ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

Grzegorz Kończak, Maria Czogała***

ON THE MULTIVARIATE TEST FOR STABILITY OF THE POPULATION PROPORTIONS

Abstract

This paper investigates the problem of testing the hypothesis of the stability proportion for multiple attributes. It is assumed that two random samples of size n_0 and n_1 respectively are selected from the population. In these samples each item is assessed with regard to its k attributes, each attribute is assessed alternatively. The verification of the hypothesis that the fractions of elements for each variable are stable over time is discussed. The applications of the Pearson chi-square test, the chi-square test with Yates continuity correction and the Fisher exact test for testing equality of proportions are analyzed. In this paper a test is considered which makes it possible to verify the hypothesis that several proportions fulfill given assumptions simultaneously.

Key words: test, equality of two proportions, Fisher exact test, chi-square test.

1. Introduction

The problem of stability proportions over time often appears in quality control in situations where items are assessed alternatively, i.e. they are assumed to be either good or bad, with regard to the fulfillment of a set of requirements. During the control, products are often evaluated in terms of meeting numerous standards, which are assessed alternatively (colour, scratches). Simultaneous verification of multiple attributes is of great importance nowadays when products have to meet even more demanding quality standards. This paper presents the advantages of the use of the proposed test rather than several tests for proportions for each attribute being assessed separately. The properties of the proposed test are compared with the results achieved in classical approaches.

* Ph.D., Department of Statistics, University of Economics Katowice.
** Master, Department of Statistics, University of Economics Katowice.

The problem of simultaneously testing stability of multiple proportions struck the authors during their cooperation with a production company in Dąbrowa Górnicza.

2. Basic definitions and notation

Let X_{0i}, X_{1i} ($i = 1, 2, \ldots, k$) be k independent pairs of binomial random variables with parameters n_0, p_{0i} and n_1, p_{1i} respectively. The probability mass function of these random variables can be written as follows:

$$P(X_{ti} = j) = p_{tij} = \binom{n}{j} p_{ti}^{j}(1 - p_{ti})^{n_t - j} \quad i = 1, 2, \ldots, k \;\; j = 0, 1, \ldots, n_t, \; t = 0, 1 \quad (1)$$

The base period we denote as $t = 0$ and the present period we denote as $t = 1$. The number of successes for random variables (1) in period t for k-*th* variable we denote by m_{tj}.

Let $n = n_0 + n_1$ be the sample size and let $m = m_0 + m_1$ be the total number of successes (S). The total number of failures (F) equals to $n - m$. We will be testing the stability of k proportions over time. We will consider the hypothesis:

$$H_0 : p_{11} = p_{01} \wedge p_{12} = p_{02} \wedge \ldots \wedge p_{1k} = p_{0k} \quad (2)$$

The alternative hypothesis is:

H_1: for at least one variable is $p_{1i} \neq p_{0i}$ $\quad i = 1, 2, \ldots, k$.

In quality control the results for the period $t = 0$ come from the base period, when the process is assumed to be stable and the results for $t = 1$ come from the present period. The probabilities of successes for i-*th* variable for the base period p_{0i} ($i = 1, 2, \ldots, k$) can be assessed on the basis of results received from previous observations:

$$\hat{p}_{0i} = \frac{m_{0i}}{n_0} \quad (3)$$

In the practice of quality control the hypothesis of the equality of fraction and approved earlier quality level is verified. In situations when the quality level had not previously been set, the fraction from the present period is usually compared with the fraction from the base period, when the process was assumed to be stable, calculated on the basis of (3). The procedure of verification of the hypothesis (2), based on exact distributions for each variable separately is presented by Levinson (2004). The area in between control limits where there is no basis for the rejection of the null hypothesis takes form of k-dimensional cubes.

However, such an approach is not always fully sufficient and satisfactory. The comparison of the above solution to the one allowing for simultaneous examination of two proportions is shown in Fig. 1. The area in between the control limits where there is no basis for the rejection of the null hypothesis obtained with the application of Levinson's proposition (2004) takes form of a rectangle. With the application of simultaneous examination of two proportions, it takes the form of the interior of an ellipse ($k = 2$), and an ellipsoid in a general case ($k > 2$). Point P which represents the quality of a which product fulfills the requirements for both of the variables separately, and at the same time does not meet the quality control standards is schematically shown in Fig. 1.

