B A D A N I A   O P E R A C Y J N E   I   D E C Y Z J E

Nr 1 2006

Grzegorz KOŃCZAK*
Janusz WYWIAŁ*

TESTING HYPOTHESIS ON STABILITY OF EXPECTED VALUE AND VARIANCE

The simple samples are independently taken from normal distribution. The two functions of the sample means and sample variances are considered. The density functions of these two statistics have been derived. These statistics can be applied for verifying the hypothesis on stability of expected value and variance of normal distribution considered, e.g., in statistical process control. The critical values for these statistics have been found using numerical integration. The tables with approximated critical values of these statistics have been presented.

Keywords: *density function, sample variance, test statistic, numerical integration, statistical process control*

1. Introduction

One of the problems of statistical process control is considered. It is a procedure (so called control charts) for monitoring stability of the expected value and the variance of diagnostic variables. We assume that during the first $k \geq 2$ periods the mean values of a diagnostic variable are the same but unknown. The same situation is with the variance of the variable. The unbiased estimators of the expected value μ and the variance σ^2 evaluated on the basis of data observed in the first k periods and the $k + 1$ period are $\overline{X}$, $\hat{S}$ and $\overline{X}_{k+1}$ and $\hat{S}_{k+1}$, respectively. Our problem is the following: is the process (characterized by the diagnostic variable) stable in all of the periods? If yes, the distances: $|\overline{X} - \overline{X}_{k+1}|$ or $|\hat{S} - \hat{S}_{k+1}|$ should not be significant. Such a problem is considered, e.g., in [4]. More formally, we have the problem of testing the hypothesis

* Department of Statistics, Katowice University of Economics, ul. Bogucicka 14, 40-226 Katowice. koncz@ae.katowice.pl, wywial@ae.katowice.pl

$$H_0: E(\hat{X}_{k+1}) = E(\hat{X}) \quad \text{and} \quad E(\hat{S}^2_{k+1}) = E(\hat{S}^2).$$

We are going to construct a test statistic for this hypothesis in the next paragraphs.

2. Basic definitions and properties

Let $\boldsymbol{J}_a$ be a column vector consisting of a element equal one and let $\boldsymbol{I}_a$ be a unit matrix of degree a. Moreover, let $\mathbf{X} = [\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X}_i, ..., \mathbf{X}_{k+1}]$, where $\mathbf{X}_i = [\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{ij}, ..., \mathbf{X}_{in_i}]$ ($i = 1, 2, \ldots, k+1$, $k \geq 2$, $j = 1, 2, \ldots, n_i$).

We will consider the following statistics:

$$\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} = \frac{1}{n_i}\mathbf{X}_i\mathbf{J}_{n_i},$$

for $i = 1, 2, \ldots, k+1$,

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{k}\overline{X}_i n_i = \frac{1}{n}\mathbf{X}\mathbf{J}_n, \quad \text{where} \quad n = \sum_{i=1}^{k} n_i,$$

$$\hat{S}_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 = \frac{\mathbf{X}_i\mathbf{M}_i\mathbf{X}_i^T}{n_i - 1}, \quad \mathbf{M}_i = \mathbf{I}_{n_i} - \frac{1}{n_i}\mathbf{J}_{\mathbf{n}_i}\mathbf{J}_{n_i}^T,$$

$$\hat{S}^2 = \frac{1}{n-k}\sum_{j=1}^{k}(n-k)\hat{S}_i^2 = \frac{1}{n-k}\mathbf{X}\mathbf{M}\mathbf{X}^\mathbf{T}, \quad \mathbf{M} = [\mathbf{M}_i],$$

where $\boldsymbol{M}_i$ is the block-diagonal matrix of degree n,

$$\widetilde{S}^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\overline{X}_i - \overline{X})^2 n_i = \frac{\mathbf{X}\mathbf{M}\mathbf{X}^T}{k-1}, \quad \mathbf{N} = \mathbf{I}_n - \mathbf{M} - \frac{1}{n}\mathbf{J}_n\mathbf{J}_n^T,$$

where $\hat{S}_i^2$ is the sample variance within the i-th group, $\hat{S}^2$ is the mean of group variances and $\widetilde{S}^2$ is the variance between groups. Moreover, let us note that $\mathbf{M}_i^\mathbf{2} = \mathbf{M}_i$, $\mathbf{M}^\mathbf{2} = \mathbf{M}$ and $\mathbf{N}^\mathbf{2} = \mathbf{N}$.

