

THE BEHRENS-FISHER PROBLEM FOR REGRESSION COEFFICIENTS

BY D. A. S. FRASER

University of Toronto

1. Summary. For two normal populations with unknown variances and means depending linearly on $p + q$ regression variables, a Behrens-Fisher generalization is to test the equality of q regression coefficients in one population with a corresponding set in the second population. When $q = 1$ a general class of similar regions is obtained for the hypothesis, and for regions restricted to this class a most powerful or most powerful unbiased test is found. When $q > 1$ several tests are presented and discussed.

2. Introduction.

2.1. *The problem.* Let U be a normally distributed random variable with mean $\sum_{r=1}^p \beta_r x_r + \sum_{s=1}^q T_s y_s$ and variance σ^2 ; the location of the distribution depends linearly on the "factors" $x_1, \dots, x_p; y_1, \dots, y_q$. The regression coefficients β_r for $r = 1, \dots, p$ and T_s for $s = 1, \dots, q$ are unknown but fixed constants while the factors $x_1, \dots, x_p; y_1, \dots, y_q$ can be fixed or recorded for each observed value of U . Similarly let U^* be normally distributed with mean $\sum_{r=1}^p \beta_r^* x_r^* + \sum_{s=1}^q T_s^* y_s^*$ and variance σ^{*2} . The problem is to test the hypothesis $H_q = \{T_s = T_s^*; s = 1, \dots, q\}$.

Consider a sample of size m from the distribution of U ; we have

$$\{U_\alpha; x_{1\alpha}, \dots, x_{p\alpha}; y_{1\alpha}, \dots, y_{q\alpha}; \alpha = 1, \dots, m\}.$$

Letting the observed values of U and of the factors be column vectors in m -dimensional Euclidean space R^m , the sample yields $\{\bar{U}; \bar{x}_1, \dots, \bar{x}_p; \bar{y}_1, \dots, \bar{y}_q\}$. For given values of the factors, \bar{U} is distributed in a spherically symmetric normal distribution about the point $\sum \beta_r \bar{x}_r + \sum T_s \bar{y}_s$ with variance σ^2 . Similarly a sample of size n from the distribution of U^* is given by $\{\bar{U}^*; \bar{x}_1^*, \dots, \bar{x}_p^*; \bar{y}_1^*, \dots, \bar{y}_q^*\}$ which is a set of vectors in R^n . \bar{U}^* has a spherically symmetric distribution about $\sum \beta_r^* \bar{x}_r + \sum T_s^* \bar{y}_s$ with variance σ^{*2} . We assume for each sample that the vectors representing the factors are linearly independent.

Consider the following orthogonality conditions on the vectors which record the observed values of the different factors: $\bar{y}_1, \dots, \bar{y}_q$ are mutually orthogonal and orthogonal to $\bar{x}_1, \dots, \bar{x}_p$; $\bar{y}_1^*, \dots, \bar{y}_q^*$ are mutually orthogonal and orthogonal to $\bar{x}_1^*, \dots, \bar{x}_p^*$. The problem can always be put in this form by replacing the given factors by suitable linear combinations thereof. For consider a set of independent vectors $\bar{x}_1, \dots, \bar{x}_p; \bar{y}_1, \dots, \bar{y}_q$. By using instead the vectors $\bar{x}_1, \dots, \bar{x}_p; \bar{y}_{1.x}, \dots, \bar{y}_{q.x}$ where $\bar{y}_{s.q}$ is the regression residual after fitting

Received 10/28/51, revised 3/24/53.