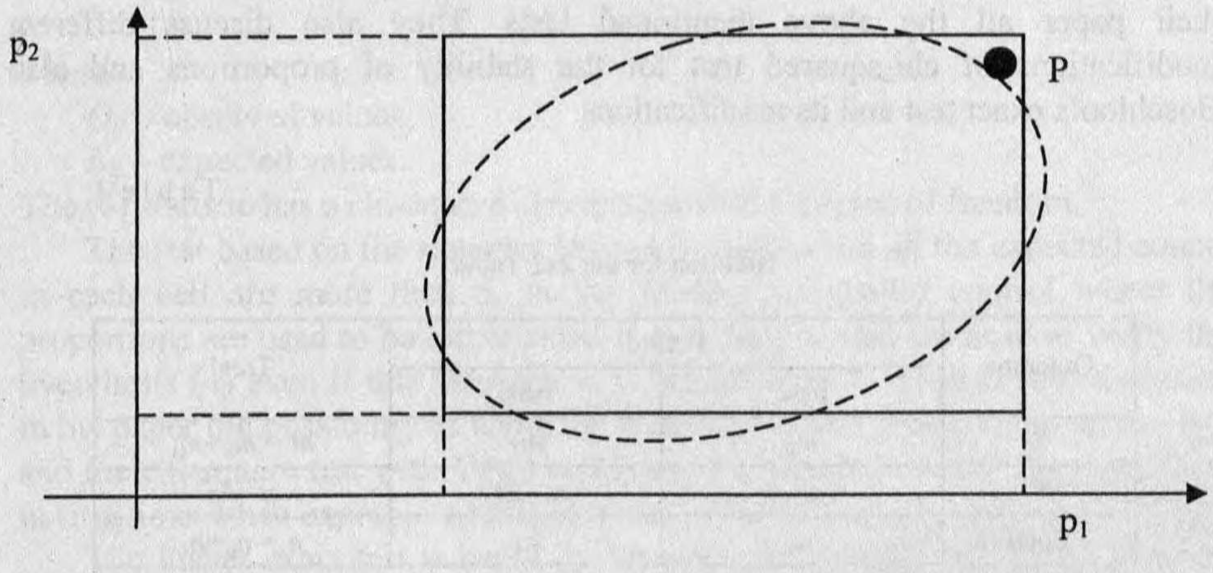

Fig. 1. The borders of the critical area for a 2-dimentional test for proportion equality (ellipse) and the exact test for each of the variables (rectangle)

Source: own study.

3. The case of the one dimension

In the procedures of quality control we are often interested in stability of proportions over time and we verify the hypothesis (2) of equality of the proportions, which are calculated on the basis of data from the previous and the present period.

In special case when $k = 1$ this hypothesis can be written:

$$H_0 : p_1 = p_0 \quad (4)$$

and the alternative hypothesis is:

$$H_1 : p_1 \neq p_0$$

The problem of verification of the hypothesis (4) in small sample case is analyzed by Agresti, Coull (1998), Little (1989), Mehrotra *et al.* (2003), Berger *et al.* (2003).

The results we obtain from the process can be presented as shown in Table 1. Different tests for verification of the hypothesis (4) for small samples are discussed in many papers and publications. The classical approach is based either on z score test or equivalent chi-square test. For small samples the Fisher exact test is used very often. Mehrotra, Chan, Berger (2003) discuss in their paper all the above mentioned tests. They also discuss different modifications of chi-squared test for the stability of proportions and also Boschloo's exact test and its modifications.

Table 1

Notation for the 2x2 Table

Outcome	Period		Total
	base	base	
S	m_0	m_1	$m = m_0+m_1$
F	$n_0 - m_0$	$n_1 - m_1$	$n - m$
Total	n_0	n_1	$n = n_0+n_1$

Source: own study.

