

PERMUTATION MOMENTS OF A MULTIVARIATE TEST STATISTIC

by

N. S. Urquhart and J. S. Williams

Cornell University and Colorado State University

ABSTRACT

Many tests for mean differences utilize statistics based on a comparison of 'hypothesis' and 'residual' sums of squares. The probability distribution of such statistics is usually derived under the assumption of normally distributed errors. The frequent occurrence of data from non-normal distributions leads one to ask how closely the actual distributions involved are approximated by the normal-derived counterparts. Permutation theory, which has been used to investigate this question in univariate situations, is applied here to the multivariate trace statistic.

PERMUTATION MOMENTS OF A MULTIVARIATE TEST STATISTIC*

by

N. S. Urquhart and J. S. Williams

Cornell University and Colorado State University

Summary

Many tests for mean differences utilize statistics based on a comparison of 'hypothesis' and 'residual' sums of squares. The probability distribution of such statistics is usually derived under the assumption of normally distributed errors. The frequent occurrence of data from non-normal distributions leads one to ask how closely the actual distributions involved are approximated by the normal-derived counterparts. Permutation theory, which has been used to investigate this question in univariate situations, is applied here to the multivariate trace statistic.

1. Introduction

Fisher (1935) noted that the random assignment of treatments to experimental units, without any distributional assumptions, suffices to provide the distribution for an exact test of significance. Pitman (1937) applied Fisher's idea to the randomized complete block design and was able to show that normal-derived approximations served quite well provided adjustments based on the data were incorporated into the standard analysis. More recently Box and Watson (1962) considered a univariate multiple regression model from a point of view parallel to Fisher's permutation theory approach. They derived the first two moments of $W = S_0 / (S_0 + S_E)$ (S_0 and S_E are the hypothesis and error sums of squares, respectively) and noted the importance of the independent-variable configuration in determining the agree-

* Part of a Ph. D. thesis completed at Colorado State University by the senior author, June 1965. Research supported in part by a grant from the National Institutes of Health.

ment between the moments derived under permutation theory and under normal theory.

This paper reports a further study of the role of the treatment (independent-variable) matrix, \underline{X} , in determining the agreement between permutation theory and normal theory regression analyses. The development is along the lines of Fisher's concept of an exact permutation test of significance and uses the methods of Pitman and Box and Watson. The model considered here is a generalization of Box and Watson's model to a multivariate response. We consider $T = \text{tr}[\underline{W}_H(\underline{W}_H + \underline{W}_E)^{-1}]$ as a multivariate generalization of the beta-form of the analysis of variance test criterion. The matrices \underline{W}_H and \underline{W}_E are the hypothesis and error sums of squares and cross products matrices, respectively. The choice of T over other multivariate generalizations of $S_0/(S_0 + S_E)$ is justified on bases other than simplicity.

2. The Permutation Model

Experimenters frequently want to compare the effects of a set of treatments to see if they produce different responses from experimental material. An experiment may be run by allocating t treatments to N experimental units. Treatments may be unrelated, for instance, varieties of sugar beets, in which case there should be more experimental units than treatments. Or the treatments may be related as are the various factor combinations of a factorial experiment. More generally the treatments may represent the application of different levels of several stimuli applied to the experimental units. There is no requirement that they constitute anything like a complete factorial experiment.

All of these situations fit within the framework of a multivariate regression experiment which requires the observation of two matrices $\underline{Y}(N \times q)$ and $\underline{X}(N \times p)$. The responses observed on the i^{th} experimental unit for each of q variates are recorded in the i^{th} row of \underline{Y} . These row elements of \underline{Y} are vector random variables, but a regression experiment requires the values of a number of nonrandom variables

associated with each experimental unit. The values of these p variables for the i^{th} experimental unit are recorded in the i^{th} row of \underline{X} . Together, the pair $(\underline{Y}, \underline{X})$ contain the observable information that can be used for statistical purposes.

This type of regression experiment may be regarded from a finite-population viewpoint and permutation theory applied to it. Each row of \underline{X} , say \underline{x}'_i , represents a treatment. Although \underline{X} has N rows, only $t \leq N$ are distinct. Since each of these treatments could be applied to any experimental unit, there are t treatment populations, each of N experimental units, to which inferences can be drawn using permutation theory.

