Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

www.czasopisma.uni.lodz.pl/foe/
4(330) 2017

DOI: http://dx.doi.org/10.18778/0208-6018.330.10

Dorota Pekasiewicz
University of Lodz, Faculty of Economics and Sociology, Department of Statistical Methods, pekasiewicz@uni.lodz.pl

Agata Szczukocka
University of Lodz, Faculty of Economics and Sociology, Department of Statistical Methods, agata@statystyk.eu

Selected Statistical Tests for Median and Their Properties

Abstract: In the paper, a selection of statistical tests for median are presented. In particular, parametric and nonparametric significance tests are considered. In the case of parametric tests the critical regions are constructed on the basis of the known population distribution and the form of the alternative hypothesis. For chosen distributions the critical values are presented. In the case of nonparametric tests we consider tests for which the sample median dispersion is estimated based on order statistics of appropriate ranks. The use of the bootstrap method for the median dispersion estimation in the test statistic construction is the author's own proposal. The simulation analysis of the nonparametric tests' properties allows to compare these tests with each other, showing better results for the bootstrap variant, especially for small samples.

Keywords: median, critical value, significance test, parametric test, nonparametric test

JEL: C12, C14, C15

1. Introduction

The population median is a measure of location and can be applied to population description. The sample median is used in statistical inference, particularly in the verification of hypothesis about the population median. Information about the population distribution allows for the application of the parametric tests of the median. The application of the parametric tests to verify hypotheses about the median of symmetrical distribution, i.e. the Cauchy, logistic and normal distribution, is presented in the article. Statistical inference about the median of the Cauchy distribution is very important, because this distribution does not have finite moments of order greater or equal to one, in particular, if the expected value is infinite.

Nonparametric tests are used when no information is available about the population distribution. In papers by Olivie (2005) and Bloch, Gastwirth (1968) the nonparametric confidence intervals for the median are presented. The construction of statistical test for the median may be performed in the same way. The test statistic is based on the sample median and its dispersion. The median dispersion can be estimated based on the values of order statistics of appropriate ranks. Other nonparametric test proposed is the test in which median dispersion is estimated by the bootstrap method. Apart from theoretical considerations, the simulation studies of nonparametric tests' properties have been carried out. The research results allow for the formulation of conclusions about the sizes of the tests considered.

2. Parametric tests for median

Let X be a random variable with a continuous distribution F and M be the median of X.

Let us formulate the null parametric hypothesis:

$$H_0\colon M = m_0,$$

and its alternative hypothesis:

$$H_1\colon M < m_0 \text{ or } H_1'\colon M > m_0 \text{ or } H_1''\colon M \neq m_0,$$

where m_0 is a fixed real value.

Let $X_1, X_2, \ldots, X_n$ be a simple random sample drawn from the population of a symmetrical distribution $F(x)$ about the median M and α be a fixed significance level.

In view of the classical median estimator properties we assume that n is an odd number (see Zieliński, 2011).

The test statistic has the following form:

$$FB_M = Me - m_0, \tag{1}$$

where $Me = X_{\left(\frac{n+1}{2}:n\right)}$ is the sample median i.e. the order statistic of rank $\frac{n+1}{2}$.

The critical region of this test for the alternative H_1 given above has the form $P(FB_M < x_\alpha(Me)) = \alpha$ and it is determined on the basis of the theorem of the sample median distribution (see Domański, Pruska, 2000). We obtain the following critical value:

$$x_\alpha(Me) = F_0^{-1}\left(B^{-1}\left(\alpha, \frac{n+1}{2}, \frac{n+1}{2}\right)\right), \tag{2}$$

where $F_0(x)$ is the distribution of the random variable $X - M$ and $B^{-1}(p; a, b)$ is a p quantile of the beta distribution with parameters a and b.

For the alternative H_1', from equation $P(FB_M > x_{1-\alpha}(Me)) = \alpha$ we obtain the critical value:

$$x_{1-\alpha}(Me) = F_0^{-1}\left(B^{-1}\left(1-\alpha, \frac{n+1}{2}, \frac{n+1}{2}\right)\right) \tag{3}$$

and for the alternative H_1'' the critical region is determined by two values:

$$\begin{cases} x_{\frac{\alpha}{2}}(Me) = F_0^{-1}\left(B^{-1}\left(\frac{\alpha}{2}, \frac{n+1}{2}, \frac{n+1}{2}\right)\right), \\ x_{1-\frac{\alpha}{2}}(Me) = F_0^{-1}\left(B^{-1}\left(1-\frac{\alpha}{2}, \frac{n+1}{2}, \frac{n+1}{2}\right)\right). \end{cases} \tag{4}$$

Let us consider two symmetrical distributions of the random variable X: the Cauchy and logistic distribution and let us find the critical regions.