Let $\hat{p}_0 = \frac{m_0}{n_0}$ and $\hat{p}_1 = \frac{m_1}{n_1}$ be the estimated proportions in two periods. Suissa and Shuster (1985) introduced the "Z-pooled" statistic

$$Z_p = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}} \quad (5)$$

where $\hat{p} = \frac{m_0 + m_1}{n_0 + n_1}$. For this statistic we can obtain exact true p-values, especially for small sample sizes, from formula (Mehrotra, Chan, Berger 2003)

$$p_{Z_p}(m_1, m_2) = \sup_{0 \le p \le 1} \left\{ \sum_{i=0}^{n_0} \sum_{j=0}^{n_1} P_{H_o}(X_0 = i, X_1 = j \mid p) \times I_{|Z_p(i,j)| \ge |Z_p(m_1, m_2)|} \right\}$$

where $I_A = \begin{cases} 0 \; A \text{ is false} \\ 1 \; A \text{ is true} \end{cases}$.

Instead of this statistic we can use the chi-square statistic

$$\chi_1^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{6}$$

where:

O_{ij} – observed values,

E_{ij} – expected values.

The (6) statistic has a chi-square distribution with 1 degree of freedom.

The test based on the statistics (6) can be used when all the expected counts in each cell are more than 5. In the practice of quality control where the proportions are used to be rather small it may happen that we need to verify the hypothesis (4) even if this assumption is not fulfilled. Little (1989) discusses in his paper the possibility of using the Fisher exact test, Pearson chi-square test and the chi-square test with Yates continuity correction to verify the hypothesis in situations when expected counts of distinguished elements are more than 1.

The Fisher exact test is based on the p-values, which can be obtained (see Mehrotra *et al.*, 2003) as follows

$$P_{H_0}(X_1 = x_1 / M = m) = \frac{\binom{n_1}{x_1}\binom{n - n_1}{m - x_1}}{\sum_{j \in G} \binom{n_1}{j}\binom{n - n_1}{m - j}} \tag{7}$$

where

$$G = \{ j : \max(0, m - n_0) \le j \le \min(m, n_1) \}$$

For small samples when the hypothesis H_0 is assumed to be true the rejection rate of H_0 using Pearson chi-square test can be to high. The Fisher exact test is too conservative especially for small samples. Little (1989) presents a table which shows the percent rejection rates under H_0 for nominal $\alpha = 0{,}05$ under the assumption that H_0 is true (see Table 2). In his paper only conditions on the (n_0, n_1) margin fixed by the sample design are imposed. Analyzing results from this table we can conclude that Fisher exact test is much too conservative in small samples.

Table 2

Percent rejection rates under H_0 of Pearson and Fisher tests, for nominal $\alpha = 0.05$

$n_1 = n_2$	$p_0 = p_1 = 0.5$		$p_0 = p_1 = 0.2, p_0 = p_1 = 0.8$	
	Pearson	Fisher	Pearson	Fisher
5	5.5	1.1	2.2	2.0
10	5.8	2.1	4.6	1.5
20	4.2	2.1	5.1	2.3
50	4.5	2.9	5.0	2.9
100	5.2	3.8	4.9	3.5
200	4.9	4.0	---	---
Infinity	5	5	5	5

Source: Little (1989).

From this table it can be clearly seen that under the assumption that H_0 is true, both the Pearson chi-square test and Fisher exact test have the rejection rates under H_0 close to α only for large samples. For small samples the rates are considerably larger than the nominal α, especially for the exact test. Agresti and Coull (1998) discuss in their paper a very similar problem which appears during construction of confidence interval on proportion and they suggest using the approximation of distributions rather than exact distributions. They explain that such solutions have very good properties in many situations including cases when we deal with small samples. Sometimes it is better to use Boschloo exact test instead of Fisher exact test. This test is based on p-values for Fisher exact test. We calculate an exact two-sided p-value for this test as

$$p_B(m_0, m_1) = \sup_{\theta \in [0,1]} \left\{ P_{H_0} \{ p_F(X_1, X_2) \le p_F(x_1, x_2) \} \mid \theta \right\} \tag{8}$$