The particular case of the theorem on independence of the quadratic or linear forms of normal vectors corresponding to our problem is as follows (see [3], pp. 224 and next).

Theorem 1. Let **X** have a non-singular normal distribution $N(\mu\mathbf{J}_n, \sigma^2\mathbf{I}_n)$ and $Q_A = \mathbf{XAX}^T$, $Q_B = \mathbf{XBX}^T$, $\boldsymbol{L} = \boldsymbol{Xa}$ where $\boldsymbol{a}$ is a column non-random vector, $\boldsymbol{A}$, $\boldsymbol{B}$ are symmetric and non-random matrices of degree n each. Then, the set of necessary and sufficient conditions for Q_A and Q_B to be independently distributed is a) $\boldsymbol{AB} = \boldsymbol{O}$ or b) $\boldsymbol{ABJ}_n = \boldsymbol{O}$. The quadratic form Q_A and the linear form $\boldsymbol{L}$ are to be independently distributed if and only if $\boldsymbol{Aa} = \boldsymbol{O}$.

On the basis of this theorem we can show that the statistics in the following pairs are independently distributed $(\overline{X}, \widetilde{S}^2)$, $(\overline{X}, \hat{S}^2)$, $(\hat{S}^2, \widetilde{S}^2)$. Moreover, the statistics $\overline{X}_{k+1}, \hat{S}^2_{k+1}$ are independent and they are independent of each of the statistics $\overline{X}, \hat{S}^2$ and $\widetilde{S}^2$.

The obtained result and the well known definitions let us derive the following distributions:

$$Z = \frac{\overline{X}_{k+1} - \overline{X}}{\sigma\sqrt{\dfrac{1}{n_{k+1}} + \dfrac{1}{n}}} : N(0,1). \tag{1}$$

Moreover,

$$U_1 = \frac{(k-1)\widetilde{S}^2}{\sigma^2} : \chi^2_{k-1}, \tag{2}$$

$$U_2 = \frac{(n-k)\hat{S}^2}{\sigma^2} : \chi^2_{n-k} \tag{3}$$

and

$$U_{0,i} = \frac{(n_i - 1)\hat{S}^2_i}{\sigma^2} : \chi^2_{n_i - 1} \quad i = 1, 2, \ldots, k+1. \tag{4}$$

On the basis of these expressions we have

$$F_1 = \frac{Z^2\sigma^2}{\widetilde{S}^2} : F(1, k-1), \tag{5}$$

$$F_2 = \frac{Z^2\sigma^2}{\hat{S}^2} : F(1, n-k), \tag{6}$$

$$F_3 = \frac{\hat{S}^2_{k+1}}{\hat{S}^2} : F(n_{k+1} - 1, n-k), \tag{7}$$

$$F_4 = \frac{\hat{S}^2_{k+1}}{\widetilde{S}^2} : F(n_{k+1} - 1, k-1), \tag{8}$$

where $F(r, m)$ denotes the well known F distribution with r and m degrees of freedom and the following density function:

$$f(g) = \frac{\Gamma\left(\frac{r}{2} + \frac{m}{2}\right)}{\Gamma\left(\frac{r}{2}\right) \cdot \Gamma\left(\frac{m}{2}\right)} \left(\frac{r}{m}\right)^{\frac{r}{2}} \frac{g^{\frac{r}{2}-1}}{\left(1 + \frac{r}{m} g\right)^{\frac{1}{2}(r+m)}} I_{(0,\infty)}(g) . \tag{9}$$

3. The statistics Q_1 and Q_2

Let us consider two statistics Q_1 and Q_2 given as follows

$$Q_1 = F_1 + (F_3 - 1)^2 = \frac{Z^2}{\tilde{S}^2} + \left(\frac{\hat{S}^2_{k+1}}{\hat{S}^2} - 1\right)^2 , \tag{10}$$

$$Q_2 = F_2 + (F_4 - 1)^2 = \frac{Z^2}{\hat{S}^2} + \left(\frac{\hat{S}^2_{k+1}}{\tilde{S}^2} - 1\right)^2 , \tag{11}$$

where $F_1 - F_4$ are defined by expressions (5)–(8), respectively.

On the basis of the previous results the distributions of the random variables Z, U_1, U_2 and $U_{0,\,k+1}$ are independent. Finally, this and the fact the samples are independent lead to the conclusion that the distributions of the statistics F_1 and F_3 as well as the distributions of F_2 and F_4 are independent.