$\bar{x}_1, \dots, \bar{x}_p$, we have orthogonality between the \bar{x}_r set and the $\bar{y}_{s,x}$ set. Doing similarly for the \bar{x}^*, \bar{y}^* we have

$$\begin{aligned}\sum \beta_r \bar{x}_r + \sum T_s \bar{y}_s &= \sum \gamma_r \bar{x}_r + \sum T_s \bar{y}_{s,x} \\ \sum \beta_r^* \bar{x}_r^* + \sum T_s^* \bar{y}_s^* &= \sum \gamma_r^* \bar{x}_r^* + \sum T_s^* \bar{y}_{s,x}^*\end{aligned}$$

for suitably defined γ_r, γ_r^* ; and the hypothesis H_q remains unchanged. Now replace the vectors $\bar{y}_{s,x} (\bar{y}_{s,x}^*)$ by linear combinations thereof, say $\sum_s a_{ts} \bar{y}_{s,x} (\sum_s a_{ts} \bar{y}_{s,x}^*)$, where the transformation matrix $\|a_{ts}\|$ is chosen so that both $\|\bar{y}'_{t,x} \bar{y}_{s,x}\|$ and $\|\bar{y}'_{t,x} \bar{y}_{s,x}^*\|$ become diagonal matrices. The vectors $\sum_s a_{ts} \bar{y}_{s,x}, (t = 1, \dots, q)$ are then mutually orthogonal and orthogonal to the \bar{x}_r ; similarly for $\sum_s a_{ts} \bar{y}_{s,x}^*$ and the \bar{x}_r^* . Also the hypothesis H_q is unchanged since the same linear combinations are used for the vectors $\bar{y}_{s,x}$ as for $\bar{y}_{s,x}^*$.

For our problem P_q , we summarize the notation and structure as follows.

I. \bar{U} is normally distributed in R^m with

$$(1) E\{\bar{U}\} = \sum_{r=1}^p \beta_r \bar{x}_r + \sum_{s=1}^q T_s \bar{y}_s,$$

and

$$(2) \text{Covariance Matrix} = \sigma^2 I \text{ (} I \text{ is the identity matrix),}$$

where

$$(3) \bar{y}_1, \dots, \bar{y}_q \text{ form an orthogonal set and each is orthogonal to the space generated by } \bar{x}_1, \dots, \bar{x}_p.$$

\bar{U}^* is normally distributed in R^n with

$$(1) E\{\bar{U}^*\} = \sum_{r=1}^p \beta_r^* \bar{x}_r^* + \sum_{s=1}^q T_s^* \bar{y}_s^*.$$

and

$$(2) \text{Covariance Matrix} = \sigma^{*2} I,$$

where

$$(3) \bar{y}_1^*, \dots, \bar{y}_q^* \text{ form an orthogonal set and each is orthogonal to the space generated by } \bar{x}_1^*, \dots, \bar{x}_p^*.$$

The hypothesis to be tested is $H_q = \{T_s = T_s^*; s = 1, \dots, q\}$.

2.2. *Examples.* The following examples of the above structure have been treated in the literature.

EXAMPLE 2.1. The original Behrens-Fisher problem [3]. This case is characterized by $p = 0, q = 1, \bar{y} = (1, \dots, 1)$ and $\bar{y}^* = (1, \dots, 1)$. Within each sample the mean is constant and the hypothesis to be tested is that the samples have the same mean. The solution in this paper reduces to that given by Scheffé [2].

EXAMPLE 2.2. The problem proposed by Barankin [1]. For this problem we have $p = 1, q = 1, \bar{x} = (1, \dots, 1), \bar{y} \equiv (\xi_1 - \sum \xi_i/m, \dots, \xi_m - \sum \xi_i/m) \cdot \bar{x}^* = (1, \dots, 1)$ and $\bar{y}^* = (\eta_1 - \sum \eta_i/n, \dots, \eta_n - \sum \eta_i/n)$. The hypothesis H is that the regression coefficients on ξ and η are equal. The test proposed reduces to that given by Barankin; however the properties proved are different. In this paper the test is shown to be most powerful or most powerful unbiased in a general class of similar tests. Barankin showed the test was most powerful among tests based on certain linear combinations of the variables (see [1]).

rows in the matrix A . Complete A by adding $m - p - q$ normalized orthogonal row vectors.

The original assumptions and structure of the problem are transformed by A to the following.