Denote the elements of these populations by Y_{ij} , $i = 1, 2, \dots, t$, $j = 1, 2, \dots, N$, and let $\bar{Y}_{i.} = \frac{1}{N} \sum_{j=1}^N Y_{ij}$. The experiment yields a response on each experimental unit for the treatment applied to that unit. Call these responses y_{ij} where the subscripts indicate that the i^{th} treatment was applied to the j^{th} experimental unit. The y_{ij} 's are a subset of the Y_{ij} 's for which the identity

$$y_{ij} = \bar{Y}_{i.} + (y_{ij} - \bar{Y}_{i.}) \quad (2.1)$$

always holds. The expressions on the right-hand side of (2.1) can be regarded as the sum of treatment and residual effects.

The observed yields are recorded in the matrix \underline{Y}_p where p indexes the permutation of treatments to experimental units. If $(\bar{Y}_{i.}) = \underline{X}\underline{B}$ has a solution for \underline{B} , we can write

$$\underline{Y}_p = \underline{X}_p \underline{B} + \underline{R}_p \quad (2.2)$$

where \underline{X}_p corresponds to the assignment of treatments to experimental units and is obtainable from \underline{X} by row interchanges. This means that $\underline{X}_p = \underline{P}\underline{X}$ where $\underline{P}(N \times N)$ is a permutation matrix (one element of one in each row and column and all other elements zero), $\underline{P}'\underline{P} = \underline{I}$, and thus $\underline{X}'_p \underline{X}_p = \underline{X}' \underline{P}' \underline{P} \underline{X} = \underline{X}' \underline{X}$. Thus (2.2) can also be

written as $\underline{Y}_p = \underline{PXB} + \underline{R}_p$. This shows that if the matrix \underline{P} is changed and the treatments applied to different experimental units, \underline{Y}_p represents a different selection of elements from $\{\underline{Y}_{1j}\}$.

If $(\underline{\bar{Y}}_1) = \underline{XB}$ does not have a solution, it is possible to reason heuristically that (2.2) should still be applicable. The experimenter includes matrix \underline{X} in the formulation since he feels that it may help explain variation in the values of \underline{Y} . It could be used in many ways, but the simplest is to hope that some linear combination, say \underline{XB} , serves as a prediction for \underline{Y} . That is, the relation (2.2) is true, but now \underline{R}_p is defined by $\underline{R}_p = \underline{Y}_p - \underline{X}_p \underline{B}$ rather than by the identity (2.1) and its statistical properties are assumed to depend only slightly on p .

The absence of treatment effects corresponds to $\underline{B} = \underline{0}$. The class of tests based on $\underline{W}_H = \underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{Y}$ ($(\underline{X}'\underline{X})^{-}$ is the generalized inverse of $\underline{X}'\underline{X}$) and $\underline{W}_E = \underline{Y}'\underline{Y} - \underline{W}_H$ should possess a minimal element of sensitivity to deviations from $H_0: \underline{B} = \underline{0}$. We consider the statistic

$$T = \text{tr}[\underline{W}_H(\underline{W}_H + \underline{W}_E)^{-}] = \text{tr}[\underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{Y}(\underline{Y}'\underline{Y})^{-}] = \text{tr}[\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{Y}(\underline{Y}'\underline{Y})^{-}\underline{Y}'] \quad (2.3)$$

as a potentially robust candidate for a test of $H_0: \underline{B} = \underline{0}$. In the last form of (2.3), \underline{X} and \underline{Y} occur in a symmetric fashion in T . This symmetry forms the basis for our subsequent discussion of how normality of \underline{X} can substitute for normality of \underline{Y} and of the potentially robust character of T relative to other statistics based on \underline{W}_H and \underline{W}_E .

In the proposed absence of treatment effects upon which significance is calculated, the observed yields constitute a uniformity trial;

$\{\underline{Y}_{1j}\} = \{\underline{Y}_{2j}\} = \dots = \{\underline{Y}_{tj}\}$ for all j so that all of the \underline{Y}_p are the same. The observed association of treatments with experimental units (and consequently with responses) manifests only the random assignment of treatments to experimental units. Each other association (allowable by the randomization scheme) of treat-

ments with experimental units is equally probable, and no more meaningful than that observed. This we emphasize here by noting that the description of the null situation under consideration is that \underline{B} is such that $\underline{Y} = \underline{P}\underline{X}\underline{B} + \underline{R}$ for all permissible permutations (\underline{P}) of the treatment labels.

