ZASTOSOWANIA MATEMATYKI
V (1960)

W. SADOWSKI (Warszawa)

A NON-PARAMETRIC TEST OF COMPARING DISPERSIONS (1)

1. Let us consider k continuous populations differing at most in their variances; the distributions of those populations, and in particular the variances σ_i^2 $(i = 1, 2, \ldots, k)$ are unknown to us.

This paper contains a *significance test* which makes it possible to verify the hypothesis that one of those populations has a greater variance than the remaining ones, i. e. that there exists a j such that

$$\sigma_j^2 > \max(\sigma_1^2, \ldots, \sigma_{j-1}^2, \sigma_{j+1}^2, \ldots, \sigma_k^2).$$

The proposed test is *non-parametric*, i. e. for its construction it is not necessary to know the distribution functions of the populations concerned. This is undoubtedly an advantage of the test. Another advantage is its simplicity, which is particularly important for applications. Its drawback, however, is its comparatively low power. Such is the price which is usually paid for the simplicity of a test and the avoidance of the assumption concerning the distribution of the general population.

As an example of the application of the test in question let us take the testing of the accuracy of measuring instruments. The test enables us to indicate the instrument with the least accuracy or to show that all the instruments in question have the same accuracy. The test can also be used to examine the accuracy of several precision machines or in other, similar cases.

The idea of the above test was suggested by Mosteller's paper [1]. Mosteller has constructed a non-parametric test to decide which of k

(1) Niniejsza praca była ogłoszona w naszym piśmie po polsku w tomie 2 (1955), str. 161-171. Obecnie ogłaszamy ją po angielsku, aby jej treść uczynić dostępniejszą obcym czytelnikom. *Redakcja.*

Данная работа была опубликована в нашем журнале на польском языке в томе 2 (1955), стр. 161-171. В настоящее время публикуем её на английском языке для того, чтобы содержание работы сделать даступным читателям, не знакомым с польском языком. *Редакция.*

This paper appeared in our periodical in Polish in vol. 2 (1955), pp. 161-171. We are now publishing it in English in order to make it easier for foreign readers to get acquainted with its contents. *Editors.*

populations has the greatest mean, under the assumption that the populations are continuous and differ at most in their means.

Both Mosteller's test and the test given in the present paper are based on the method of permuting the elements of a sample, called also the *randomization method*. That method is very often used in constructing non-parametric tests. We shall briefly discuss it in chapter 2 (²).

2. We are given N random variables $X_1, X_2, \ldots, X_N$ with a joint distribution function $F_N(x_1, x_2, \ldots, x_N)$. Let W denote the space of samples, i. e. the space of the points $E = (x_1, x_2, \ldots, x_N)$. Every statistical hypothesis that is usually verified can be written in the form $F_N \epsilon \omega$ where $\omega \epsilon \Omega$, Ω being the class of admissible distribution functions. A statistical hypothesis is thus a certain supposition concerning the shape of the distribution function F_N. The statistical test used for testing hypotheses is a rule according to which we decide on the basis of a sample whether the hypothesis in question is to be rejected or accepted. The construction of the test consists in finding a region $w \subset W$ such that if the random point $\boldsymbol{E} = (X_1, X_2, \ldots, X_N)$ is found in region w the hypothesis is rejected and in the opposite case it is accepted. According to the Neyman-Pearson's theory—regarded today as classical—we choose the region w in such a way that for a number α $(0 < \alpha < 1)$, selected beforehand, we have the inequality

$$\text{(1)} \qquad \mathrm{P}(\boldsymbol{E} \epsilon w | F_N) \leqslant \alpha \quad \text{for every} \quad F_N \epsilon \omega.$$

A region w satisfying the above relation is called a *region similar* to the sample space. The letter α denotes here the probability of committing an error consisting in rejecting the hypothesis $F_N \epsilon \omega$ which is being tested when it is the true one. Since in most cases there exist infinitely many regions satisfying condition (1) we choose that one of those regions for which

$$\text{(2)} \qquad \mathrm{P}(\boldsymbol{E} \epsilon w | F_N) = \max \quad \text{for every} \quad F_N \epsilon (\Omega - \omega).$$

In this way the region w ensures the greatest probability of rejecting the hypothesis which is being tested when it is actually false.