For the random variable X with the Cauchy distribution $Ca(m, \lambda)$, the critical values for the test of the median m are defined by the following formulae:

$$x_\alpha(Me) = \lambda \tan\left(\pi\left(B^{-1}\left(\alpha; \frac{n+1}{2}, \frac{n+1}{2}\right) - \frac{1}{2}\right)\right) \tag{5}$$

and

$$x_{1-\alpha}(Me) = -\lambda \tan\left(\pi\left(B^{-1}\left(\alpha; \frac{n+1}{2}, \frac{n+1}{2}\right) - \frac{1}{2}\right)\right) \tag{6}$$

for the hypotheses H_1: $M < m_0$ and H_1': $M > m_0$, respectively, and for the alternative H_1'': $M \neq m_0$:

$$\begin{cases} x_{\frac{\alpha}{2}}(Me) = \lambda \tan\left(\pi\left(B^{-1}\left(\frac{\alpha}{2};\frac{n+1}{2},\frac{n+1}{2}\right)-\frac{1}{2}\right)\right), \\ x_{1-\frac{\alpha}{2}}(Me) = -\lambda \tan\left(\pi\left(B^{-1}\left(\frac{\alpha}{2};\frac{n+1}{2},\frac{n+1}{2}\right)-\frac{1}{2}\right)\right). \end{cases} \quad (7)$$

For the random variable X with the logistic distribution $Lg(\mu, s)$ the critical values of statistical test for median have the following forms:

$$x_{\alpha}(Me) = -s\ln\left(\frac{1}{B^{-1}\left(\alpha;\frac{n+1}{2},\frac{n+1}{2}\right)}-1\right) \text{ for } H_1\text{: } M < m_0, \quad (8)$$

$$x_{1-\alpha}(Me) = s\ln\left(\frac{1}{B^{-1}\left(\alpha;\frac{n+1}{2},\frac{n+1}{2}\right)}-1\right) \text{ for } H_1\text{': } M > m_0 \quad (9)$$

and

$$\begin{cases} x_{\frac{\alpha}{2}}(Me) = -s\ln\left(\frac{1}{B^{-1}\left(\frac{\alpha}{2};\frac{n+1}{2},\frac{n+1}{2}\right)}-1\right), \\ x_{1-\frac{\alpha}{2}}(Me) = s\ln\left(\frac{1}{B^{-1}\left(\frac{\alpha}{2};\frac{n+1}{2},\frac{n+1}{2}\right)}-1\right), \end{cases} \text{ for } H_1\text{'': } M \neq m_0. \quad (10)$$

In the same way we can construct critical regions for the test of the median for random variable with the normal distribution.

In the case of an even sample size, the median estimator is the arithmetic mean of the middle two order statistics and it is not an unbiased estimator. Thus, in this case, we ignore one element, or, if possible, we draw one additional element and use the procedure presented above or we choose one of two order statistics $X_{\left(\frac{n}{2}:n\right)}$ or $X_{\left(\frac{n}{2}+1:n\right)}$ (see Zieliński, 2000).

3. Nonparametric tests for median

The application of the parametric tests is not always possible, because the population distribution is not always known. Then nonparametric test for the median has to be used.

In papers by Olivie (2005), Baszczyńska, Pekasiewicz (2010) the nonparametric confidence intervals for median are given. In the construction of one of the intervals, the dispersion of the sample median is determined by the following formula:

$$s_{Me} = 0.5(X_{(k_n:n)} - X_{(l_n:n)}), \tag{11}$$

where $X_{(k_n:n)}, X_{(l_n:n)}$ are order statistics with ranks:

$$l_n = \lceil n/2 \rceil - \left\lceil \sqrt{n/4} \right\rceil + 1, \tag{12}$$

$$k_n = n - l_n + 1. \tag{13}$$

The $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

The sample median and description of the sample median (11) can be applied to construct nonparametric, significance test for median. The test statistics has the following form:

$$t_M = \frac{Me - m_0}{S_{Me}}, \tag{14}$$

where Me is a sample median.