The density function of random variable F_1 is as follows:

$$f_1(g) = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{k-1}{2}\right)} \frac{1}{\sqrt{(k-1)g}} \frac{1}{\left(1 + \frac{g}{k-1}\right)^{\frac{k}{2}}} I_{(0,\infty)}(g) , \tag{12}$$

where

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} .$$

The density function of random variable F_2 is:

$$f_2(g) = \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{n-k}{2}\right)} \frac{1}{\sqrt{(n-k)g}} \frac{1}{\left(1+\frac{g}{n-k}\right)^{\frac{n-k+1}{2}}} I_{(0,\infty)}(g) . \tag{13}$$

The density functions of F_3 and F_4 are as follows:

$$f_3(g) = c_3 \frac{g^{\frac{r}{2}-1}}{\left(1+\frac{r}{n-k}g\right)^{\frac{r+n-k}{2}}} I_{(0,\infty)}(g) , \tag{14}$$

$$f_4(g) = c_4 \frac{g^{\frac{r}{2}-1}}{\left(1+\frac{r}{k-1}g\right)^{\frac{r+k-1}{2}}} I_{(0,\infty)}(g) . \tag{15}$$

where $r = n_{k+1} - 1$ and

$$c_3 = \frac{\Gamma\left(\frac{r+n-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\cdot\Gamma\left(\frac{n-k}{2}\right)}\left(\frac{r}{n-k}\right)^{\frac{r}{2}}, \quad c_4 = \frac{\Gamma\left(\frac{r+k-1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\cdot\Gamma\left(\frac{k-1}{2}\right)}\left(\frac{r}{k-1}\right)^{\frac{r}{2}} .$$

Let us derive the distribution of random variable $(F_3 - 1)^2$. If $b = g - 1$ then we have $g = b + 1$ and $dg = db$ and

$$f_{3a}(b) = c_3 \frac{(b+1)^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k}+\frac{r}{n-k}b\right)^{\frac{r+n-k}{2}}} I_{(0,\infty)}(g) . \tag{16}$$

When $v = b^2$, then we have $b = \pm\sqrt{v}$ and $|db| = \frac{1}{2\sqrt{v}}$. If $b \in (-1,0]$, then $b = -\sqrt{v}$ and for $b \in (0,1)$, we have $b = \sqrt{v}$.

If $v \in [0,1)$

$$f_{3b}(v) = c_3 \frac{1}{2\sqrt{v}} \left(\frac{(1-\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} - \frac{r}{n-k}\sqrt{v} \right)^{\frac{r+n-k}{2}}} + \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} + \frac{r}{n-k}\sqrt{v} \right)^{\frac{r+n-k}{2}}} \right). \quad (17)$$

If $v \in [1,\infty)$ then

$$f_{3b}(v) = c_3 \frac{1}{2\sqrt{v}} \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} + \frac{r}{n-k}\sqrt{v} \right)^{\frac{r+n-k}{2}}}. \quad (18)$$

We can write the density function of $(F_3 - 1)^2$ in the following way

$$f_{3b}(v) = c_3 \frac{1}{2\sqrt{v}} \left[\frac{(1-\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} - \frac{r}{n-k}\sqrt{v} \right)^{\frac{r+n-k}{2}}} I_{(0,1)}(v) + \right.$$

$$\left. + \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} + \frac{r}{n-k}\sqrt{v} \right)^{\frac{r+n-k}{2}}} I_{(0,\infty)}(v) \right]. \quad (19)$$

If W: $F(1, k-1)$, then the density function of W is given by expression (12). Let V be the random variable of F distribution with r and n-k degrees of freedom. The density function of V is given by (19). Now, we are going to evaluate the density functions of the random variable: $Q_1 = W + V$. The density function of the statistic Q_1 is as follows

$$h_1(q) = \int_0^\infty f_{3b}(v) f_1(q-v)\,dv \quad (20)$$

$$h_1(q) = c_1 \int_0^\infty \left[\frac{(1-\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} - \frac{r}{n-k}\sqrt{v}\right)^{\frac{r+n-k}{2}}} I_{(0,1)}(v) \right.$$

$$\left. + \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{n-k+r}{n-k} + \frac{r}{n-k}\sqrt{v}\right)^{\frac{r+n-k}{2}}} I_{(0,\infty)}(v) \right] \cdot \frac{1}{\sqrt{(k-1)v(q-v)\left(1+\frac{g-v}{k-1}\right)^{\frac{k}{2}}}}, \tag{21}$$

where

$$c_1 = \frac{\Gamma\left(\frac{k}{2}\right)}{2\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{k-1}{2}\right)} \frac{\Gamma\left(\frac{r+n-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\cdot\Gamma\left(\frac{n-k}{2}\right)} \left(\frac{r}{n-k}\right)^{\frac{r}{2}}.$$