This test is not as conservative as the Fisher exact test, but it is difficult to find p-values even using modern computer techniques. As the tests we mentioned in this paper have many faults, such as the assumption of quite high expected counts in Table 1, the conservatism of exact test or difficult calculations, we are to look for new solutions which would allow us to verify the hypothesis (2) and would be much easier to use in practice. To verify the hypothesis (2) we can also use the chi-squared test with either Yates of Dandekar continuity correction. The Yates continuity correction can be written as follows

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(\left|O_{ij}-E_{ij}\right|-0,5\right)^2}{E_{ij}} \tag{9}$$

and the Dandekar correction (see R a o, 1982) can be written

$$\chi_c^2 = \chi_0^2 - \frac{\chi_0^2-\chi_{-1}^2}{\chi_1^2-\chi_{-1}^2}(\chi_1^2-\chi_0^2) \tag{10}$$

where $\chi_0^2, \chi_1^2, \chi_{-1}^2$ denote the statistics which is derived using (6) but after adding 0, 1 or –1 to the least values in the Table 1 with constant unconditional values.

4. Multivariate tests

All the above mentioned tests for the equality of proportions can be generalized to multidimensional analyze. Let assume that k properties of an object are assessed alternatively in two periods. In practical approach of quality control the tests for comparing the response proportions are usually used for each property independently to verify the hypothesis (2). L e v i n s o n (2004) in his paper presents a proposition of such a test. It is also possible to use k Fisher exact tests simultaneously, but as we mentioned earlier, L i t t l e (1989) and A g r e s t i and C o u l l (1998) note in their papers that these tests are very conservative.

When we monitor k variables simultaneously and we use the Fisher exact test to verify the hypothesis H_0 of equality of proportions for each of the variables separately, we reject the hypothesis for fixed nominal α when the rejection rate for any variable in the exact test equals to $1-\sqrt[k]{1-\alpha}$ .The rejection rates under H_0 assuming H_0 to be true for the simultaneous use of 1, 3 and 5 exact tests are presented in Table 3.

Table 3

Percent rejection rates under H_0 of Fisher's test ($k = 1, 3, 5$) and nominal $\alpha = 0.05$

$n_1 = n_2$	$p = 0.5$			$p = 0.2, p = 0.8$		
	$k = 1$	$k = 3$	$k = 5$	$k = 1$	$k = 3$	$k = 5$
5	1.1	3.2	0.5	2.0	0.7	1.0
10	2.1	1.8	2.9	1.5	1.6	2.7
20	2.1	1.4	2.4	2.3	4.3	2.4
50	2.9	3.1	3.0	2.9	2.6	4.2

Source: own study.

The results presented in Table 3 point out the problem of even more conservatism of using k Fisher exact tests simultaneously for each of the k variables than in one dimensional analyze. The results obtained from the analysis of k variables can be presented in contingency tables. Table 4 presents the results for i-th variable ($i = 1, 2, \ldots, k$).

Table 4

The results of analyze of i-th variable

Outcome	Period		Total
	base	present	
S	m_{0i}	m_{1i}	$m_i = m_{0i}+m_{1i}$
F	$n_0 - m_0$	$n_1 - m_{1i}$	$n - m_i$
Total	n_0	n_1	$n = n_0+n_1$

Source: own study.

As the Fisher exact test is that conservative, when veryfying the hypothesis (3) for k variables, we are to find alternative solutions of verification of this hypothesis. One of the possible solutions is a generalization of the above mentioned one dimensional tests. We are presenting now three statistics, which can be used for verification of the hypothesis (2).

- The first of the proposed statistics T_1 is the chi-squared test for the equality of proportions expanded for k proportions

$$T_1 = \sum_{s=1}^{k} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij}^{(s)} - E_{ij}^{(s)})^2}{E_{ij}^{(s)}} \tag{11}$$

where:

$O_{ij}^{(s)}$ – observed frequencies for s-th variable in row i and column j,

$E_{ij}^{(s)}$ – expected frequency for s-th variable for i-th row and j-th column, $s = 1, 2, \ldots, k$.