II. \bar{L} is normally distributed in R^m with

$$\begin{aligned} (1) \quad E(L_1) &= \gamma_1, \dots, & E(L_p) &= \gamma_p, \\ E(L_{p+1}) &= T_1 n_1, \dots, & E(L_{p+q}) &= T_q n_q, \\ E(L_{p+q+i}) &= 0, i = 1, \dots, m - p - q, \end{aligned}$$

and

$$(2) \quad \text{Covariance Matrix } \sigma^2 I.$$

The constants $\gamma_1, \dots, \gamma_p$ are independent linear functions of the constants β_1, \dots, β_p . Since the β 's are unknown regression coefficients in this problem, so also are the coefficients $\gamma_1, \dots, \gamma_p$.

Similarly we introduce a transformation A^* in the space R^n . The new structure for L_1^*, \dots, L_n^* is given by the above equations where each T, γ, n, L is given an asterisk.

The hypothesis to be tested is $H_q = \{T_s = T_s^*; s = 1, \dots, q\}$.

We further transform the problem by the following transformation of the variables.

$$\begin{aligned} T_1 &= \frac{L_{p+1}}{n_1} - \frac{L_{p+1}^*}{n_1^*}, \dots, T_q = \frac{L_{p+q}}{n_q} - \frac{L_{p+q}^*}{n_q^*}, \\ S_1 &= \frac{L_{p+1}}{n_1} + \frac{L_{p+1}^*}{n_1^*}, \dots, S_q = \frac{L_{p+q}}{n_q} + \frac{L_{p+q}^*}{n_q^*}, \\ V_r &= L_r & r &= 1, \dots, p \\ V_r^* &= L_r^* & r &= 1, \dots, p \\ z_i &= L_{p+q+i} & i &= 1, \dots, m - p - q \\ z_j^* &= L_{p+q+j}^* & j &= 1, \dots, n - p - q. \end{aligned}$$

The formulation of the original problem P_q in terms of these transformed random variables $T_s, S_s, V_r, V_r^*, z_i, z_j^*$ is:

III. $z_i; i = 1, \dots, m'$ are normally and independently distributed with means 0 and variances σ^2 .

$z_j^*; j = 1, \dots, n'$ are normally and independently distributed with means 0 and variances σ^{*2} .

$V_r; r = 1, \dots, p$ are normally and independently distributed with means γ_r and variances σ^2 .

$V_r^*; r = 1, \dots, p$ are normally and independently distributed with means γ_r^* and variances σ^{*2} .

(T_1, S_1) is normal with mean (d_1, α_1) and covariance matrix

$$\begin{bmatrix} c_1 \sigma^2 + c_1^* \sigma^{*2} & c_1 \sigma^2 - c_1^* \sigma^{*2} \\ c_1 \sigma^2 - c_1^* \sigma^{*2} & c_1 \sigma^2 + c_1^* \sigma^{*2} \end{bmatrix}.$$

(T_q, S_q) is normal with mean (d_q, α_q) and covariance matrix

$$\begin{bmatrix} c_q \sigma^2 + c_q^* \sigma^{*2} & c_q \sigma^2 - c_q^* \sigma^{*2} \\ c_q \sigma^2 - c_q^* \sigma^{*2} & c_q \sigma^2 + c_q^* \sigma^{*2} \end{bmatrix}.$$

A priori Hypothesis: $c_1, c_1^*, \dots, c_q, c_q^*$ are known positive real numbers; $\sigma^2, \sigma^{*2}, \gamma_1, \gamma_1^*, \dots, \gamma_p, \gamma_p^*, \alpha_1, \dots, \alpha_q$ are real numbers with values unknown.

Hypothesis H_q to be tested: $\{d_1 = d_2 = \dots = d_q = 0\}$. Alternative Hypothesis: $\{(d_1, \dots, d_q) \in D \subset R^q; (0, \dots, 0) \notin D\}$.

Note that $m' = m - p - q, n' = n - p - q$. Also the γ 's, γ^* 's, α 's, and d 's are simply related to the original regression coefficients β 's, β^* 's, T 's, T^* 's in terms of constants which depend on the values of the original regression variables.