Similarly, the density function of Q_2 is derived in the following way.

$$h_2(q) = \int_0^\infty f_{4b}(v) f_2(q-v)dv, \tag{22}$$

where f_{4b} is the density function of $(F_4 - 1)^2$

$$f_{4b}(v) = c_4 \frac{1}{2\sqrt{v}} \left[\frac{(1-\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{k-1+r}{k-1} - \frac{r}{k-1}\sqrt{v}\right)^{\frac{r+k-1}{2}}} I_{[0,1)}(v) \right.$$

$$\left. + \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{k-1+r}{k-1} + \frac{r}{k-1}\sqrt{v}\right)^{\frac{r+k-1}{2}}} I_{(0,\infty)}(v) \right]. \tag{23}$$

Finally, the density function $h_2(q)$ of the statistic Q_2 is as follows

$$h_2(q) = c_2 \int_0^{\infty} \left[\frac{(1-\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{k-1+r}{k-1} - \frac{r}{k-1}\sqrt{v}\right)^{\frac{r+k-1}{2}}} I_{[0,1)}(v) + \frac{(1+\sqrt{v})^{\frac{r}{2}-1}}{\left(\frac{k-1+r}{k-1} + \frac{r}{k-1}\sqrt{v}\right)^{\frac{r+k-1}{2}}} I_{(0,\infty)}(v) \right] \cdot \frac{1}{\sqrt{(n-k)v(q-v)}} \frac{1}{\left(1+\frac{g-v}{n-k}\right)^{\frac{n-k+1}{2}}}, \quad (24)$$

where

$$c_2 = \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{2\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{n-k}{2}\right)} \frac{\Gamma\left(\frac{r+k-1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\cdot\Gamma\left(\frac{k-1}{2}\right)} \left(\frac{r}{k-1}\right)^{\frac{r}{2}}.$$

The distribution functions of Q_1 or Q_2 are evaluated by means of the following integral

$$H_i(q) = P(Q_i < q) = \int_0^q h_i(s)ds, \quad I = 1, 2. \quad (25)$$

For the given significance level α the quantil $q_{1-\alpha}$ is determined on the basis of the integral:

$$\int_{q_{1-\alpha}}^{\infty} h_i(s)ds = \alpha\ . \quad (26)$$

4. Numerical computations

An exact solution of equation (26) is very difficult. In this situation, the quantiles of Q_1 and Q_2 can be found using numerical integration (see, e.g., [2], [1]). The quantiles were found for three significance levels (α = 0.01, 0.05, 0.1). Table 1 presents quantiles of the statistic Q_1 and Table 2 presents quantiles of the statistic Q_2. These quantiles were evaluated for the case $n_1 = n_2 = \ldots = n_k = n_{k+1} = 5$. These quantiles have been evaluated for the numbers of groups k from 4 to 10 and for 15, 20, 25 and 30.

Table 1

Quantiles of statistic Q_1

Number of groups k	Significance level α		
	0.10	0.05	0.01
4	5.85	10.38	31.19
5	4.86	8.08	21.43
6	4.36	6.96	16.75
7	4.06	6.31	14.21
8	3.86	5.90	12.67
9	3.72	5.61	11.64
10	3.62	5.40	10.92
15	3.34	4.85	9.15
20	3.23	4.62	8.45
25	3.16	4.50	8.08
30	3.12	4.42	7.85

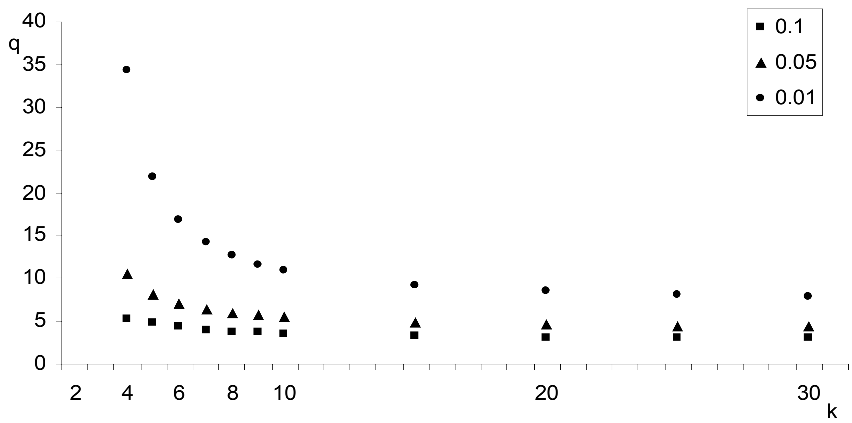

Fig. 1. The graphic display of quantiles of statistic Q_1 for significance level α = 0.1, 0.05 and 0.01