The critical region for this test is described by the critical values of the t-Student distribution with $Df = k_n - l_n$ degrees of freedom (see Olive, 2005). The critical values have the following forms:

$$x_\alpha(Me) = -t_\alpha \text{ for } H_1 : M < m_0,$$

$$x_{1-\alpha}(Me) = t_\alpha \text{ for } H_1' : M > m_0,$$

$$\left.\begin{array}{l} x_{\frac{\alpha}{2}}(Me) = -t_{\frac{\alpha}{2}}, \\ x_{1-\frac{\alpha}{2}}(Me) = t_{\frac{\alpha}{2}}, \end{array}\right. \text{ for } H_1'' : M \neq m_0.$$

The second considered nonparametric test statistic has the form:

$$t_M = \frac{Me - m_0}{S^*_{Me}}, \tag{15}$$

where S^*_{Me} is calculated on the basis of the bootstrap estimation method.

The bootstrap estimation methods can be found in Efron, Tibshirani (1993), Davison, Hinkley (1997) and Kisielińska (2015).

The bootstrap method involves generating, on the basis of the random sample $X_1, X_2, \ldots, X_n$, N bootstrap samples $X^*_1, X^*_2, \ldots, X^*_n$, , according to the bootstrap distribution $P\left(X^* = x_i\right) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$, where $x_1, x_2, \ldots, x_n$ are the elements of the original sample. Following that, on the basis of each bootstrap sample, the order statistics $X^{*i}_{(k_n:n)}, X^{*i}_{(l_n:n)}$ ($i = 1, 2, \ldots, N$) are determined, where l_n, k_n are given by (12) and (13), respectively.

The estimator S^*_{Me} is given by the formula:

$$S^*_{Me} = \frac{1}{N}\sum_{i=1}^{N} 0.5\left(X^{*i}_{(k_n:n)} - X^{*i}_{(l_n:n)}\right).. \tag{16}$$

The critical region for this test is determined by critical values of the t Student distribution.

4. Simulation studies

In the simulation analysis we considered:

1) two nonparametric tests:
   – test I – with statistic (14),
   – test II – with statistic (15),
2) parametric test – with statistic (1), for symmetric population distribution.

Some properties of the median test for selected classes of distributions, among others: Cauchy (*Ca*), logistic (*Lg*), normal (*N*), exponential (*Exp*) have been studied by means of the Monte Carlo method carried out in the Mathematica environment.

We started the simulation experiment by attempting to estimate the distributions of the test statistics defined by formulas (14) and (15). To this end random samples were drawn from the populations ($n \geq 100$) and the null hypothesis H_0: $M = m_0$ was verified against the alternative H_1'': $M \neq m_0$, where m_0 was the true value of the population median.

Each of these tests was applied to the verification of hypotheses $R = 10\ 000$ times. For test II the number of generating bootstrap samples was fixed as $N = 1000$.

The distributions of nonparametric test statistics were approximated by the t-Student distribution with degrees of freedom equal to $Df = k_n - l_n$ (see (12), (13)). For the population with the Cauchy distribution $Ca(2,1)$ and the random sample $n = 100$, 300 results of the approximations are presented in Figure 1 and Figure 2.

a) b)

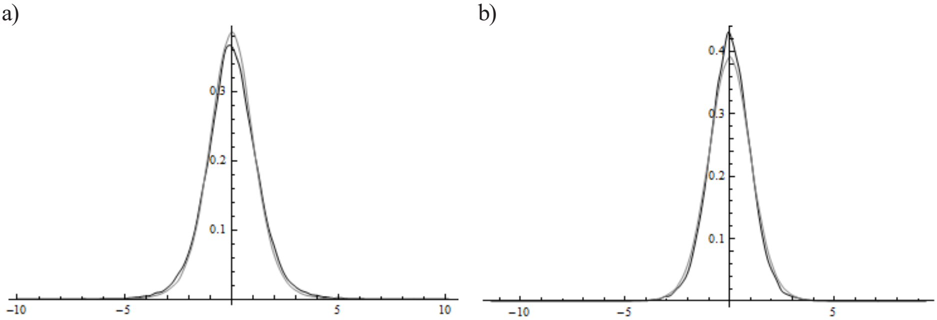

Figure 1. Density functions of test statistic (14) for sample size n and density distribution of t-Student distribution with $Df = k_n - l_n$ degree of freedom: a) $n = 100$, b) $n = 300$

Source: own investigations

a) b)

