Canad. Math. Bull. Vol. **18** (3), 1975

AN APPLICATION OF PAIRWISE INDEPENDENCE OF RANDOM VARIABLES TO REGRESSION ANALYSIS

BY
RONALD A. SCHAUFELE

ABSTRACT. The exact probability of making an error at the first stage of a stepwise regression is found, both when the independent variables are fixed and when they are random. The probabilities differ in the two cases because the random variables involved are pairwise independent but not jointly independent.

1. **Introduction.** In linear regression, an F-statistic (or t-statistic) is used to test the hypothesis that the slope is zero. The null distribution of each of these statistics does not depend on whether the independent variable is random or is fixed by the experimenter. The purpose of this paper is to show that the property does not carry over to multilinear regression and that the explanation for this lies in the pairwise independence but not joint independence of the various random variables that are involved. As well, it may serve to bring more attention to a "natural" example of random variables that are pairwise independent but not jointly independent, an example that does not seem to have achieved much notice.

Lastly, a 1971 result of Draper, Guttman and Kanemasu [3] is generalized to include non-orthogonal estimators in the case where the independent variables are fixed.

2. **Preliminaries.** Throughout this paper, we will use the following convention. Capital letters will represent random variables. Small letters will represent fixed variables or values of random variables. For example, if X is a random variable, $E(Y \mid x)$ will represent the conditional expected value of the random variable Y, given $X=x$. Otherwise, $E(Y \mid x)$ will imply that the expected value of Y depends functionally on a (non-random) variable x.

If $Y, X_1, \ldots, X_k$ are random variables, we define $\mu_0=EY, \mu_i=EX_i (i=1, \ldots, k)$, $\sigma_{00}=VY, \sigma_{ii}=VX_i\ (i=1, \ldots, k), \sigma_{0i}=\mathrm{Cov}(Y, X_i)\ (i=1, \ldots, k), \sigma_{ir}=\mathrm{Cov}(X_i, X_r)$ $(i, r=1, \ldots, k; i\neq r)$. We let $\mu=(\mu_0, \ldots, \mu_k)$ and $\underset{\sim}{\Sigma}$ be the $k+1$ by $k+1$ matrix whose (i,j)th entry is $\sigma_{ij} (i, j=0, \ldots, k)$. We let Σ_{ij} be the cofactor of σ_{ij} in $\underset{\sim}{\Sigma}$ and let Σ be determinant of $\underset{\sim}{\Sigma}$.

Received by the editors December 6, 1972 and, in revised form, September 9, 1974.

Assume the following model. Y is a normal random variable such that

$$E(Y \mid \mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \tag{2.1}$$

$$V(Y \mid \mathbf{x}) = \sigma_{0|1\cdots k} \tag{2.2}$$

where $\mathbf{x}=(x_1, \ldots, x_k)$. If $(Y, \mathbf{X})$ has a multivariate normal distribution with parameters $\boldsymbol{\mu}$ and $\underset{\sim}{\Sigma}$ ($\underset{\sim}{\Sigma}$ positive definite), then

$$\beta_0 = \mu_0 - \sum_{i=1}^{k} \beta_i \mu_i, \tag{2.3}$$

$$\beta_i = [(\sigma_{01}, \ldots, \sigma_{0k})\underset{\sim}{\Sigma}_0^{-1}]_i \qquad i = 1, \ldots, k, \tag{2.4}$$

$$\sigma_{0|1,\ldots,k} = \sigma_{00} - (\sigma_{01}, \ldots, \sigma_{0k})\underset{\sim}{\Sigma}_0^{-1}(\sigma_{01}, \ldots, \sigma_{0k})^T \tag{2.5}$$

where $\underset{\sim}{\Sigma}_0$ is obtained from $\underset{\sim}{\Sigma}$ by deleting its first row and first column. In any case, we assume that $-\infty<\beta_i<+\infty$ $(i=0, 1, \ldots, k)$ and that $\sigma_{0|1,\ldots,k}>0$. Within this model, let