- Statistics T_2 is the chi-squared test with Yates continuity correction expanded for testing equality of k variables

$$T_2 = \sum_{s=1}^{k}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(|O_{ij}^{(s)} - E_{ij}^{(s)}| - 0,5)^2}{E_{ij}^{(s)}} \tag{12}$$

- Statistics T_3 is chi-squared test with Dandekar continuity correction expanded for testing equality of proportions of k variables

$$T_3 = \sum_{s=1}^{k}\chi_{c(s)}^2 \tag{13}$$

where $\chi_{c(s)}^2$ calculated on the basis of (6) for s-th variable.

All the statistics have the asymptotic chi-square distribution with k degree of freedom. We are the most interested in deciding which of these three statistics can be used in situations when the expected values are small. We will present the results of computer simulations, which allow us to compare the properties of these statistics T_1, T_2, T_3 assuming that the expected values are small.

5. Monte Carlo study

The three proposed statistics T_1, T_2 and T_3 have the asymptotic chi-square distribution with k degrees of freedom. When we deal with small samples there can occur significance differences in results of these three statistics and it can lead to differences in final conclusions in terms of the hypothesis H_0. In the computer simulations a situation was analyzed in which these three above mentioned statistics were used for testing the stability of three proportions ($k = 3$).

Under the assumption that hypothesis (2) of equality of proportions in two periods is true the probabilities of successes were equal. In the next stage of simulations we decided the probabilities of success in the present period to be 10%, 20%, 30%, 40% and 50% respectively higher than in the base period. In first series of simulations (A) these changes were applied to all variables, to first two variables in second series of simulations (B) and to the first variable only in third series of simulations (C). The sample from the base period was of a size 900 elements and the sample from the present period was of a size 100 elements. The hypothesis of the stability of three proportions was verified using three statistics T_1 – based on chi-squared statistics, T_2 – the chi-squared statistics with Yates continuity correction and with Dandekar continuity correction (T_3). The simulation process was repeated 10 000 times. The statistics T_1, T_2 and T_3 were calculated on the basis of the results obtained from the simulations and then the decisions about either rejection or not rejection the hypothesis (2) were taken. The percent rejection rates under H_0 received in the simulation process are presented in Table 5.

Table 5

The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variable p_{11}, p_{12}, p_{13})

Statystic	The increase of p_1 in % in thf period t_1					
	0	10	20	30	40	50
T_1 (Pearson)	0.0485	0.0763	0.1076	0.1574	0.2170	0.2946
T_2 (Yates)	0.0166	0.0321	0.0526	0.0801	0.1273	0.1775
T_3 (Dandekar)	0.0402	0.1040	0.1429	0.2051	0.2730	0.3642

Source: own study.

Analyzing the results in Table 5 we can notice that the test based on T_2 statistics with Yates continuity correction is very conservative. The best results were achieved for the statistics T_3, which uses the chi-squared statistics with Dandekar continuity correction. In this case the percent rejection rate under H_0, under the assumption that H_0 was false, was the highest. The results of the simulations are also presented in Figure 2.