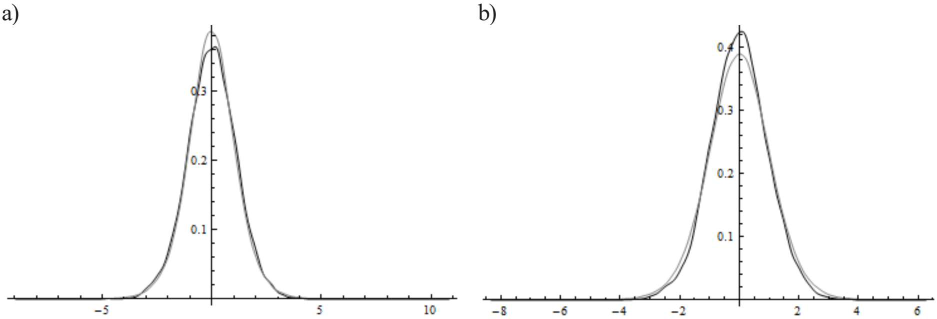

Figure 2. Density functions of test statistic (15) for sample size n and density distribution of t-Student distribution with $Df = k_n - l_n$ degree of freedom: a) $n = 100$, b) $n = 300$

Source: own investigations

We observe the divergence between the empirical distribution and the theoretical one, especially near the mode, but we are more interested in the tail of distribution.

The results may raise doubts about the use of the test, so further simulations were carried out. In the next steps the test power and size of these nonparametric tests were studied.

Figure 3 presents the estimated power of these tests for H_0: $M = 2$ against H_1'': $M \neq 2$, for the population with the Cauchy distribution $Ca(2,1)$. The significance level was equal to $\alpha = 0.05$ and sample size was $n = 100$.

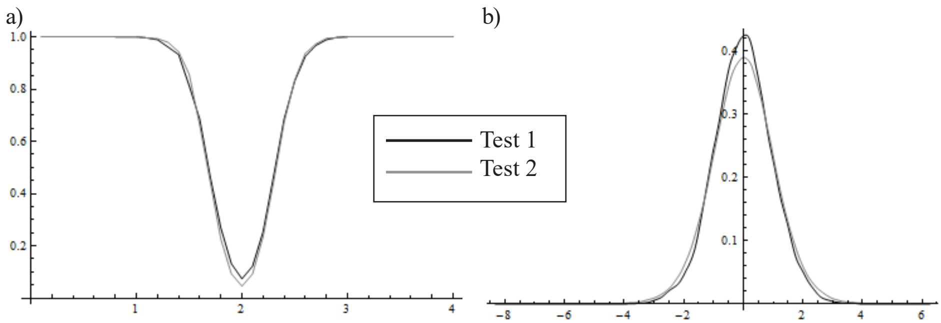

Figure 3. Estimated power of test I and test II: a) $n = 100$, b) $n = 300$

Source: own investigations

Table 1. Estimated sizes of nonparametric tests for selected population distributions and selected sample sizes

Distribution	Sample size (n)	Test I	Test II
$Ca(2,1)$	100	0.0753	0.0463
	300	0.0559	0.0453
$Ca(2,4)$	100	0.0768	0.0517
	300	0.0553	0.0434
$Lg(2,1)$	100	0.0795	0.0545
	300	0.0588	0.0460
$Lg(2,4)$	100	0.0761	0.0547
	300	0.0603	0.0502
$N(2,1)$	100	0.0782	0.0560
	300	0.0570	0.0480
$N(2,4)$	100	0.0732	0.0489
	300	0.0574	0.0450
$Exp(1/3)$	100	0.0843	0.0566
	300	0.0620	0.0503
$Exp(1)$	100	0.0739	0.0538
	300	0.0606	0.0479

Source: own calculations

Sizes of the nonparametric tests I and II, for selected population distribution and selected sample sizes are presented in table 1.

Table 2. Estimated test sizes and critical regions of parametric tests for selected population distributions and sample sizes

Distribution	Sample size (*n*)	$x_{\frac{\alpha}{2}}(Me)$	$x_{1-\frac{\alpha}{2}}(Me)$	Test size
Ca(2,1)	100	–0.3139	0.3139	0.0482
	300	–0.2198	0.2198	0.0470
Ca(2,4)	100	–1.2557	1.2557	0.0468
	300	–0.8793	0.8793	0.0473
Lg(2,1)	100	–0.3923	0.3923	0.0498
	300	–0.2773	0.2773	0.0499
Lg(2,4)	100	–1.5690	1.5690	0.0496
	300	–1.1091	1.1091	0.0487
N(2,1)	100	–0.0245	0.0245	0.0488
	300	–0.1417	0.1417	0.0480
N(2,4)	100	–0.9806	0.9806	0.0491
	300	–0.5669	0.5669	0.0492

Source: own calculations

The bootstrap method of the median dispersion estimation (test II) improved some characteristics of the nonparametric test for median. The size of test II is lower than the size of test I and the test powers are similar.

