Marie Hušková
Simultaneous rank test procedures

_Aplikace matematiky_, Vol. 25 (1980), No. 1, 33–38

Persistent URL: [http://dml.cz/dmlcz/103835](http://dml.cz/dmlcz/103835)

Terms of use:

© Institute of Mathematics AS CR, 1980

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project _DML-CZ: The Czech Digital Mathematics Library_ [http://dml.cz](http://dml.cz)
SIMULTANEOUS RANK TEST PROCEDURES

MARIE HUŠKOVÁ

(Received February 15, 1978)

1. INTRODUCTION

Let $X_j = (X_{ij}, \ldots, X_{pj})^t$, $j = 1, \ldots, N$, be independent $p$-dimensional random variables with continuous distribution functions. Consider the hypotheses of randomness associated with some marginal distributions:

$$H_v : F_j^v(x^v) = F^v(x^v), \quad j = 1, \ldots, N, \quad v = 1, \ldots, r,$$

where $F_j^v(x^v)$ is the marginal distribution of the subvector $X^v$, $v = 1, \ldots, r$, $x^1, \ldots, x^r$ is a partition of the vector $x$, i.e., $x = (x^1, \ldots, x^r)^t$. We are interested in testing hypotheses $H_1, \ldots, H_r$ and $H_0 = \bigcap_{v=1}^r H_v$ against alternatives $A_1, \ldots, A_r$ and $A_0 = \bigcup_{v=1}^r A_v$, resp., where $A_v : F_j^v(x^v) = F^v(x^v; \theta_j^v)$, $j = 1, \ldots, N$, $v = 1, \ldots, r$, with $\theta_j = (\theta_j^1, \ldots, \theta_j^r)^t$ being a vector of unknown parameters.

Krishnaiah and some others (see [5]—[8]) developed several simultaneous test procedures for the classical multivariate normal theory. As for simultaneous rank test procedures, Krishnaiah and Sen [9] dealt with this problem for some MANOVA models, Jensen [3] for multivariate random blocks, Hušková [2] suggested a method for the problem considered in the present paper (see method I below).

Here we give three test procedures analogous to those proposed by Krishnaiah in [5—6] and based on the asymptotic distributions of quadratic rank statistics (for definition see (3) below).

Put

$$S_c = (S_{c1}, \ldots, S_{cp})^t,$$

$$(2) \quad S_{ci} = \sum_{j=1}^N (c_{ij} - \bar{c}_i) a_{N_i} R_{ij}, \quad i = 1, \ldots, p,$$

with $R_{ij}$ being the rank of $X_{ij}$ in the sequence $X_{i1}, \ldots, X_{IN}$, $c_{ij}$ regression constants, $a_{N_i}$ scores and $\bar{c}_i = N^{-1} \sum_{j=1}^N c_{ij}$. Denote by $S_c^v$ the subvector of $S_c$ corresponding to $X^v$, $v = 1, \ldots, r$. Define

$$Q_c = S_c^t (\text{var}_p S_c)^{-1} S_c,$$

$$Q_c^v = S_c^v (\text{var}_p S_c^v)^{-1} S_c^v, \quad v = 1, \ldots, r,$$

$$Q_c^v = S_c^v (\text{var}_p S_c^v)^{-1} S_c^v, \quad v = 1, \ldots, r.$$
where the matrix $\text{var}_p S_c$ is regular with elements

$$(N - 1)^{-1} \sum_{j=1}^{N} (c_{ij} - \bar{c}_i)(c_{ij} - \bar{c}_i) \sum_{m=1}^{N} (a_{ni}(R_{im}) - \bar{a}_{ni})(a_{ni}(R_{ij}) - \bar{a}_{ni})$$

if

$$i, t \in I_k, \quad k = 1, \ldots, r,$$

and

$$\sum_{j=1}^{N} (c_{ij} - \bar{c}_i)(c_{ij} - \bar{c}_i) (a_{ni}(R_{ij}) - \bar{a}_{ni})(a_{ni}(R_{ij}) - \bar{a}_{ni})$$

if

$$i \in I_k, \quad t \notin I_k, \quad k = 1, \ldots, r,$$

where $I_1, \ldots, I_r$ is the partition of the set $I = \{1, \ldots, p\}$ considered in hypotheses $H_v$ and $\text{var}_p S^v$ is the submatrix of $\text{var}_p S_c$ corresponding to $S^v$ and $\bar{a}_{ni} = N^{-1} \sum_{j=1}^{N} a_{ni}(j)$.