In the case of parametric hypotheses, i. e. when the distribution functions belonging to the set of admissible hypotheses Ω differ only in their parameters, it is often possible—with the use of suitable propo-

---

(²) We shall discuss it only in outline. More detailed information can be found above all in the works of R. A. Fisher [2], pp. 96-99 and [3], pp. 43-47, who has himself created the method in question. Interesting generalizations and a precise formulation of the randomization method can be found in H. Scheffé [4] and [5] and E. Lehman and C. Stein [6].

sitions—to construct a region w satisfying conditions (1) and (2). For non-parametric hypotheses, however, i. e. when any distribution functions can be elements of the set Ω, we not only are unable to construct effectively regions w satisfying the above two conditions but we also ignore whether such regions exist. However, there are methods permitting construction of regions similar to the sample space, i. e. regions w satisfying condition (1). One of the most important of those methods is the randomization method mentioned in chapter 1.

Denote by S the set of those permutations of the coordinates x_1, $x_2, \ldots, x_N$ in the sample space W for which the values of the individual distribution functions $F_N \epsilon \omega$ do not change. Let s denote the number of permutations belonging to set S. We establish a correspondence between each point E of space W and the set (E') comprising s points obtained by all the permutations from set S performed on the coordinates of point E.

We assign to the region w constructed by the randomization method q points ($q < s$) picked out from each set (E') corresponding to all points E. Regions constructed by the randomization method prove to be similar regions, i. e., they satisfy condition (1), and we have $\mathrm{P}(\boldsymbol{E} \epsilon w | F_N) = q/s$. It can be shown, under fairly weak assumptions, that the randomization method is the only method permitting the construction of similar regions in the case of testing non-parametric hypotheses. It can also be shown that similar regions exist only if class Ω contains exclusively continuous distribution functions ([3]). That is why in most non-parametric tests we must assume the continuity of the general population.

It is easy to observe that there can be a large number of similar regions constructed by the randomization method. The problem arises how to select one of them. In non-parametric cases it is very difficult to make use of condition (2) since it would then be necessary to consider a functional $\mathrm{P}(\boldsymbol{E} \epsilon w | F_N)$ defined for $F_N \epsilon (\Omega - \omega)$. It has been usual so far in practice to choose a similar region from among the regions satisfying condition (1) by means of a suitably constructed statistics $T(X_1, X_2, \ldots$ $\ldots, X_N)$. We shall now give an example of using the randomization method and a suitable statistics T.

Example. From a population with an unknown continuous distribution $G(x, y)$ we take a sample (x_i, y_i) $(i = 1, 2, \ldots, m)$ consisting of m independent pairs. On this basis we are to check the hypothesis that the variables X and Y are independent.

The class of all continuous (two-dimensional) distribution functions

---

([3]) All those theorems in precise form can be found in the works of H. Scheffé [4] and [5] and E. Lehman and C. Stein [6], which have already been mentioned.

constitutes here the class Ω of admissible distribution functions, while the subset ω contains all distribution functions F_N of the form

$$F_N = \prod_{i=1}^{m} J(x_i) \prod_{j=1}^{m} K(y_j), \tag{3}$$

where J and K are arbitrary continuous distribution functions and $N = 2m$.

The set S contains in this case those permutations performed on the coordinates $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_m$ which do not change the value of any of the distribution functions $F_N \epsilon \omega$, i. e. any of the distribution functions of form (3). Now, F_N does not change its value if the coordinates x_i are permuted with one another and the coordinates y_j are permuted separately; however, joint permutation of the coordinates x_i and y_j may affect the value of F_N. Thus the set S consists in our example of $(m!)^2$ elements. Consequently there is a correspondence between each point E of the $2m$-dimensional sample space W and the set (E') consisting of $(m!)^2$ points with coordinates determined by the elements of the set S. Choosing q points from each of the sets (E') we obtain a similar region w for which

$$\mathrm{P}(\boldsymbol{E} \epsilon w | F_N) = q/(m!)^2 \quad \text{for every} \quad F_N \epsilon \omega.$$

We decide which q points of each set (E') should be assigned to the region w by means of a suitably chosen statistics T. In our case it can be the following statistics:

$$T(E) = \frac{\left|\sum_{i=1}^{m} x_i y_i\right|}{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i^2}.$$

In accordance with intuition, we shall assign to region w those q points of each set (E') for which $T(E)$ assumes values greater than for the remaining $s-q$ points.

For instance, suppose that the sample is following: $x_1 = 2$, $x_2 = 3$, $x_3 = 5$, $y_1 = 1$, $y_2 = 4$, $y_3 = 8$. The set S contains here $(3!)^2 = 36$ elements. We are to find on the basis of a sample 36 possible values which can be assumed by $T(E)$. It can easily be verified that in 6 cases $T(E) = 54$, in the next 6 cases $T(E) = 51$, and further, each time for 6 cases, we have $T(E) = 46$, $T(E) = 39$, $T(E) = 37$, $T(E) = 33$.