The randomness in the value assumed by T (or any other statistic based on \underline{X} and \underline{Y}) is a consequence of the random assignment of treatments to experimental units. Given \underline{Y} and \underline{X} , the probability distribution of T , or certain properties of this distribution, e.g. its mean and variance, can be evaluated by recourse to the original source of the randomness in T , namely that inherent in the random assignment of treatments to experimental units. Extensive calculations are required for evaluating the permutation distribution of T . Following Pitman (1937) we approximate the permutation distribution by a beta distribution with the same first two moments.

3. The Permutation Moments of T

The mean and variance of T when there is no treatment effect, i.e. $\underline{B} = \underline{0}$, are evaluated by averaging the values of T and T^2 for each possible association of treatments with experimental units. The various associations of treatments with experimental units is effected by interchanging rows of the \underline{X} matrix. This can be accomplished algebraically by replacing \underline{X} in T by $\underline{P}\underline{X}$. Thus the randomness in T lies in the permissible variation of \underline{P} . An examination of (2.3) shows that the permutation matrix can be grouped with \underline{X} or with \underline{Y} . It is more convenient in developing the moments of T to group it with \underline{Y} , i.e. leave the rows of \underline{X} fixed and permute the rows of \underline{Y} .

Notation will be simplified if we write $\underline{M} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$, $\underline{Q}'\underline{Q} = (\underline{Y}'\underline{Y})^{-1}$, where \underline{Q}' is $q \times b$, $b = \text{rank}(\underline{Y})$, and let $\underline{Z} = (\underline{z}_1, \dots, \underline{z}_b) = \underline{Y}\underline{Q}'$. Then $\underline{Z}\underline{Z}' = \underline{Y}(\underline{Y}'\underline{Y})^{-1}\underline{Y}'$ and $\underline{Z}'\underline{Z} = (\underline{z}'_r \underline{z}_s) = \underline{I}_b$. Without loss of generality for statistical purposes, we

assume that all variates are corrected for variate means, that is, that $\underline{j}'\underline{z} = 0$ and $\underline{j}'\underline{X} = 0$. The statistic T is now written as

$$\begin{aligned} T &= \text{tr}[(\underline{Y}'\underline{M}\underline{Y})(\underline{Y}'\underline{Y})^{-1}] = \text{tr}[\underline{Z}'\underline{M}\underline{Z}] \\ &= \text{tr}[\underline{M}\underline{Z}\underline{Z}'] = \sum_{s=1}^b \text{tr}[\underline{M}\underline{z}_s\underline{z}_s'] = \sum_{s=1}^b \underline{z}_s'\underline{M}\underline{z}_s. \end{aligned} \quad (3.1)$$

3.1. $E_p(T)$

The expectation of T over its permutation distribution requires only the evaluation of $E_p(\underline{z}_s\underline{z}_s')$. Since each diagonal position of $\underline{z}_s\underline{z}_s'$ will be occupied by a specific z_{si}^2 $(N-1)!$ times in the $N!$ permutations of the rows of \underline{z}_s ,

$$E(\text{a diagonal element of } \underline{z}_s\underline{z}_s') = \frac{(N-1)!}{N!} \sum_{i=1}^N z_{si}^2 = \frac{1}{N} \underline{z}_s'\underline{z}_s = \frac{1}{N}.$$

Likewise,

$$\begin{aligned} E(\text{an off-diagonal element of } \underline{z}_s\underline{z}_s') &= \frac{(N-2)!}{N!} \sum_{i \neq j} z_{si}z_{sj} = \frac{1}{N(N-1)} [(\sum z_{si})^2 - \sum z_{si}^2] \\ &= -\frac{1}{N(N-1)}. \end{aligned}$$

Combining the above,

$$E(\underline{z}_s\underline{z}_s') = \frac{1}{N} \underline{I}_N - \frac{1}{N(N-1)} (\underline{J} - \underline{I}) = \frac{1}{N-1} (\underline{I} - \frac{1}{N} \underline{J})$$

and thus

$$E_p(T) = \sum_s \text{tr}[\underline{M}E_p(\underline{z}_s\underline{z}_s')] = \frac{b}{N-1} \text{tr}[\underline{M}] - \frac{b}{N(N-1)} \text{tr}[\underline{M}\underline{J}] = \frac{ab}{N-1}, \quad (3.2)$$

where $a = \text{tr}[\underline{M}] = \text{rank}(\underline{X})$. It should be noted here that in the full rank re-

gression case, $a = p$ and with sufficient observations $b = q$. When $q = 1$, this result agrees with that of Box and Watson (1962).