Denote by $m_v$ the number of components of $\mathbf{x}^v, \nu = 1, \ldots, r$.

We shall impose usual conditions on scores, regression constants and the matrix $\text{var}_p S_c$:

a. The scores $a_{ni}(j)$ are generated by a nonconstant square integrable functions $\varphi_i, \ i = 1, \ldots, p$, i.e.,

$$\int_0^1 (\varphi_i(u) - a_{ni}(\lceil uN \rceil + 1))^2 \, du \to 0 \quad \text{for} \ N \to \infty, \ i = 1, \ldots, p.$$

b. The regression constants fulfil:

$$\max_{1 \leq j \leq N} (c_{ij} - \bar{c}_i)^2 \left( \sum_{j=1}^{N} (c_{ij} - \bar{c}_i)^2 \right)^{-1} \to 0, \quad i = 1, \ldots, p.$$

c. The matrices $\text{var}_p S_c$ are regular and any accumulation point of the set $\{E \text{var}_p S_c; c_i's satisfy (5)\}$ is a regular matrix.

In the sequel we shall often use the following results:

A. Under hypothesis $H_0$ and assumptions a, b, c the asymptotic distribution of $S_c$ is multivariate normal $\mathcal{M}(\mathbf{0}, \text{var} S_c)$, where $\text{var} S_c$ is the variance matrix of $S_c$ under hypothesis $H_0$ (see [2]).

B. Under hypothesis $H_0$ and assumptions a, b, c the asymptotic distributions of $Q_c^1, \ldots, Q_c^r$ are $\chi^2$ with $p$ and $m_1, \ldots, m_r$ degrees of freedom, resp. (see [2]).

C. Under hypothesis $H_0$ and assumptions a, b, c the matrix $S_c S_c^*$ has asymptotically central Wishart distribution with 1 degree of freedom and positive definite matrix $\text{var} S_c$ (it follows from A).

D. Under hypothesis $H_0$ and assumptions a, b, c the joint asymptotic distribution of $Q_c^1, \ldots, Q_c^r$ is the generalized multivariate $\chi^2$-distribution defined by Jensen in [4], where the corresponding density is derived (it follows from C and [4]).

E. For an arbitrary subvector $S^*_c$ of $S_c$ the relation

$$S^*_c (\text{var}_p S^*_c)^{-1} S^*_c = \max_{u \neq 0} \frac{(u S^*_c)^2}{u' \text{var}_p S^*_c u},$$

34
where \( \mathbf{u} \) are nonzero real vectors, holds and thus
\[
\mathbf{S}_c^*(\text{var}_p, \mathbf{S}_c^*)^{-1} \mathbf{S}_c^* \leq Q_c
\]
(as follows by Schwarz inequality).

F. Bonferroni inequality: For arbitrary events \( A_1, \ldots, A_r \) the inequality
\[
P(\bigcap_{i=1}^r A_i) \geq 1 - \sum_{i=1}^r (1 - P(A_i))
\]
is true.

G. Let a random \( p \)-vector \( \mathbf{Y} = (Y_1, \ldots, Y_p)' = (\mathbf{Y}_1', \ldots, \mathbf{Y}_r')' \) have the normal distribution \( \mathcal{N}(\mathbf{0}, \Sigma) \), where
\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \cdots & \Sigma_{1r} \\
\vdots & \ddots & \vdots \\
\Sigma_{r1} & \cdots & \Sigma_{rr}
\end{pmatrix}
\]
Assume that there exist vectors \( \mathbf{b}_i \) with \( m_i \) components, \( i = 1, \ldots, r \), \( \sum_{i=1}^r m_i = p \), such that
\[
(6) \quad \Sigma_{ij} = \mathbf{b}_i \mathbf{b}_j', \quad i \neq j, \quad i, j = 1, \ldots, r
\]
\[
(7) \quad \Sigma_{ii} - \mathbf{b}_i \mathbf{b}_i' \geq \mathbf{0}, \quad i = 1, \ldots, r
\]
then for arbitrary convex sets \( C_1, \ldots, C_r \) symmetric about origin, \( C_i \subseteq \mathbb{R}^{m_i} \), the inequality
\[
P(\mathbf{Y}_i' \in C_i, i = 1, \ldots, r) \geq \prod_{i=1}^r P(\mathbf{Y}_i' \in C_i)
\]
holds (see [1]).
The inequality always holds for \( m_i = 1, \quad i = 1, \ldots, r \) (see [10]).