Suppose that $\alpha = 0{,}2$; then, assigning the first six points to the critical region we have

$$\mathrm{P}(\boldsymbol{E} \epsilon w | F_N) = \tfrac{6}{36} = 0{,}166\ldots$$

Assigning 12 points we obtain

$$\mathrm{P}(\boldsymbol{E} \epsilon w | F_N) = \tfrac{12}{36} = 0{,}333\ldots$$

Thus for $\alpha = 0{,}2$ we must assign to region w the 6 points giving the greatest value of $T(E)$. Since in our example $T(E) = 54$, the independence hypothesis must be rejected.

3. We shall now give the construction of the test mentioned in chapter 1. Suppose that the samples from all k populations have the same size and contain n elements each (4). From among k samples we select one containing the greatest element of kn observations and at the same time the smallest element. In the sample thus selected we establish the joint number of elements greater and smaller than the elements in the remaining $k-1$ samples. Those elements will be termed *protruding elements*. Their number in a chosen sample (if such a sample exists) will be denoted by r. We fix a certain number r_0 in such a way that if $r \geqslant r_0$, we reject the hypothesis of equality of variances and assume that the sample containing the greatest and the smallest element, i. e. the selected sample, comes from the population with the greatest variance. Otherwise, if $r < r_0$ or if there is no selected sample, we assume that the variances of all the populations are equal (5).

The construction of the test consists in using the randomization method, which appears here in a particularly simplified form. It follows from the fact that statistics r which we use here does not depend directly on the magnitude of the individual observations but on the order of the indices of those observations when ordered according to their increasing (or decreasing) magnitudes, e. g.

$$x_{a_1} < x_{a_2} < \ldots < x_{a_{kn}}, \tag{4}$$

where $(a_1, a_2, \ldots, a_{kn})$ is a certain permutation of numbers $1, 2, \ldots, kn$. Of course, all points satisfying condition (4) give the same value of the statistics r and thus it is not necessary to distinguish them. The probability of obtaining any point belonging to the region determined by (4)(6) is equal to $1/s$ (where s is the number of permutations of the indices a which do not change the value of F_N), which immediately follows from the general considerations of chapter 2. It should be observed, moreover, that a change of the order of indices within the individual k samples, i. e. in the groups

$$(a_1, a_2, \ldots, a_n),\ (a_{n+1}, a_{n+2}, \ldots, a_{2n}),\ \ldots,\ (a_{(k-1)n+1}, a_{(k-1)n+2}, \ldots, a_{kn}),$$

(4) With samples of different size the test would involve more complicated calculations.

(5) After this paper had been handed in for publication, there appeared a paper by S. Rosenbaum [7], in which the author gives an analogically constructed test for the case of two populations ($k = 2$), but does not assume the same number of elements in the samples.

(6) Itsi thus the probability of a point giving a definite value of the statistics r.

has likewise no effect upon the size of the statistics r. In our case therefore it is worth while only to consider the number of those permutations of the indices which can give different values of r. It is easy to show that the number of those permutations is $(kn)!/(n!)^k$.

Our task is to find the probability of the number of protruding elements in the selected sample being equal or greater than a given number i ($i = 2, 3, \dots, n$).

The calculation must be made under the assumption that the zero hypothesis ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$) is true. We must namely find the number of those permutations from among the $(kn)!/(n!)^k$ for which $r \geqslant i$.

Let us begin with the easiest case—of $i = n$. The composition of the chosen sample can be shown diagrammatically as follows:

Variants		1	2	3	...	$n-3$	$n-2$	$n-1$
Number of elements in the selected sample	greater than elements in other samples	1	2	3	...	$n-3$	$n-2$	$n-1$
	smaller than elements in other samples	$n-1$	$n-2$	$n-3$	...	3	2	1
Total number of protruding elements		n	n	n	...	n	n	n

Each variant of this kind can be connected with

$$\frac{[(k-1)n]!}{(n!)^{k-1}0!}$$

permutations arising from the permutation of the remaining elements. Thus the total number $L(n)$ of permutations in which the selected sample has n protruding elements is

$$L(n) = (n-1)\frac{[(k-1)n]!}{(n!)^{k-1}0!}.$$

Let us now find the number of those permutations in which the number of protruding elements in the selected sample is $n-1$ or more. The composition of the selected sample containing $n-1$ protruding elements may be one of the following variants:

Variants		1	2	3	...	$n-3$	$n-2$
Number of elements in the selected sample	greater than elements in other samples	1	2	3	...	$n-3$	$n-2$
	smaller than elements in other samples	$n-2$	$n-3$	$n-4$	...	2	1
Total number of protruding elements		$n-1$	$n-1$	$n-1$	...	$n-1$	$n-1$