For symmetric distributions the parametric test with statistic (1) was applied to verify the hypotheses about median. We considered the Cauchy, logistic and normal population. For each of the distributions the critical regions are determined analytically, in particular the formula (7) defines the critical values for the Cauchy distribution and formula (10) – for the logistic distribution and alternative hypothesis H_1'': $M \neq m_0$. The parametric procedures of verification of the hypothesis H_0 against H_0'' was repeated R = 10 000 times. The sizes of the parametric tests are shown in table 2. Additionally, for $\alpha = 0.05$, the critical values for the parametric tests are presented.

Comparing the sizes of the nonparametric tests with the size of the appropriate parametric test, we notice that the size of test II is approximately equal to the size of the parametric tests. Then, the bootstrap method of the median dispersion estimation improved the properties of the nonparametric non-bootstrap test.

5. Conclusions

In the paper significance parametric and nonparametric tests for the median were considered. The bootstrap test was one of the tests considered, but the bootstrap approach was not associated with the critical value estimation (percentile bootstrap test), but with the median dispersion estimation. This test was proposed by the authors and it is the subject of this research. The simulation analysis involved nonparametric tests in which the dispersion of the sample median was estimated by the order statistics of appropriate ranks (non-bootstrap and bootstrap methods) and parametric test (for comparison). The test with the bootstrap estimation of the sample median dispersion has proven to be effective due to its size being smaller than the size of the non-bootstrap, nonparametric test. For bigger sample sizes the size of proposed test is comparable with the size of parametric test for median (for the Cauchy, logistic and normal distribution of populations). Other bootstrap tests for median, not considered in this paper, are presented in Domański et al. (2014).

References

Baszczyńska A., Pekasiewicz D. (2010), *Selected Methods of Interval Estimation of the Median. The Analysis of Accuracy of Estimation*, "Acta Universitatis Lodziensis. Folia Oeconomica", no. 235, pp. 21–30.

Białek J. (2014), *Simulation Study of an Original Price Index Formula*, "Communications in Statistics, Simulation and Computation", no. 43(2), pp. 285–297.

Bloch D.A., Gastwirth J.L, (1968), *On a Sample Estimate of Reciprocal of the Density Function*, "Annals of Mathematical Statistics", no. 39, pp. 1083–1085.

Davison A.C., Hinkley D.V. (1997), *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge.

Domański C., Pekasiewicz D., Baszczyńska A., Witaszczyk A. (2014), *Testy statystyczne w procesie podejmowania decyzji*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Domański C., Pruska K. (2000), *Nieklasyczne metody statystyczne*, PWE, Warszawa.

Efron B., Tibshirani R.J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Kisielińska J. (2015), *Bootstrapowy estymator mediany dla próby o nieparzystej liczbie elementów*, "Metody Ilościowe w Badaniach Ekonomicznych", no. 14(3), pp. 172–182.

Olivie D.J. (2005), *A Simple Confidence Interval for the Median*, http: http://lagrange.math.siu.edu/Olive/ppmedci.pdf [accesed: 5.09.2016].

Zieliński R. (2000), *A Median – Unbiased Estimator of the Characteristic Exponent of a Symmetric Stable Distribution*, „Statistics", no. 39(1), pp. 67–71.

Zieliński R. (2011), *Statystyka matematyczna stosowana. Elementy*, Wydawnictwo Centrum Studiów Zaawansowanych Politechniki Warszawskiej, Warszawa.

Wybrane testy statystyczne dla mediany i ich własności

Streszczenie: W pracy rozważane są wybrane testy pozwalające weryfikować hipotezy o wartości mediany populacji, w szczególności parametryczne i nieparametryczne testy istotności. W przypadku testów parametrycznych obszary odrzucenia hipotezy zerowej konstruuje się w oparciu o znany rozkład populacji oraz postać hipotezy alternatywnej. W pracy zaprezentowane są testy dla mediany (statystyki i obszary odrzucenia hipotezy zerowej) dla wybranych rozkładów populacji: Cauchy'ego i logistycznego. Rozważania dotyczące testów nieparametrycznych obejmują testy, w których zróżnicowanie mediany z próby szacowane jest w oparciu o statystyki pozycyjne odpowiednich rzędów.

Zastosowanie metody bootstrapowej do szacowania zróżnicowania mediany i konstrukcja statystyki testowej przy jej wykorzystaniu stanowi autorską propozycję. Wyniki symulacyjnej analizy własności testów nieparametrycznych wskazują na lepsze własności testu wykorzystującego bootstrapową estymację zróżnicowania mediany.

Słowa kluczowe: kwantyl, mediana, test istotności, parametryczny test, nieparametryczny test

JEL: C12, C14, C15

Received: 2017-01-14; verified: 2017-04-09. Accepted: 2017-08-02