$$\Omega = \{\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k): -\infty < \beta_i < +\infty, \qquad i = 1, \ldots, k\},$$

let $\omega_i^1=\{\boldsymbol{\beta} \in \Omega: \beta_i=0\}$ and $\omega_i^{-1}=\{\boldsymbol{\beta} \in \Omega: \beta_i \neq 0\}$ $(i=1, 2, \ldots, k)$ and consider the 2^k subsets of Ω given by $\bigcap_{i=1}^{k} \omega_i^{x_i}$ as $x_i=\pm 1$, $i=1, 2, \ldots, k$. These 2^k subsets form a partition of Ω from which we wish to choose one subset and assert that $\boldsymbol{\beta}$ lies in it, i.e. we wish to assert some subset of the β's are zero and the rest non-zero. This choice will be based on independent observations $Y_1, Y_2, \ldots, Y_n$ associated with vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ $(n>k)$ where $\mathbf{x}_i$ is the value of $\mathbf{X}_i$ if the vector $(Y_i, \mathbf{X}_i)$ is multivariate normal $(i=1, 2, \ldots, n)$.

In Chapter 6 of Draper and Smith [4], six possible methods are listed for choosing one of the subsets of Ω. In this paper, we discuss some problems that arise from the forward selection procedure, the stepwise procedure, and variations thereof. These procedures successively add one more β at each stage to those that are stated as being non-zero until a stage is reached at which any remaining β's are stated to be zero. At stage one, either one β is asserted to be non-zero or all β's are asserted to be zero.

To describe the mechanics of stage one, we assume, for the moment, that $(Y, \mathbf{X})$ is multivariate normal with $\underset{\sim}{\Sigma}=\underset{\sim}{D}$ where $\underset{\sim}{D}$ is a diagonal matrix and that $(Y_1, \mathbf{X}_1), (Y_2, \mathbf{X}_2), \ldots, (Y_n, \mathbf{X}_n)$ are a random sample on $(Y, \mathbf{X})$. We let

$$S_{00} = \sum_{j=1}^{n} (Y_j - \bar{Y})^2, \qquad S_{ii} = \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2, \qquad S_{0i} = \sum_{j=1}^{n} (Y_j - \bar{Y})(X_{ij} - \bar{X}_i)$$

and

$$S_{ir} = \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)(X_{rj} - \bar{X}_r)$$

https://doi.org/10.4153/CMB-1975-073-5 Published online by Cambridge University Press

where $i, r=1, \ldots, k$ $(i \neq r)$. Now let

$$R_i = S_{0i}/(S_{ii}^{1/2} S_{00}^{1/2}) \tag{2.6}$$

$$F_i = \frac{(n-2)S_{0i}^2/S_{ii}}{S_{00}-S_{0i}^2/S_{ii}} \tag{2.7}$$

for $i=1, \ldots, k$. It is easily seen that $F_i=(n-2)R_i^2/(1-R_i^2)$ so that F_i and R_i^2 are 1-1 monotone increasing functions of each other. At stage one, one looks at $S_{01}^2/S_{11}, \ldots, S_{0k}^2/S_{kk}$ and chooses the largest quantity, say S_{0i}^2/S_{ii}. It is conventional to assert that $\beta_i \neq 0$ if and only if, for the corresponding F_i, $F_i \geq f_{1,n-2,\alpha}$ where $f_{1,n-2,\alpha}$ is the upper 100α percent point for the central $F(1, n-2)$ distribution. Otherwise, all the β's are said to be zero.

If we let $S^*=\max(S_{01}^2/S_{11}, \ldots, S_{0k}^2/S_{kk})$ and let $F^*=(n-2)S^*/(S_{00}-S^*)$, then

$$P(F^* \geq f_{1,n-2,\alpha} \mid \underset{\sim}{\Sigma} = \underset{\sim}{D}) \tag{2.8}$$

is the probability of making the error at stage one of saying that some β is non-zero, when, in fact, all the β's are zero.

