Computational Statistics and Data Analysis 54 (2010) 2306–2316

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A bootstrap test for equality of variances

Dexter O. Cahoy *

Program of Mathematics and Statistics, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, United States

ARTICLE INFO

Article history:
Received 17 November 2009
Received in revised form 16 April 2010
Accepted 16 April 2010
Available online 27 April 2010

Keywords:
Homogeneity of variances
ANOVA
Dirichlet distribution
Bootstrap

ABSTRACT

We introduce a bootstrap procedure to test the hypothesis H_o that $K+1$ variances are homogeneous. The procedure uses a variance-based statistic, and is derived from a normal-theory test for equality of variances. The test equivalently expressed the hypothesis as $H_o : \boldsymbol{\eta} = (\eta_1, \ldots, \eta_{K+1})^T = \mathbf{0}$, where η_i's are log contrasts of the population variances. A box-type acceptance region is constructed to test the hypothesis H_o. Simulation results indicated that our method is generally superior to the Shoemaker and Levene tests, and the bootstrapped version of the Levene test in controlling the Type I and Type II errors.

1. Introduction

Testing the homogeneity of variances arises in many scientific applications. It is increasingly used now to determine uniformity in quality control, in biology, in agricultural production systems, and even in the development of educational methods (see Boos and Brownie, 2004). It is also a prelude to testing the equality of population means such as the analysis of variance (ANOVA) (see Scheffe, 1959), dose- response modeling or discriminant analysis. The literature for testing equality of variances is huge and we refer the readers to the comprehensive review of Conover et al. (1981).

More recently, procedures for testing equality of variances that are robust to non-normality have been categorized into three major approaches. These strategies are based on the following: (1) Kurtosis adjustment of normal-theory tests (Box and Andersen, 1955; Shoemaker, 2003), (2) Analysis of variance (ANOVA) on scale variables such as the absolute deviations from the median or mean (Levene, 1960; Brown and Forsythe, 1974), and (3) Resampling methods to obtain p-values for a given test statistic (Box and Andersen, 1955; Boos and Brownie, 1989). Descriptions of these methods are summarized in Boos and Brownie (2004).

The main focus of our research is on resampling methods, as they have been shown to improve the Type I and possibly the Type II error rates (see Boos and Brownie, 1989; Lim and Loh, 1996). More specifically, our goal is to propose a variance-based procedure to test the homoscedasticity of variances for a wide variety of distributions. It is also our objective to validate whether resampling methods improve Type I and Type II error rates. An important attribute of our proposed method is its ability to control better the Type I and Type II error rates for small sample (both equal and unequal) sizes. Our test uses a box-type acceptance region rather than a p-value which distinguishes it from other resampling methods. It is solely based on a variance-based statistic without applying any transformation to the observed data such as smoothing, fractional trimming, or replacing original observations by the scale or residual estimates. The variance-based procedure is also shown to be more sensitive to deviations from the null conditions.

Just like Boos and Brownie (1989), we prefer variance-based procedures as they are more appealing to practitioners, easier to interpret, and variances are of interest in many areas. We also hope that with constantly improving state-of-the-art computing machinery, this research will encourage the use of resampling-based tests for equality of variances

* Tel.: +1 318 257 3529; fax: +1 318 257 2182.
E-mail address: dcahoy@latech.edu.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.04.012

by practitioners, and the integration of these procedures into major statistical software packages. The descriptions of the bootstrap and non-bootstrap tests for equality of variances are given in Section 2. Section 3 shows the small-to-moderate sample size performance of the tests. We close the article with a summary and an outline of possible future extensions.

2. Description of tests to be compared

Given $K+1$ samples from the populations $F\{(d_i-\mu_i)/\sigma_i\}, i = 1, \ldots, K+1$ with equal kurtosis, the ith sample $d_{i1}, d_{i2}, \ldots, d_{in_i}$ having size n_i, and $n = n_1 + \cdots + n_{K+1}$, consider a test of the hypothesis

$$H_o : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_{K+1}^2$$

against the alternative hypothesis H_a that at least two of the $K+1$ variances are unequal. Let s_i^2 denote the sample variance based on n_i observations from the ith sample. We now describe the tests that will be compared.