The Figure 1 presents quantiles of the statistic Q_1 for the significance levels 0.1, 0.05 and 0.01. These quantiles are presented for the same cases as in Table 1. Figure 2 presents the same results as Figure 1 but for the statistic Q_2.

Table 2

Quantiles of statistic Q_2

Number of groups k	Significance level α		
	0.10	0.05	0.01
4	3.42	5.00	10.71
5	3.30	4.76	9.41
6	3.23	4.62	9.21
7	3.18	4.53	8.34
8	3.15	4.46	8.07
9	3.12	4.41	7.88
10	3.10	4.37	7.74
15	3.04	4.26	7.39
20	3.01	4.20	7.23
25	2.99	4.17	7.21
30	2.98	4.15	7.10

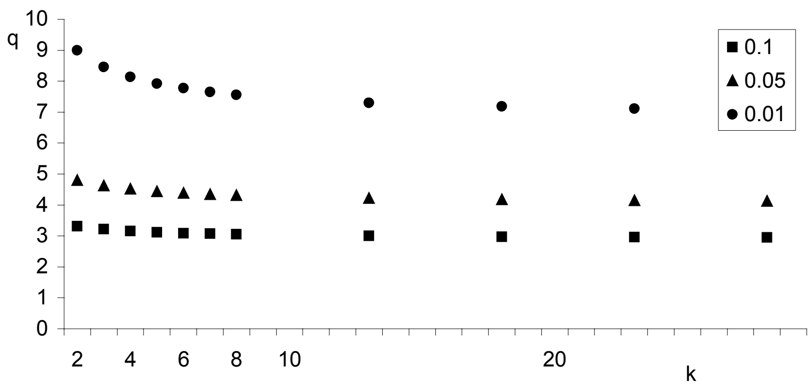

Fig. 2. The graphic display of quantiles of statistic Q_2 for significance level α = 0.1, 0.05 and 0.01

References

[1] BRANDT S., *Statistical and Computational Methods in Data Analysis*, Springer Verlag, New York 1997.

[2] DAHLQUIST G., BJORCK A., *Numerical Methods* (in Polish), PWN, Warszawa 1992.

[3] MATHAI A.M., PROVOST S.B., *Quadratic forms in random variables*, Marcel Dekker, Inc., New York, Basel, Hong Kong 1992.

[4] THOMPSON J.R., KORONACKI J., *Statistical Process Control: The Deming Paradigm and Beyond*, Chapman and Hall/CRC, New York, London 2001.

Testowanie hipotezy o stabilności wartości oczekiwanej i wariancji

W pracy jest rozważane zagadnienie jednoczesnej stabilności wartości przeciętnej i wariancji. Próby proste są pobierane niezależnie z populacji o rozkładzie normalnym. Rozważa się dwie funkcje średniej i wariancji z próby. Dla rozważanych statystyk zostały wyprowadzone funkcje gęstości. Proponowane statystyki mogą być wykorzystane do weryfikacji hipotezy o stabilności wartości oczekiwanej i wariancji dla rozkładu normalnego. Hipoteza taka może być rozważana np. w statystycznym sterowaniu procesem przy konstrukcji kart kontrolnych. Bardzo trudne jest dokładne wyznaczenie kwantyli rozważanych statystyk. Dlatego wartości krytyczne dla tych statystyk zostały wyznaczone dla trzech zwykle używanych poziomów istotności (α = 0,01, 0,05 i 0,1) dla prób o liczebnościach od 4 do 30 z wykorzystaniem całkowania numerycznego. Zaprezentowano tablice wartości krytycznych dla tych statystyk. Zaproponowane statystyki i wyznaczone wartości krytyczne mogą być również przydatne do wykrywania zmian w procesach produkcyjnych.

Słowa kluczowe: *funkcja gęstości, wariancja z próby, test, numeryczne całkowanie, statystyczna kontrola procesu*