By inspection, it is seen that S_{0i}^2/S_{ii} is largest if and only if R_i^2 is largest (amongst the R's) if and only if F_i^2 is largest (amongst the F's). Thus,

$$\{F^* < f_{1,n-2,\alpha}\} = \bigcap_{i=1}^{k} \{F_i < f_{1,n-2,\alpha}\}. \tag{2.9}$$

In later sections, we show that the F_i are jointly independent central $F(1, n-2$ random variables when $\underset{\sim}{\Sigma}=\underset{\sim}{D}$. Thus,

$$\begin{aligned} P(F^* \geq f_{1,n-2,\alpha} \mid \underset{\sim}{\Sigma} = \underset{\sim}{D}) &= 1-[P(F < f_{1,n-2,\alpha})]^k \\ &= 1-(1-\alpha)^k \\ &\cong k\alpha. \end{aligned} \tag{2.10}$$

The procedure for choosing a non-zero β in the fixed case is the same but we show that the probability corresponding to (2.8) is different and is a function of the fixed x's, as well as n and α.

3. **Bilinear regression.** Since the formulas are easiest when $k=2$, we study this problem in more detail. Recall that $n>2$. Let H_0 be $\omega_1^1 \cap \omega_2^1$, i.e., $H_0: \beta_1 = \beta_2 = 0$.

THEOREM 3.1. *If (Y, X_1, X_2) is a multivariate $N(\mu, \underset{\sim}{\Sigma})$ random vector with $\underset{\sim}{\Sigma}$ positive definite and (Y_1, Y_{11}, X_{21}), $(Y_2, X_{12}, X_{22}), \ldots, (Y_n, X_{1n}, X_{2n})$ are a random sample on (Y, X_1, X_2), then R_1 and R_2 are independent under H_0, with joint density*

$$f(r_1, r_2) = f(r_1)f(r_2) \tag{3.1}$$

on $-1<r_1, r_2<+1$ where $f(r_i)=\Gamma((n-1)/2)(1-r_i^2)^{(n-4)/2}/\Gamma(\frac{1}{2})\Gamma((n-2)/2)$, $i=1, 2$. Likewise F_1, F_2 are independent $F(1, n-2)$ random variables under H_0.

https://doi.org/10.4153/CMB-1975-073-5 Published online by Cambridge University Press

Proof. Using (2.4), we can write

$$\beta_1 = \frac{\sigma_{01}\sigma_{22}-\sigma_{02}\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \tag{3.2}$$

$$\beta_2 = \frac{\sigma_{02}\sigma_{11}-\sigma_{01}\sigma_{12}}{\sigma_{01}\sigma_{22}-\sigma_{12}^2} \tag{3.3}$$

where the denominators are positive since $\underset{\sim}{\Sigma}$ is positive definite. Under H_0, the numerators are zero. If $\sigma_{01}\neq 0$, $\sigma_{02}\neq 0$, and $\sigma_{12}\neq 0$, then $\sigma_{12}^2=\sigma_{11}\sigma_{22}$, as can be quickly computed. Since this is impossible, $\sigma_{12}=0$ and this implies $\sigma_{01}=\sigma_{02}=0$. Thus, from (29.13.1) on p. 411 of Cramer [2], the joint density of R_1, R_2 and $R=S_{12}/(S_{11}^{1/2}S_{22}^{1/2})$, under H_0 is

$$f(r_1, r_2, r) = \frac{\Gamma\left(\frac{n-1}{2}\right)^2}{\Gamma(\tfrac{1}{2})^3\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-3}{2}\right)}\,(1-r_1^2-r_2^2-r^2+2r_1r_2r)^{(n-5)/2} \tag{3.4}$$

on $-1<r_1, r_2, r<1$ and $1-r_1^2-r_2^2-r^2-2r_1r_2r>0$. Integrating out r [using the change of variable $u=(r-r_1r_2)/(1-r_1^2)^{1/2}(1-r_2^2)^{1/2}$] yields (3.1). Thus, R_1 and R_2 are independent under H_0. Hence, F_1 and F_2 are independent $F(1, n-2)$ random variables under H_0 because of the functional relationship between F and R.