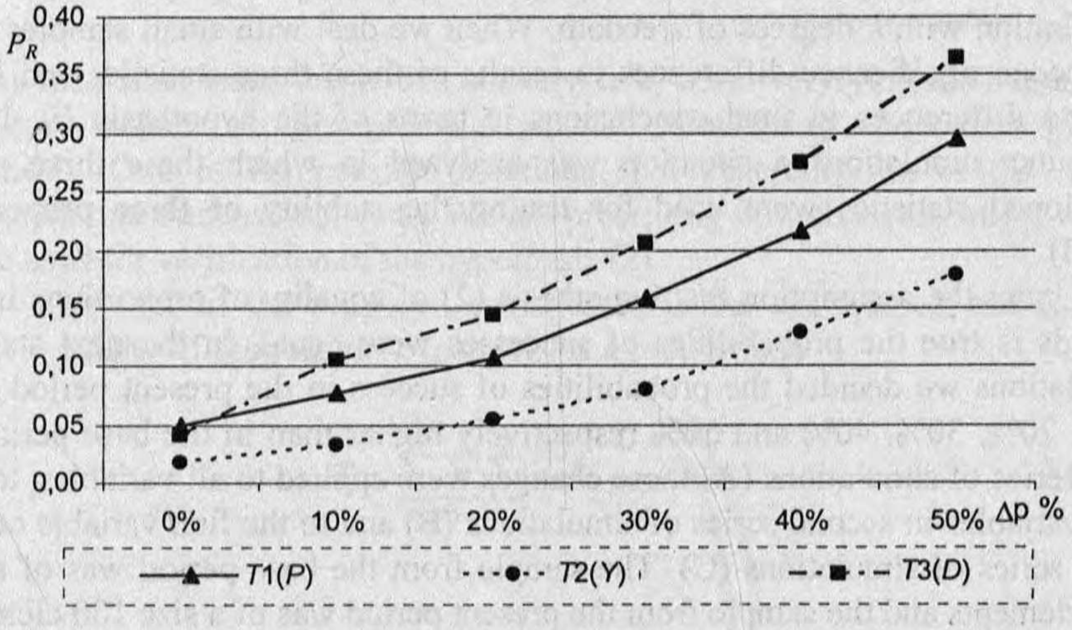

Fig. 2. The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variables p_{11}, p_{12}, p_{13})

In the second stage of the simulations it was assumed that the probabilities of successes of the first two variables were increasing and the probability of success of the third variable was constant. The results of the simulations under these assumptions are presented in Table 6. Analyzing the results from the Table 6 it can be seen that the best results were achieved for the statistics T_3 also in this case. Under the assumption that H_0 is true the probabilities of rejection the hypothesis H_0 are close to the nominal $\alpha = 0.05$ only for statistics T_1.

Table 6

The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variables p_{11}, p_{12})

Statystic	The increase of p_1 in % in the period t_1					
	0	10	20	30	40	50
T_1 (Pearson)	0.0485	0.0705	0.0870	0.1081	0.1442	0.1851
T_2 (Yates)	0.0166	0.0297	0.0371	0.0514	0.0748	0.0999
T_3 (Dandekar)	0.0402	0.0947	0.1189	0.1467	0.1874	0.2363

Source: own study.

In the third stage of simulations it was assumed that the probabilities of successes of second and third variables were constant in the base and the present period and that the probability of success of the first variable was increasing. The results of the simulations under these assumptions are presented in Table 7.