3.2. $\text{Var}_p(T)$

The first step in deriving the permutation variance of T , $\text{Var}_p(T)$, is to expand it in terms of variances and covariances of the variate components from

(3.1):

$$\text{Var}_p(T) = \sum_{s=1}^b \text{Var}_p(\underline{z}'_s \underline{Mz}_s) + \sum_{s \neq t} \text{Cov}_p(\underline{z}'_s \underline{Mz}_s, \underline{z}'_t \underline{Mz}_t). \quad (3.3)$$

The variance terms in the first sum can be derived simply from results given by Box and Watson (1962). They obtained $\text{Var}_p(\underline{z}' \underline{Mz})$ when $\underline{M} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$, i.e. when \underline{X} was full rank. They used only the relations $\underline{Mj} = \underline{0}$, $\underline{M}^2 = \underline{M}$, and $\text{tr}[\underline{M}] = p$ in their argument. Since here we have $\underline{Mj} = \underline{0}$, $\underline{M}^2 = \underline{M}$ and $\text{tr}[\underline{M}] = a$, their result is applicable to $\text{Var}_p(\underline{z}'_s \underline{Mz}_s)$ for $s = 1, 2, \dots, b$ with p replaced by a :

$$\text{Var}_p(\underline{z}'_s \underline{Mz}_s) = \frac{2a(N-1-a)}{(N+1)(N-1)^2} + k_4^s \left[g - \frac{a(a+2)(N-1)}{N(N+1)} \right]. \quad (3.4)$$

Following through details similar to Box and Watson's, we get

$$\text{Cov}_p(\underline{z}'_s \underline{Mz}_s, \underline{z}'_t \underline{Mz}_t) = - \frac{2a(N-1-a)}{(N+1)(N-1)^2(N-2)} + k_{22}^{st} \left[g - \frac{a(a+2)(N-1)}{N(N+1)} \right]. \quad (3.5)$$

In (3.4) and (3.5), g is the sum of squares of the diagonal elements of \underline{M} and k_4^s and k_{22}^{st} are multivariable generalizations of Fisher's k statistics as given by Kendall and Stewart (1963). Subsequently the bracketed expression in (3.5) will be denoted by u .

Combining (3.4) and (3.5),

$$\text{Var}_p(T) = \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} + u \left[\sum_s k_{44}^s + \sum_{s \neq t} k_{22}^{st} \right],$$

and setting the bracketed term equal to K gives

$$\text{Var}_p(T) = \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} + uK. \quad (3.6)$$

Except for a multiplier, u and K are the same function of the matrices \underline{X} and \underline{Y} respectively, i.e. $uK = cf(\underline{X})f(\underline{Y})$.

4. Approximate Distribution of T

If we define D by

$$\frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} D = uK, \quad (4.1)$$

equation (3.6) can be written as

$$\text{Var}_p(T) = (1+D) \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)}. \quad (4.2)$$

Two-moment agreement between the permutation distribution of T and the beta distribution

$$\frac{x^{m-1} (1-x)^{n-1}}{B(m,n)}, \quad 0 \leq x \leq 1,$$

is obtained for $m = \delta a$, $n = \delta(N-1-a)$ by setting

$$\delta = \frac{(Nb-2)(N-1) - 2D(N-b-1)}{2(N-1)(N-b-1)(1+D)}. \quad (4.3)$$

This beta distribution is the normal theory distribution of T for a or b = 1 and

$D = 0$. For $D = 0$ and p and q both at least two, this beta distribution is essentially the approximation to the normal theory distribution of T suggested by Pillai (1955). The variance form (4.2) and the adjustment factor (4.3) are multivariate generalizations of the corresponding quantities from Box and Watson (1962).

The permutation theory and normal theory distributions will thus be approximated by the same distribution whenever $D = 0$. This will happen if either u or K is zero. Box and Watson show that u can be represented as a sum of standardized fourth order k -statistics of the same form as K . If either \underline{X} or \underline{Y} were normal conditional upon the other, then $E(D) = 0$.

Thus if either K or u is near zero, the usual normal theory should provide a reasonable approximation to permutation theory. The statistic K can be zero by every element being zero (which is their expected value under normal theory). Or the various addends of K can be cancelling positive and negative values. This says the net kurtosis of all the variates is zero. Contrast this to univariate analysis where with a single variable, an analogous phenomenon does not exist.