From Theorem 3.1, we see that (2.8) can be written as

$$\begin{aligned} P_{H_0}(F^* \geq f_{1,n-2,\alpha}) &= P_{H_0}(F_1 \geq f_{1,n-2,\alpha} \quad \text{or} \quad F_2 \geq f_{1,n-2,\alpha}) \\ &= 1-(1-\alpha)^2. \end{aligned} \tag{3.5}$$

Formula (3.5) does not follow when the $\mathbf{x}$'s are fixed as the following theorem shows. Let $Y_1, \ldots, Y_n$ be independent normal random variables associated with vectors $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})$. With some abuse of notation, let $S_{00}=\sum_{i=1}^n (Y_i-\bar{Y})^2$, $S_{0i}=\sum_{j=1}^n (Y_j-\bar{Y})(x_{ij}-\bar{x}_i)$, $i=1, 2$, $s_{ii}=\sum_{j=1}^n (x_{ij}-\bar{x}_i)^2$, $i=1, 2$, and $s_{12}=\sum_{j=1}^n (x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)$. Let $R_i=S_{0i}/S_{00}^{1/2}s_{ii}^{1/2}$, $i=1, 2$, $r=s_{12}/s_{11}^{1/2}s_{22}^{1/2}$, and $F_i=(n-2)R_i^2/(1-R_i^2)$, $i=1, 2$. With this notation, we have the following theorem.

THEOREM 3.2. *If* $(1, 1, \ldots, 1)$, $\mathbf{x}_1=(x_{11}, x_{12}, \ldots, x_{1n})$, *and* $\mathbf{x}_2=(x_{21}, x_{22}, \ldots, x_{2n})$ *are fixed (linearly independent) n-vectors such that* $Y_1, Y_2, \ldots, Y_n$ *are independent normal random variables with*

$$E(Y_j \mid \mathbf{x}_1, \mathbf{x}_2) = \beta_0+\beta_1x_{ij}+\beta_2x_{2j} \tag{3.6}$$

$$V(Y_j \mid \mathbf{x}_1, \mathbf{x}_2) = \sigma_1^2, \qquad j = 1, 2, \ldots, n \tag{3.7}$$

then R_1, R_2, *and* S_{00} *have joint density*

$$f(r_1, r_2, s \mid \mathbf{x}_1, \mathbf{x}_2) = \frac{(1-r_1^2-r_2^2-r^2+2r_1r_2r)^{(n-5)/2}s^{(n-3)/2}e^{-(Q)/2\sigma_1^2}}{(2\sigma_1^2)^{(n-1)/2}\Gamma(\tfrac{1}{2})^2\Gamma\left(\frac{n-3}{2}\right)(1-r^2)^{(n-4)/2}} \tag{3.8}$$

https://doi.org/10.4153/CMB-1975-073-5 Published online by Cambridge University Press

where $Q=s-2\beta_1 s_{11}^{1/2} r_1^{1/2} s^{1/2}-2\beta_2 s_{22}^{1/2} r_2^{1/2} s^{1/2}+2\beta_1\beta_2 s_{12}+\beta_1^2 s_{11}+\beta_2^2 s_{22}$ *on* $s>0$, $-1<r_1, r_2<1$, $1-r_1^2-r_2^2-r^2+2r_1r_2r>0$.

In particular, under H_0, R_1 *and* R_2 *(hence* F_1 *and* F_2*) are dependent random variables and* R_1 *and* R_2 *have joint density*

$$f(r_1, r_2 \mid \mathbf{x}_1, \mathbf{x}_2) = \frac{\Gamma\left(\frac{n-1}{2}\right)(1-r_1^2-r_2^2-r^2+2r_1r_2r)^{(n-5)/2}}{\Gamma(\tfrac{1}{2})^2\Gamma\left(\frac{n-3}{2}\right)(1-r^2)^{(n-4)/2}} \tag{3.9}$$

on $-1<r_1, r_2<1$ *and* $1-r_1^2-r_2^2-r^2+2r_1r_2r>0$.