Each variant of this kind can be connected with

$$\frac{[(k-1)n+1]!}{(n!)^{k-1}1!}$$

permutations formed from the remaining elements. It should be observed here that a sample with composition $(1, n-2)$ can give rise by a suitable change of one element to samples with composition $(1, n-1)$ or $(2, n-2)$. A sample with composition $(2, n-3)$ can give rise to samples with composition $(2, n-2)$ or $(3, n-3)$, etc. Consequently we obtain

$$L(n-1) = (n-2)\frac{[(k-1)n+1]!}{(n!)^{k-1}1!} - (n-1-2)\frac{[(k-1)n]!}{(n!)^{k-1}0!}.$$

Following an analogical argument we can find the number of permutations in which the number of protruding elements in the selected sample is $n-2$ or more:

$$L(n-2) = (n-3)\frac{[(k-1)n+2]!}{(n!)^{k-1}2!} - (n-2-2)\frac{[(k-1)n+1]!}{(n!)^{k-1}1!}.$$

It is now easy to show that

$$L(n-u) = (n-u-1)\frac{[(k-1)n+u]!}{(n!)^{k-1}u!} - (n-u-2)\frac{[(k-1)n+u-1]!}{(n!)^{k-1}(u-1)!}$$

for $1 \leqslant u \leqslant n-2$. Writing $n-u = i$ we eventually obtain

$$L(i) = (i-1)\frac{(kn-i)!}{(n!)^{k-1}(n-i)!} - (i-2)\frac{(kn-i-1)!}{(n!)^{k-1}(n-i-1)!} \quad \text{for} \quad i = 2, 3, \ldots, n-1, \tag{5}$$

$$L(i) = (i-1)\frac{(kn-i)!}{(n!)^{k-1}(n-i)!} \quad \text{for} \quad i = n. \tag{5'}$$

The formulas obtained permit us to find the number of those permutations among $(kn)!/(n!)^k$ for which in a given sample (i. e. in a sample coming from a given population) $r \geqslant i$. Consequently the probability that in a given sample $r \geqslant i$ is

$$\mathrm{P}(r \geqslant i) = \frac{L(i)(n!)^k}{(kn)!}.$$

We are not interested, however, in the probability of $r \geqslant i$ in a given sample (or a sample taken at random) but in the probability of $r \geqslant i$ in a selected sample. Thus the probability desired is

$$\mathrm{P}(r \geqslant i) = k\frac{L(i)(n!)^k}{(kn)!}.$$

Substituting (5) or (5′) in the last expression and simplifying, we obtain

$$\text{(6)}\qquad P(r \geqslant i) = (i-1)k\frac{\binom{kn-i}{n-i}}{\binom{kn}{n}} - (i-2)k\frac{\binom{kn-i-1}{n-i-1}}{\binom{kn}{n}}$$

for $i = 2, 3, \ldots, n-1$;

$$\text{(6′)}\qquad P(r \geqslant i) = (i-1)k\frac{\binom{kn-i}{n-i}}{\binom{kn}{n}} \quad \text{for} \quad i = n.$$

It is worth observing here that for constant i and k and $n \to \infty$ the above formula is considerably simplified. Using Sterling's formula for $n!$ we can show that

$$\text{(7)}\qquad \lim_{n\to\infty} P(r \geqslant i) = \frac{1}{k^{i-1}}\left(i-1-\frac{i-2}{k}\right).$$

4. On the basis of formulas (6), (6′) and (7) it is possible to construct tables facilitating the application of the proposed test. We give below a few tables for different values of k, n and i. These tables give the probabilities that for given k and n the selected sample will have i or more protruding elements.

EXAMPLE. The same object has been measured by three micrometers, five times by each. It is to be decided whether all three micrometers are equally accurate or whether one (and which one) is less accurate than the remaining ones. The results of the measurements of each micrometer have been ordered and are the following:

I micrometer	II micrometer	III micrometer
4,077	4,070	4,069
4,078	4,079	4,071
4,082	4,080	4,075
4,084	4,081	4,083
4,085	4,086	4,087

We apply the test, for instance, at the significance level 0,05. The selected sample are the measurement results obtained by means of the third micrometer. The value of the statistics r, i. e. the number of protruding elements in the selected sample is 2. We then use the table for $k = 3$ and find for $n = 5$ and $i = 2$ the probability 0,2857, i. e. $P(r \geqslant 2) =$ $= 0,2857$. Thus there are no grounds for asserting that the third micrometer is less accurate than the remaining two. The critical value of r_0 is 4, which can easily be verified in the tables.

