigma_i^2\right)^2$ are constants (for the proof see Cojbasic and Tomovic, 2007).

In order to eliminate the effect of the skewness, we can use expansion (10) and got this transformed test statistic:

$$g_2(T)=T+\frac{1}{3}\frac{1}{\sqrt{n}}\hat{A}T^2+\frac{1}{27}\frac{1}{n}\hat{A}^2T^3+\frac{1}{6}\frac{1}{\sqrt{n}}\hat{A}^3, \tag{12}$$

where $\hat{A}$ is a moment estimator of $A$ (for details see Cojbasic and Tomovic, 2007; Hall, 1992a).

## 4. A simulation study

In this section we will deal with the sensitivity of the tests, which are introduced in Sections 2 and 3. We will test the variance in case of one sample and two samples. Furthermore, we will consider the application of these tests on the Weibull and Exponential distribution. The same approach can be applied to any other skewed distributions (e.g. Lognormal, Gamma and others). Among parametric methods we will use $\chi^2$ and $F$ test, while among nonparametric tests we will use bootstrap test (where we use statistics (4) and (9)) and bootstrap test with Hall's transformation approach (where we use statistics (7) and (12)).

In the case of one sample we have considered various sample sizes (10, 20, 50, 100). From each sample we generated 1000 bootstrap samples. Resulting $p$-values are shown in Table 1. All calculations are performed in programming language Fortran.

Based on results from the Table 1, we see that the bootstrap test with Hall's transformation approach shows advantages over the $\chi^2$ test for small samples. For instance, if the sample size is 20 and the distribution is exponential, the bootstrap test with Hall's transformation approach rejects the null hypothesis (which is not true) at the level of significance 0.05, while other tests do not reject the null hypothesis. In large samples, the use of any test gives similar conclusions.

In the case of two samples, we considered random samples of the following size (10, 10), (10, 20), (10, 100), (20, 20), (50, 50), and (100, 100). We have selected some of the combinations of pairs of sample sizes and subsequently generated 1000 bootstrap samples, out of already generated samples. We considered cases when both samples are small, when one sample is small and the other is large,

**Table 1**
$p$-value when testing hypothesis $H_0:\sigma^2=\sigma_0^2$ against $H_1:\sigma^2<\sigma_0^2$ (for example, let $\sigma_0^2$ be two times the population variance).

| Distributions | $n$ | Bootstrap test | Bootstrap with Hall's approach | $\chi^2$test |
|---|---|---|---|---|
| Exponential with mean 1 | 10 | 0.260 | 0.220 | 0.378 |
| | 20 | 0.065 | 0.045 | 0.082 |
| | 50 | 0.050 | 0.035 | 0.020 |
| | 100 | 0.010 | 0.005 | 0.001 |
| Weibull with shape parameter 2 | 10 | 0.335 | 0.265 | 0.436 |
| | 20 | 0.150 | 0.130 | 0.407 |
| | 50 | 0.010 | 0.005 | 0.001 |
| | 100 | 0.002 | 0.000 | 0.000 |