Proof. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are fixed, the joint density of F_1 and F_2 can be found directly by an extension of the matrix method that shows the independence of $\bar{X}$ and s^2 when $X_1, X_2, \ldots, X_n$ are a random sample on X, a normal random variable. Let A be the orthonormal matrix formed from $(1, 1, \ldots, 1)$, $\mathbf{x}_1$, $\mathbf{x}_2$, and $n-3$ additional linearly independent n-vectors. (Recall that $(1, 1, \ldots, 1)$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are linearly independent vectors.) Using the fact that $\mathbf{U}=\underset{\sim}{A}\mathbf{Z}$ is $N(\mathbf{O}, I)$ if $\mathbf{Z}$ is $(\mathbf{O}, I)$, changing variables carefully, and integrating out extraneous variables yields the joint density of R_1, R_2, and S_{00} as

$$f(r_1, r_2, s \mid \mathbf{x}_1, \mathbf{x}_2) = \frac{(1-r_1^2-r_2^2-r^2+2r_1r_2r)^{(n-5)/2} s^{(n-3)/2} e^{-Q/2\sigma_1^2}}{(2\sigma_1^2)^{(n-1)/2}\Gamma(\tfrac{1}{2})^2\Gamma\left(\frac{n-3}{2}\right)(1-r^2)^{(n-4)/2}} \tag{3.10}$$

where $Q=s-2\beta_1 s_{11}^{1/2} r_1^{1/2} s^{1/2}-2\beta_2 s_{22}^{1/2} r_2^{1/2} s^{1/2}+2\beta_1\beta_2 s_{12}+\beta_1^2 s_{11}+\beta_2^2 s_{22}$ on the appropriate ranges. Note that this density depends on $\mathbf{x}_1$ and $\mathbf{x}_2$ only through s_{11}, s_{22}, and s_{12}.

Equation (3.8) can be obtained in an alternate way by assuming that the $\mathbf{x}$'s, though fixed, are values of corresponding normal random variables. That is, write out the Wishart distribution for S_{00}, S_{01}, S_{02}, S_{11}, S_{12}, and S_{22} in terms of variances and covariances, divide it by the Wishart distribution of S_{11}, S_{12}, and S_{22}, and obtain the distribution of S_{00}, S_{01}, S_{02}, given $(S_{11}, S_{12}, S_{22})=(s_{11}, s_{12}, s_{22})$. Use (2.4) and (2.5) to change the constants from variances and covariances to β's and $\sigma_1^2=\sigma_{0|1,\ldots,k}^2$. Finally, use the mapping $(S_{00}, S_{01}, S_{02})\to(S_{00}, R_1, R_2)$ to show that the conditional density of S_{00}, R_1, and R_2 is (3.8).

If H_0 is true, inspection of (3.10) shows that S_{00} is independent of R_1 and R_2, the joint density of latter depends on $\mathbf{x}_1$ and $\mathbf{x}_2$ only through r, their sample correlation coefficient, and the density is given by

$$f(r_1, r_2 \mid \mathbf{x}_1, \mathbf{x}_2) = \frac{\Gamma\left(\frac{n-1}{2}\right)(1-r_1^2-r_2^2-r^2+2r_1r_2r)^{(n-5)/2}}{\Gamma(\tfrac{1}{2})^2\Gamma\left(\frac{n-3}{2}\right)(1-r^2)^{(n-4)/2}} \tag{3.11}$$

https://doi.org/10.4153/CMB-1975-073-5 Published online by Cambridge University Press

on the appropriate range. Thus, under H_0, R_1 and R_2 (hence F_1 and F_2) are dependent and the proof is complete.

The reason for this basic difference between the random and fixed cases can be seen by looking at equations (3.1) and (3.4). From these, it follows that the random variables R_1, R_2, and R are pairwise independent but not jointly independent as was first shown by Geisser and Mantel [6] in 1962. When the independent variables are random, R_1 and R_2 (hence F_1 and F_2) are independent, under H_0. When the independent variables are fixed, r is known and the conditional distribution of R_1 and R_2, obtainable by dividing (3.4) by the density of R, is not factorable, even under H_0.