If there is reason to suspect that K will not be close to zero, then instead \underline{X} can be chosen to make $u = 0$. The beta form of Hotelling's T^2 test for the difference between two multivariate means can serve as an illustration of this point. If the first sample mean is based on n_1 observations and the second on n_2 , we can take

$$\underline{X} = \begin{bmatrix} \underline{j}_{n_1} & \underline{0} \\ \underline{0} & \underline{j}_{n_2} \end{bmatrix} - \frac{1}{N} \underline{j}_N (n_1, n_2) = \frac{1}{N} \begin{bmatrix} n_2 \underline{j}_{n_1} & -n_2 \underline{j}_{n_1} \\ -n_1 \underline{j}_{n_2} & n_1 \underline{j}_{n_2} \end{bmatrix} .$$

The statistic u is then a constant times $N^2 + N - 6n_1 n_2$. If n_1 is taken to be larger

than n_2 , then $N^2 + N - 6n_1n_2 = 0$ when approximately $n_1 = 3.7n_2 - 1$.

Thus the normal theory distribution of T should provide a satisfactory approximation to the permutation distribution of T whenever

- i. The distribution of \underline{Y} for any \underline{X} is normal
- ii. The distribution of \underline{X} for any \underline{Y} is normal
- iii. The responses \underline{Y} give $K = 0$
- iv. The regressors \underline{X} give $u = 0$.

Otherwise the approximate beta from (4.3) should be used.

5. Discussion

We wish to make two basic points concerning approximations to the percentiles of the permutation distribution of T . The first concerns the nature of the significance distribution of T when the null hypothesis does not hold and the degree to which this distribution can be approximated with the distribution derived under normal theory. The second is to identify a characteristic of T which separates it from other multivariate test statistics as a likely prospect for a robust test statistic.

We feel that our first point is frequently missed by those who try to equate significance and test of hypothesis procedures. Even when the null hypothesis fails and perhaps as well the primary alternative in mind, significance is calculated from the permutation distribution of the observed response. However, power, the salient value one calculates in the test of hypothesis approach, is derived from a set of outcomes which characterize the alternative and cannot be constructed from the observed data. The difference between these stands out well in the example of the comparison of two treatment means from samples of equal size when a strictly additive model holds. If \underline{y}_j , $j = 1, 2, \dots, 2n$, is the response vector of the j^{th} experimental unit in the absence of any treatment and if the

differential treatment effect is $2t$, the power of a mean-difference test is obtained from the set of $\binom{2n}{n}$ differences

$$\frac{1}{n} \left(\sum_{j=1}^{2n} y_{rj} - \sum_{j'=1}^{2n} y_{rj'} \right) + 2t, \quad j \neq j', \quad r = 1, 2, \dots, \binom{2n}{n}.$$

Significance of the same statistic is calculated from the set of $\binom{2n}{n}$ differences

$$\frac{1}{n} \left(\sum_{j=1}^{2n} y_{rj} - \sum_{j'=1}^{2n} y_{rj'} \right) + \left(\frac{2m_r - n}{n} \right) (2t)$$

where m_r is the number of units receiving treatment 1 which are called just that in the r^{th} permutation of labels to the observed responses; $m_r = n$ only for the observed configuration.

The Neyman-Pearson alternative population and the significance population obviously are different. How well the first of these is approximated by a normal array depends on normal properties of the permuted residual mean differences. Whatever normal characteristic these have are carried over and enhanced in the significance population by the contributions from the distribution of $(2m_r - n)/n$ since for large n this scaled binomial is well approximated by the normal.

The importance of this type of distinction lies in a precise evaluation of significance when the null hypothesis is not true and the mode of treatment response is not strictly additive. The form of the power population then is unknown while the significance population still is completely specified. In order for percentiles calculated from a normal-derived distribution to serve as an adequate approximation for significance probabilities, it is necessary that the test statistic have some normal-derived sampling characteristics whether or not the null hypothesis is true. The act of randomization by itself does not introduce these.

Normal-type properties can be introduced into the distribution of the test statistic either by "nature" through the responses from the sampling units or through the choice of the experimental design. Yates (1964) has endorsed Fisher's view (1935) that the former is the source on which we rely for normality. Nor would one be inclined to look any further when the observed response is an average of responses on several experimental entities drawn randomly from some large population. It is at the other extreme of experimental types, when the units form the whole population to which inference can be made, when they are contiguous in blocks, and when the effect of treatments is a marked stratification of the response, that we feel one may question the normal-derived approximations to the permutation distribution percentiles. This is particularly true and important to consider when treatments convey a strong effect. When the latter is the case, there are three avenues open for solving the inference problem.