2.1. Levene's test L

Levene (1960) first proposed ANOVA on the scale variables $e_{ij} = |d_{ij} - \overline{d}_i|$, where $\overline{d}_i$ is the mean of the ith sample but Miller (1968) showed that e_{ij} is asymptotically correct for asymmetric populations if the median is used instead of the mean. Brown and Forsythe (1974) formally studied Levene's method where the median was used instead of the mean to center the variables. Boos and Brownie (1989) and Lim and Loh (1996) provide more details on the features of Levene's test.

We consider Levene's test as it is widely used in practice even if it is not a variance- and resampling-based procedure. It is also recommended by Conover et al. (1981). Levene's procedure is a test for equality of means applied to the scale quantities $e_{ij} = |d_{ij} - \tilde{\mu}_i|$, where $\tilde{\mu}_i$ is the median of the ith sample $\{d_{ij}, j = 1, \ldots, n_i\}$. The test statistic is

$$L_{ts} = \frac{\sum_{i=1}^{K+1} n_i(e_{i\cdot} - \bar{e}_{\cdot\cdot})^2/K}{\sum_{i=1}^{K+1}\sum_{j=1}^{n_i}(e_{ij} - \bar{e}_{i\cdot})^2/(n-K+1)},$$

where $e_{i\cdot} = \sum_{j=1}^{n_i} e_{ij}/n_i$ and $\bar{e}_{\cdot\cdot} = \sum_{i=1}^{K+1}\sum_{j=1}^{n_i} e_{ij}/n$. We reject the null hypothesis H_o if L_{ts} exceeds the $100(1-\alpha)$th quantile $F_{K,n-K+1}(\alpha)$ of the F-distribution with K and $n-K+1$ degrees of freedom.

Some variants of Levene's test are proposed by O'Brian (1978), Hines and Hines (2000), and Good (2000). But these modified versions of Levene's test are still inferior in terms of level and power to its bootstrap version (see Liu, 2006), which is discussed in Section 1.

2.2. Shoemaker's test S

We also consider Shoemaker's test S, which not only provides good insights to our procedure but was the test recommended by Shoemaker (2003) after comparing its performance with some kurtosis-adjusted normal-theory tests. The test statistic is

$$S_{ts} = \sum_{i=1}^{K+1}\left(\ln s_i^2 - \overline{\ln s^2}\right)^2 / \widehat{\mathrm{var}}(\ln s_i^2),$$

where $\overline{\ln s^2} = \sum_{i=1}^{K+1} \ln s_i^2/(K+1)$, $\widehat{\mathrm{var}}(\ln s_i^2) = [\widehat{\mu_4'}/\widehat{\sigma}^4 - (h_i-3)/(h_i-1)]/h_i$, h_i is the harmonic mean $(K+1)/\sum_{i=1}^{K+1} 1/n_i$, $\widehat{\mu_4'} = \sum_{i=1}^{K+1}\sum_{j=1}^{n_i}(d_{ij} - \overline{d}_{i\cdot})^4/n$ is the estimator of the fourth moment about the population mean, and $\widehat{\sigma}^2 = \sum_{i=1}^{K+1}(n_i-1)s_i^2/n$. He recommended the estimator of an asymptotically equivalent formula, which is $\widehat{\mathrm{var}}(\ln s_i^2) = [\widehat{\mu_4'}/\widehat{\sigma}^4 - (h_i-3)/h_i]/(h_i-1)$, to improve simulation accuracy. The null hypothesis is rejected when S_{ts} exceeds the $100(1-\alpha)$th percentile of the chi-square distribution with K degrees of freedom.

2.3. Lim and Loh's test BL

Lim and Loh (1996) compared several bootstrap and non-bootstrap tests for heterogeneity of variances. A bootstrap version of Levene's test was recommended because of its superiority in terms of power and Type I error robustness. The procedure used the technique of Boos and Brownie (1989), and is given below.

1. Compute the test statistic from the given data d_{ij}, $i = 1, \ldots, K+1, j = 1, \ldots, n_i$.
2. Initialize $l = 0$.
3. Compute the residuals $e_{ij} = d_{ij} - \widehat{\mu}_i$, $i = 1, \ldots, K+1, j = 1, \ldots, n_i$ where $\widehat{\mu}_i$ is the median of group i.
4. Draw n data points e_{ij}^*'s from the pooled residuals $\overline{R} = \{d_{ij} - \widehat{\mu}_i, i = 1, \ldots, K+1, j = 1, \ldots, n_i\}$.