Most elementary texts in probability and statistics stress the fact that pairwise independence does not, in general, imply total independence. But natural examples are hard to find, particularly examples involving continuous random variables. (See Feller [5] p. 116–117.) Thus, Geisser and Mantel's example is of valuable pedagogical use and, in the present setup, explains why different results should be expected in the fixed and random cases.

For $k=2$, we have found P_{H_0} (H_0 is rejected) both in the random and in the fixed cases. In the former case, the error probability can be set at α by setting the critical level at $f_{1,n-2,1-\sqrt{1-\alpha}}$. In the latter case, for any critical level c, we have

$$\begin{aligned} P_{H_0}(H_0 \text{ is rejected}) &= P_{H_0}(F_1 \geq c \quad \text{or} \quad F_2 \geq c) \\ &= P_{H_0}(R_1^2 \geq c' \quad \text{or} \quad R_2^2 \geq c') \\ &= 1 - \int_{-c'}^{c'} \int_{-c'}^{c'} f(r_1, r_2 \mid \mathbf{x}_1, \mathbf{x}_2)\, dr_1\, dr_2 \end{aligned} \tag{3.12}$$

where $c'=c/(n-2+c)$ and this probability depends on c, n, and r (the correlation coefficient of $\mathbf{x}_1$ and $\mathbf{x}_2$). Note that $r^2<1$ by our assumption that $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent.

In 1971, Draper *et al.* [3] found the probability given in (3.12) for a special case. They assume that the estimators of the β's are orthogonal. Inspection of the covariance matrix shows that this implies that r, the sample correlation coefficient of the x_1's and x_2's, is zero. Orthogonality actually implies more but (3.12) depends on the x's only through r. Thus, for various n's and α's, they find that the error probabilities behave approximately as they would if the F's were independent when $r=0$.

If the independent variables are fixed by the experimenter, the probabilities and/or critical levels must be found by numerical means but it would seem wiser to compute these quantities as needed rather than preparing tables since these would be simply too large. However, computations to see how the probability in (3.12) depends on the value of r would be of interest.

4. **Multilinear regression.** For $k>2$ and $n>k$, similar results hold but the equations become unwieldy. If $(Y, X_1, \ldots, X_k)$ is a multivariate $N(\mu, \Sigma)$ random

https://doi.org/10.4153/CMB-1975-073-5 Published online by Cambridge University Press

vector with $\underset{\sim}{\Sigma}=\underset{\sim}{D}$, a diagonal matrix that is positive definite, then $R_1, R_2, \ldots, R_k$ are jointly independent. (See, for example, Anderson [1], p. 177, Problem 9). Thus, the problem of finding $P_{H_0}(F_1 \geq c \text{ or } F_2 \geq c \text{ or } \cdots \text{ or } F_k \geq c)$ for fixed c and the problem of finding c so as to fix the above probability are easily solved. There are, however $k(k+1)/2$ correlation coefficients and they are jointly dependent. Since Y holds no special place in this vector, it follows that amongst the $k(k+1)/2$ jointly dependent random variables, there are $k+1$ sets of k random variables that are jointly independent, each correlation coefficient appearing exactly twice in the $k+1$ sets. This implies that the $k(k+1)/2$ correlation coefficients are pairwise independent, as pointed out by Geisser and Mantel [6]. More importantly, it shows that k-independence of certain subsets of $k(k+1)/2$ random variables is still not enough to ensure the joint independence of the full set.

Thus, in the general case where the dependent variables are random, if H_0 is $\bigcap_{i=1}^{k} \omega_1^i$, i.e., $H_0: \beta_1 = \cdots = \beta_k = 0$, then (2.8) can be written as

$$
\begin{aligned}
P_{H_0}(H_0 \text{ is rejected}) &= P_{H_0}(F^* \geq f_{1,n-2,\alpha}) \\
&= 1 - P_{H_0}(F_1 < f_{1,n-2,\alpha}, \ldots, F_k < f_{1,n-2,\alpha}) \\
&= 1 - (1-\alpha)^k \\
&\cong k\alpha,
\end{aligned} \tag{4.1}
$$

for α small.