2. TEST PROCEDURES

**Procedure I.** The author [2] proposed the test procedure with critical regions
\[
Q_c > \chi^2_{1}\left( p \right),
\]
where \( \chi^2_{1}(p) \) is 100\(\alpha\)% critical value of the central \( \chi^2 \)-distribution with \( p \) degrees of freedom. This test can be used for a class of hypotheses that contain \( H_0 \) as a sub-hypothesis, e.g. for hypothesis that all \( X_{ij}, j = 1, \ldots, N, \) have the same distributions.

**Procedure II.** We base the test procedure on the statistics \( Q_1', \ldots, Q_r' \) given by (4). We reject the hypothesis \( H_v \) if
\[
Q_c' > d_v,
\]
where the \( d_v \)'s are chosen so that
\[
limit_{c} P(Q_c' < d_v, \quad v = 1, \ldots, r) = 1 - \alpha.
\]
The total hypothesis \( H_0 \) is rejected if at least one of the hypotheses \( H_1, \ldots, H_r \) is
rejected. The optimal choice of the $d'_v$'s is not known. Consistently with the classical normal case the values $d_1, \ldots, d_r$ are chosen either to be equal (i.e. $d_1 = \ldots = d_r = d$) or the individual critical regions are of equal sizes (denote them by $d'^*_1, \ldots, d'^*_r$). When $m_v = m$, $v = 1, \ldots, r$ then $d'^*_v = d_v$, $v = 1, \ldots, r$. To find $d, d'^*_1, \ldots, d'^*_r$ with the requested properties is also very difficult for the asymptotic distribution of $(Q^1, \ldots, Q^r)$ includes numerous parameters. This problem was discussed by Jensen in [4] where some approximations are suggested.

We shall suggest here three approximations of $d, d'^*_1, \ldots, d'^*_r$. First consider the approximation of $d$. Using Bonferroni inequality we get an approximative value $\chi^2_{a,r}(\max_{1 \leq i \leq r} m_i)$ and the critical region for testing $H_v$ against $A_v$

$$Q^*_c > \chi^2_{a,r}(\max_{1 \leq i \leq r} m_i).$$

When the assumptions in $G$ are satisfied then the critical region is

$$Q^*_c > \chi^2_{1-(1-a)^{1/r}}(\max_{1 \leq i \leq r} m_i).$$

Utilizing assertion E we get the third possible approximation of $d$. Then we reject the hypothesis $H_v$ if

$$Q^*_c > \chi^2_{a}(p).$$

Similarly we obtain the approximations of $d'^*_1, \ldots, d'^*_r$. By Bonferroni inequality and by $G$ (if possible) we have the critical regions for testing $H_v$ against $A_v$

$$Q^*_c > \chi^2_{a,r}(m_v)$$

and

$$Q^*_c > \chi^2_{1-(1-a)^{1/r}}(m_v),$$

respectively.

If $m_i = 1$, $i = 1, \ldots, p$, the test procedure can be based on the statistics $S_{c1}, \ldots, S_{cp}$. Similarly, in the general case we get critical regions

$$|S_{ci}| > (N^{-1} \sum_{v=1}^{N} (a_N(v) - \bar{a}_{N_i})^2)^{1/2} u \left(1 - \frac{\alpha}{2p}\right),$$

$$|S_{ci}| > (N^{-1} \sum_{v=1}^{N} (a_N(v) - \bar{a}_{N_i})^2)^{1/2} u\left(\frac{1}{2} + \frac{1}{2}(1 - \alpha)^{1/p}\right),$$

$$|S_{ci}| > (N^{-1} \sum_{v=1}^{N} (a_N(v) - \bar{a}_{N_i})^2)^{1/2} (\chi^2_{a}(p))^{1/2},$$

where $u(\cdot)$ is the $100\alpha\%$ quantile of the normal distribution $(0, 1)$.