(1) One may approximate the distribution of standardized, quadratic, test statistics with the beta distribution where the first two moments agree with those of the permutation distribution. The justification for this rests largely on the study by Pitman (1937) and a scattering of empirical verifications. The results are quite good for a univariate response and the only extension for a multivariate response (a paired Hotelling's T^2 test) which was investigated up to four moments by Arnold (1964), appears to justify the procedure.

(2) One may evaluate the significance of the observed outcome from the permutation distribution. Complete enumeration is possible only for very small numbers of experimental units. Monte Carlo approximations will handle somewhat larger types of experiments, but really large numbers and many restrictions appear at the present to be out of hand for enumeration or sampling techniques.

(3) For test statistics in which the only permutation-variable elements arise from the product $\underline{Y}'\underline{X}$, a permutation of the association, $\underline{Y}'\underline{P}'\underline{X}$, is achieved

either by permuting the rows of \underline{Y} or the rows of \underline{X} . The rows of \underline{X} can be chosen to approximate a normal distribution to a given degree in a discrete manner. When there are a distinct row vectors in \underline{X} then one may choose these to fit the cumulants or k-statistics of \underline{X} up to the a^{th} order with those of the normal. This method works very well when p is small, say 2, and the range of the rows of \underline{X} is not restricted. It does so however at the expense of efficient use of estimation resources when the mean response has a known polynomial form. It may be possible to use this approach in experiments of blocked treatment plans. If the loss in sensitivity compared to a well balanced design is nominal, such designs would be valuable in those situations where a fine degree of approximation to the distribution of the design responses is needed. We are now investigating these problems.

In the realm of questions about a good test statistic for the multivariate analysis of variance we are inclined to rely on the leading statement in (3) to justify the choice of T . The statistics which have so far received considerable attention,

$$\lambda = | \underline{I} - \underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}(\underline{Y}'\underline{Y})^{-1} |$$

$$T_0^2 = \text{tr} \left\{ [\underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}] [\underline{Y}'(\underline{I}-\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X})\underline{Y}]^{-1} \right\}$$

$$c = \max \text{ch} \left\{ [\underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}] [\underline{Y}'(\underline{I}-\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X})\underline{Y}]^{-1} \right\}$$

$$T = \text{tr} [\underline{Y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}(\underline{Y}'\underline{Y})^{-1}]$$

all vary under randomization according to the manner in which the product $\underline{Y}'\underline{X}$ varies. Because $\underline{Y}'\underline{Y}$ and $\underline{X}'\underline{X}$ do not vary, e.g. $\underline{Y}'\underline{P}'\underline{P}\underline{Y} = \underline{Y}'\underline{Y}$, the elements of the inverse matrices in T and λ do not vary while those in T_0^2 and c do. The point in favor of T is that it is the only statistic which is quadratic in the elements of

$\tilde{X}'\tilde{Y}$. If either through the rows of \tilde{X} or the rows of \tilde{Y} the first k -statistics of $\tilde{Y}'\tilde{X}$ agree closely with those of the normal, then to a corresponding degree of approximation those of T fit the normal-derived distribution of T . This closeness of agreement, available only in statistics which vary as standardized quadratics, should give T an advantage in robustness over λ , T_0^2 , and c .

References

- Arnold, H. J. (1964), "Permutation support for multivariate techniques", Biometrika 51:65-70.
- Box, G. E. P. and Watson, G. S. (1962), "Robustness to non-normality of regression tests", Biometrika 49:93-106.
- Fisher, R. A. (1935, 1960), The Design of Experiments. First and Seventh Editions. Oliver and Boyd, Edinburgh, 252 p.
- Kendall, M. G. and Stuart, A. (1963), The Advanced Theory of Statistics, Vol. I, Second Edition, Griffin and Company, London, 433 p.
- Pillai, K. C. S. (1955), "Some new test criteria in multivariate analysis", Ann. math. Statist. 26:117-121.
- Pillai, K. C. S. and Mijares, T. A. (1959), "On the moments of the trace of a matrix and approximations to its distributions", Ann. math. Statist. 30:1135-1140.
- Pitman, E. J. G. (1937), "Significance tests which can be applied to samples from any population III. The analysis of variance test", Biometrika 29:322-335.
- Yates, F. (1964), "Sir Ronald Fisher and the design of experiments", Biometrics 20:307-321.