If the independent variables are fixed by the experimenter, the easiest way to obtain the conditional distribution of $R_1, R_2, \ldots, R_m$, under H_0, is to divide the joint density of all $k(k+1)/2$ correlation coefficients by the joint density of the $(k-1)k/2$ correlation coefficients that do not involve Y. If $\underset{\sim}{R}=(r_{ij})$ is a $k+1$ by $k+1$ matrix $(i, j=0, 1, \ldots, k)$ with determinant $R=|\underset{\sim}{R}|$ and R_{00} is the cofactor of r_{00}, the conditional density of $R_1, R_2, \ldots, R_k$ is given by

$$
f(r_{01}, \ldots, r_{0k} \mid r_{12}, \ldots, r_{k-1,k}) = \frac{C(k+1, n)}{C(k, n) R_{00}^{1/2}} \left(\frac{R}{R_{00}}\right)^{(n-k-3)/2} \tag{4.2}
$$

on the set where R is positive, the density being defined whenever R_{00} is positive, and $C(k+1, n)/C(k, n)$ being a constant so that the integral of (8) over $r_{01}, \ldots, r_{0k}$ is one.

For a fixed significance level c, it is now theoretically possible to compute the probability of rejecting H_0, given H_0 is true. Likewise, it is theoretically possible to determine c so that the stated probability is α. But both problems can only be solved by numerical methods, preferably on a computer. In [3], it is shown that the R's behave as if they are approximately independent when the estimators are orthogonal.

5. **Problems.** The results of this paper leave many problems. Amongst them are the following:

(1) Is the error probability relatively uninfluenced by the choice of r when the

independent variables are fixed ($k=2$), i.e., are the F's approximately independent no matter what the choice of the x's.

(2) For fixed n and $\alpha(k=2)$, can r be chosen, in the fixed case, so as to minimize the error probability? Are there values of r for which the error probability is less than the probability when the independent variables are random?

(3) More generally, this paper only investigates the probability of one error for one test. The other errors are functions of the non-zero β's. Does this test perform well, at least as $|\boldsymbol{\beta}|$ becomes large? The procedures mentioned in the introduction use at least k tests in stage one. Can a procedure be devised that uses only k tests in total and, for fixed probability of rejecting H_0 when H_0 is true, improves the probability of finding other hypotheses when they are true?

(4) In some applied areas, e.g. economics, the independent variables are usually considered random and k is often large. In such a situation, use of any common α, e.g. 0·05 or 0·01, will lead often to the choosing of some β at stage one and the assertion that it is significantly different from zero when, in fact, all the β's are zero. How can this be avoided? At one extreme, one can raise the critical level in 4.1. This will lower the error probability we have been discussing but may make the discovery of true non-zero β's very difficult. Is there an optimal way to choose the critical level, as a function of k, to keep all probabilities low, at least for large n?

ACKNOWLEDGEMENTS. I am indebted to Professor R. Hamilton for bringing this problem to my attention. I wish to thank the referee for his useful criticisms and for drawing my attention to the work of Draper *et al.* [3].

REFERENCES

1. T. W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, 1958.
2. Harald Cramer, *Mathematical methods of statistics*, Princeton University Press, 1946.
3. N. R. Draper, I. Guttman, and A. Kanemasu, *The distribution of certain regression statistics*, Biometrika, 1971, **58** 295–298.
4. Norman Draper and Harry Smith, *Applied regression analysis*, Wiley, 1966.
5. W. Feller, *An introduction to probability theory and its applications* (*2nd edition*), Wiley, 1957.
6. Seymour Geisser and Nathan Mantel, *Pairwise independence of jointly dependent variables*, Ann. Math. Statist., 1965, **36** 1313–17.

YORK UNIVERSITY