TABLES

Values of probabilities $P(r \geqslant i)$ depending on the number of populations k and sizes n

$k = 2$

n \ i	2	3	4	5	6
2	0,3333				
3	,4000	0,2000			
4	,4286	,2572	0,0857		
5	,4444	,2857	,1270	0,0317	
6	,4545	,3030	,1515	,0541	0,0108
7	,4615	,3147	,1678	,0699	,0210
8	,4667	,3231	,1795	,0816	,0294
9	,4706	,3294	,1882	,0905	,0362
10	,4737	,3344	,1950	,0975	,0418
11	,4762	,3383	,2005	,1032	,0464
12	,4783	,3416	,2050	,1079	,0504
13	,4800	,3444	,2087	,1118	,0537
14	,4815	,3467	,2119	,1162	,0565
15	,4828	,3487	,2146	,1180	,0590
16	,4839	,3504	,2169	,1205	,0612
17	,4848	,3519	,2190	,1227	,0631
18	,4857	,3532	,2208	,1246	,0648
19	,4865	,3544	,2224	,1264	,0663
20	,4872	,3555	,2238	,1279	,0677
21	,4878	,3565	,2252	,1293	,0690
22	,4884	,3574	,2263	,1306	,0701
23	,4889	,3582	,2274	,1318	,0711
24	,4894	,3589	,2284	,1328	,0721
25	,4898	,3595	,2293	,1337	,0730
26	,4902	,3602	,2301	,1346	,0738
27	,4906	,3607	,2309	,1355	,0745
28	,4909	,3612	,2315	,1362	,0752
29	,4912	,3617	,2322	,1369	,0758
30	,4915	,3622	,2328	,1376	,0764
∞	0,5000	0,3750	0,2500	0,1563	0,0938

$k = 3$

n \ i	2	3	4	5	6
2	0,2000				
3	,2500	0,0714			
4	,2727	,1030	0,0182		
5	,2857	,1209	,0310	0,0050	
6	,2941	,1324	,0399	,0079	0,0008
7	,3000	,1404	,0464	,0112	,0018
8	,3044	,1462	,0514	,0139	,0028
9	,3077	,1508	,0553	,0162	,0038
10	,3104	,1544	,0584	,0180	,0046
11	,3125	,1573	,0609	,0196	,0053
12	,3143	,1597	,0630	,0209	,0060
13	,3158	,1617	,0648	,0221	,0066
14	,3171	,1634	,0664	,0231	,0071
15	,3182	,1650	,0677	,0240	,0075
16	,3191	,1662	,0689	,0247	,0079
17	,3200	,1674	,0699	,0254	,0083
18	,3208	,1684	,0708	,0260	,0086
19	,3214	,1693	,0717	,0266	,0089
20	,3220	,1701	,0724	,0271	,0092
∞	0,3333	0,1852	0,0864	0,0370	0,0151

$k = 4$

n \ i	2	3	4	5	6
2	0,1429				
3	,1818	0,0364			
4	,2000	,0550	0,0066		
5	,2105	,0661	,0119	0,0010	
6	,2174	,0734	,0158	,0022	0,0001
7	,2222	,0786	,0188	,0032	,0004
8	,2258	,0825	,0211	,0041	,0006
9	,2286	,0856	,0230	,0048	,0008
10	,2308	,0880	,0245	,0055	,0010
11	,2326	,0899	,0258	,0060	,0012
12	,2340	,0916	,0268	,0065	,0013
13	,2353	,0930	,0277	,0069	,0015
14	,2364	,0942	,0285	,0073	,0016
15	,2373	,0952	,0292	,0076	,0017
∞	0,2500	0,1194	0,0391	0,0127	0,0039

Bibliography

[1] F. Mosteller, *A k-sample slippage test for an extreme population*, Ann. of Math. Statist. 19 (1948), pp. 58-65.

[2] R. A. Fisher, *Statistical methods for research workers*, 10th ed., London 1948.

[3] — *The design of experiments*, 5th ed., London 1948.

[4] H. Scheffé, *Statistical inference in the non-parametric case*, Ann. of Math. Statist. 14 (1943), pp. 305-332.

[5] — *On a measure problem arising in the theory of non-parametric tests*, Ann. of Math. Statist. 14 (1943), pp. 227-233.

[6] E. L. Lehman and C. Stein, *On the theory of some non-parametric hypotheses*, Ann. of Math. Statist. 20 (1949), pp. 28-45.

[7] S. Rosenbaum, *Tables for a non-parametric test of dispersion*, Ann. of Math. Statist. 24 (1953), pp. 663-668.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK
MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES

(*The Polish version of this paper has been received on 16. 12. 1953*)