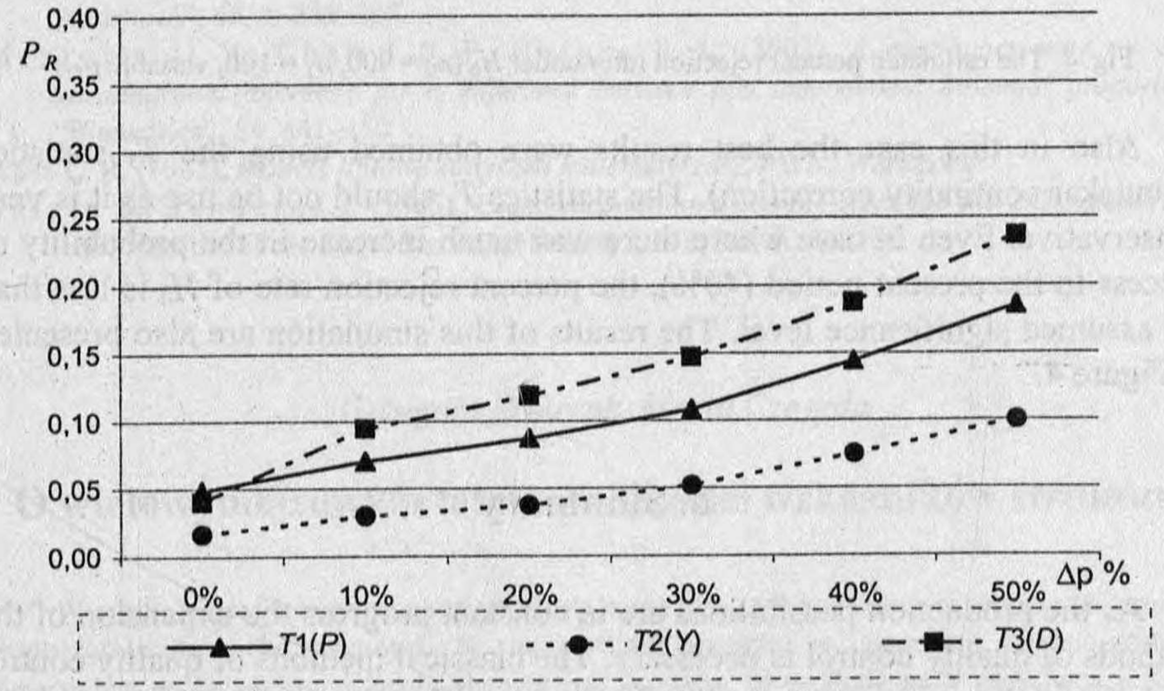

Fig. 3. The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variables p_{11}, p_{12})

Table 7

The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variable p_{11})

Statystic	The increase of p_1 in % in the period t_1					
	0	10	20	30	40	50
T_1 (Pearson)	0.0485	0.0556	0.0672	0.075	0.0879	0.1118
T_2 (Yates)	0.0166	0.0224	0.0251	0.0330	0.0382	0.0526
T_3 (Dandekar)	0.0402	0.0779	0.0908	0.0984	0.1188	0.1450

Source: own study.

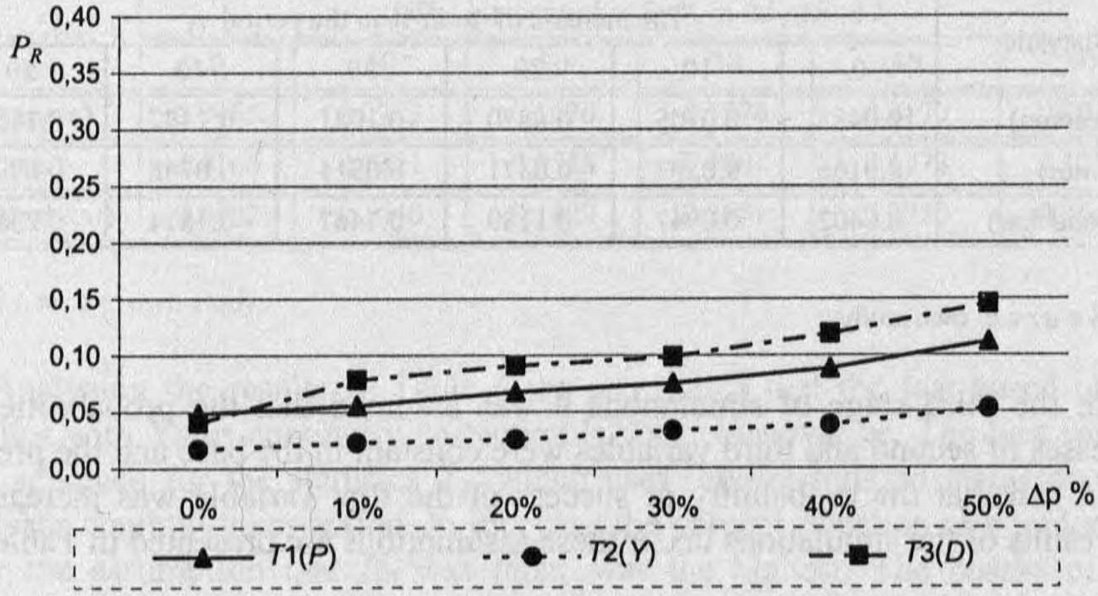

Fig. 4. The estimated percent rejection rates under H_0 ($n_0 = 900$, $n_1 = 100$, variable p_{11})

Also in this case the best results were obtained using the T_3 statistics (Dandekar continuity correction). The statistics T_2 should not be use as it is very conservative. Even in case where there was much increase in the probability of success in the present period (40%), the percent rejection rate of H_0 is less than the assumed significance level. The results of this simulation are also presented in Figure 4.