As for the comparison of the critical regions $(9-10)$, $(12-13)$, we can easily get the following relations among the approximations of $d_1, \ldots, d_r$

$$\chi^2_{1-(1-a)^{1/r}}(\max_{1 \leq i \leq r} m_i) \geq \chi^2_{1-(1-a)^{1/r}}(m_v),$$

$$\chi^2_{a,r}(\max_{1 \leq i \leq r} m_i) \geq \chi^2_{a,r}(m_v) \geq \chi^2_{1-(1-a)^{1/r}}(m_v), \quad v = 1, \ldots, r.$$
Thus the critical region (13) is larger than (9), (10) and (12). The comparison of (11) with the other critical regions is more complicated, e.g.

if $\alpha \leq 0.05$, $p - \max m_i \geq 5$ then $\chi^2(p) > \chi^2_{\alpha}(\max m_i)$,

if $\alpha = 0.05$, $p = 22$, $\max m_i \leq p - 2$ then $\chi^2_{0.05}(p) < \chi^2_{1 - (0.95)^{1/r}}(\max m_i)$.

When $m_i = 1$ then the largest critical region is (15).

**Procedure III.** Define

$$Q^*_c = S^*_c (\var_p S^*_c)^{-1} S^*_c, \quad v = 1, \ldots, r,$$

where

$$S^*_c = S^*_c,$$

$$S^*_c = S^*_c - \cov_p (S^*_c, \ldots, S^*_c) (\var_p (S^*_1, \ldots, S^*_r))^{-1} .$$

$$(S^*_1, \ldots, S^*_r), \quad v = 1, \ldots, r - 1 ,$$

$$(\cov_p (S^*_1, \ldots, S^*_r))'.$$

with $\var_p (\ldots)$ and $\cov_p (\ldots)$ denoting the corresponding submatrices of $\var_p S_c$.

The assertion $A$ implies that the asymptotic distribution of $S_c$ (under hypothesis $H$ and assumptions a, b, c) is multivariate normal with mean $\theta$ and the variance matrix

$$\var S^*_c = \var S^*_c - \cov (S^*_c, \ldots, S^*_c) (\var (S^*_1, \ldots, S^*_r))^{-1} .$$

$$(\cov (S^*_1, \ldots, S^*_r))'.$$

and $Q^*_c$ has asymptotically $\chi^2$-distribution with $m_i$ degrees of freedom. By direct computations we get that $S^*_1, \ldots, S^*_r$ are asymptotically independent and thus so are $Q^*_c, \ldots, Q^*_r$.

Using these arguments one can assert that

$$\lim P(Q^*_c \leq \chi^2_{1 - (1 - \alpha)^{1/r}}(\max m_i), \quad v = 1, \ldots, r) \geq$$

$$\geq \lim P(Q^*_c \leq \chi^2_{1 - (1 - \alpha)^{1/r}}(m_i), \quad v = 1, \ldots, r) = 1 - \alpha .$$

Thus the critical region for testing the hypothesis $H_v$ against $A_v$ can be chosen in either of the following ways:

(17)  \[ Q^*_c > \chi^2_{1 - (1 - \alpha)^{1/r}}(\max m_i), \]

(18)  \[ Q^*_c > \chi^2_{1 - (1 - \alpha)^{1/r}}(m_v). \]

Obviously, the critical region (18) contains (17).

We reject the hypothesis $H_0$ if we reject at least one of $H_1, \ldots, H_r.$
If \( m_i = 1, i = 1, \ldots, p \) the test procedure can be based on the statistics \( S_{ev}^* \), \( v = 1, \ldots, p \). We reject the hypothesis \( H_v \) if

\[
|S_{ev}^*| > (\text{var} S_{ev}^*)^{1/2} n(\frac{1}{2} + \frac{1}{2}(1 - \alpha)^{1/p})
\]

References


Souhrn

MARIE HUŠKOVÁ

SIMULTÁNNÍ PROCEDURY POŘADOVÝCH TESTŮ

Nechť \( X_j, j = 1, \ldots, N \) jsou nezávislé r-prozměrné náhodné vektory se spojitou distribuční funkcí \( F_j \). V článku jsou navržena tři testová kritéria založená na pořadích pro test nezávislosti marginálních rozdělení \( X_j \) na indexu \( j \). Výchozím bodem pro konstrukci testových kritérií byl článek P. R. Krishnaiah (Ann. Inst. Statist. Math. 17, 35—53, 1965).

Author’s address: RNDr. Marie Hušková, CSc. Matematicko-fyzikální fakulta Karlovy univerzity, Sokolovská 83, 186 00 Praha 8.