6. Summary

As the production possibilities are in constant progress the expansion of the methods of quality control is necessary. The classical methods of quality control used in an alternative control allowed for classification the elements as either good or bad. Nowadays during the process of quality control we are able to obtain very detailed information about single element and its defectives. This information may be very useful in the quality control process. It requires though finding new methods, which could be successfully applied in such cases.

In this paper three tests are introduced which can be used in situations when multiple attributes are assessed simultaneously. The expansion of chi-square test, chi-squared test with Yates continuity correction and chi-squared test with Dandekar continuity correction results in the proposed tests. The analyze and comparisons of these tests with one dimensional tests and the exact tests point out that the application of these solutions in quality control can be the source of many benefits. The application of chi-squared test with Dandekar continuity correction seems to be especially interesting. This test allows us to control simultaneously multiple attributes, which are assessed alternatively. The advantages of this test are that it can be applied also in situations when we deal with small samples and it is not as conservative as the other exact tests. The disadvantage of this test is that when deciding about rejection of the hypothesis H_0 it is not easy to find out which of the controlled attributes causes the problem.

References

Agresti A., Coull B. A. (1998), *Approximate is better than "exact" for interval estimation of binomial proportions*, "The American Statistican", **52**, 2, 119–126.

Levinson W. A. (2004), *Control charts control multiple attributes*, "Quality Software & Analysis", **43**, iss. 9, 40–43.

Little (1989), *Testing the equality of two independent binomial proportions*, "The American Statistican", **43**, 4, 283–288.

Mehrotra D. V., Chan I. S. F., Berger R. L. (2003), *A cautionary note on exact unconditional inference for a difference between two independent binomial proportions*, "Biometrics", **59**, 441–450.

Rao C. R. (1982), *Modele liniowe statystyki matematycznej*, PWN, Warszawa.

Suissa S., Schuster J. (1985), *Exact unconditional sample sizes for the 2x2 comparative trial*, "Journal of the Royal Statistical Society", **148**, Ser. A, 86–105.

Grzegorz Kończak, Maria Czogała

O wielowymiarowym teście stabilności wskaźników struktury

W artykule analizowano zagadnienie testowania hipotezy o stabilności wskaźnika struktury jednocześnie dla wielu zmiennych. W rozważaniach przyjęto, że z populacji pobrana została n elementowa próba, w której każdy element jest oceniany ze względu na *k* właściwości. Każda właściwość jest oceniana alternatywnie. Rozważany jest problem weryfikacji hipotezy głoszącej

niezmienność w czasie frakcji wyróżnionych elementów dla każdej zmiennej. Przedstawiono możliwości zastosowania klasycznych testów chi kwadrat Pearsona, testu chi kwadrat z poprawką Yatesa oraz dokładnego testu Fishera dla weryfikacji hipotezy o równości wskaźników struktury. W artykule przedstawiono propozycję testu pozwalającego na weryfikację hipotezy o jednoczesnej zgodności z założeniami wielu wskaźników struktury. Porównano własności proponowanego testu z klasycznymi rozwiązaniami.

Problem przedstawiony w artykule jest spotykany w zagadnieniach statystycznej kontroli jakości, gdy elementy są sprawdzane alternatywnie (dobry lub zły) pod kątem zgodności z szeregiem wymogów. Produkty często sprawdzane są pod względem zgodności z normami wielu charakterystyk ocenianych alternatywnie (kolor, zarysowania, docisk itd.). Jednoczesna ocena wielu właściwości ma szczególne znaczenie w obecnych czasach, gdy wobec produkowanych wyrobów są stawiane coraz to wyższe wymagania jakościowe. W artykule zwrócono uwagę na korzyści zastosowania proponowanego testu zamiast wielu testów dla wskaźników struktury dla każdej ocenianej właściwości z osobna.