3.2. *Reduction of the class of similar regions.* We look for a general class of similar regions under the hypothesis H_q . The following theorem allows us to restrict our choice.

THEOREM 1. Any region in R^{m+n} similar of size α under variation of $\gamma_1, \dots, \gamma_p, \gamma_1^*, \dots, \gamma_p^*$ has almost everywhere conditional size α given $V_1, \dots, V_p, V_1^*, \dots, V_p^*$.

The proof follows immediately from the completeness of the multiple Laplace transform.

Under the hypothesis H_q and its alternative the conditional distribution given the V 's and V^* 's is independent of the V 's, V^* 's and γ 's, γ^* 's. Thus any similar test most powerful against a simple alternative can be chosen independent of the V 's and V^* 's.

3.3. *The principle of invariance.* The reduction obtained from Theorem 1 can also be obtained on the basis of the principle of invariance, which in fact provides an additional reduction. Consider the following linear transformation

$$\begin{bmatrix} \bar{V}_r = V_r + \lambda_r \\ \bar{V}_r^* = V_r^* + \lambda_r^* \\ \bar{S}_s = S_s + \mu_s \end{bmatrix} \quad \begin{matrix} r = 1, \dots, p \\ s = 1, \dots, q \end{matrix}.$$

This induces the following transformation on the parameter space:

$$\begin{bmatrix} \bar{\gamma}_r = \gamma_r + \lambda_r \\ \bar{\gamma}_r^* = \gamma_r^* + \lambda_r^* \\ \bar{\alpha}_s = \alpha_s + \mu_s \end{bmatrix} \quad \begin{matrix} r = 1, \dots, p \\ s = 1, \dots, q \end{matrix}.$$

It is easily seen that the problem is invariant under such transformations since the values of the parameters $\{d_s\}$ are unaffected. Thus on the basis of the prin-

ciple of invariance we look for a test which is independent of the values of the V 's, V^* 's, and S 's.

Our problem has now been considerably simplified; it can be summarized as follows.

IV. $T_1, \dots, T_q; z_1, \dots, z_{m'}; z_1^*, \dots, z_{n'}^*$ are normally and independently distributed with means $d_1, \dots, d_q; 0, \dots, 0; 0, \dots, 0$ and variances $c_1\sigma^2 + c_1^*\sigma^{*2}, \dots, c_q\sigma^2 + c_q^*\sigma^{*2}; \sigma^2, \dots, \sigma^2; \sigma^{*2}, \dots, \sigma^{*2}$. The values $c_1, c_1^*, \dots, c_q, c_q^*$ are prescribed, and σ^2, σ^{*2} are unknown. The hypothesis to be tested is: $H_q = \{d_1 = \dots = d_q = 0\}$.

4. Test for H_1 . In this section the Scheffé test [2] is shown to be most powerful or most powerful unbiased in a general class of similar tests.

4.1. *The class of similar regions.* The problem P_1 may be slightly modified by changing the scale of the z 's by a factor $c^{\frac{1}{2}}$ and the z^* 's by a factor $c^{*\frac{1}{2}}$. Letting $c\sigma^2$ and $c^*\sigma^{*2}$ be respectively σ^2 and σ^{*2} , we have $P_1: T; z_i; z_j^* (i = 1, \dots, m'; j = 1, \dots, n')$ are normally and independently distributed with means $d; 0; 0$ and variances $\sigma^2 + \sigma^{*2}; \sigma^2; \sigma^{*2}$. The hypothesis is $H_1 = \{d = 0\}$.

We now define a class of regions similar of size α for the probability distributions of P_1 assuming H_1 ; the distributions have $d = 0$ and no restrictions on σ^2 and σ^{*2} . Assume $m' \leq n'$ without loss of generality. Considering $(T; z_i; z_j^*)$ as a point in $(m' + n' + 1)$ -dimensional Euclidean space $R^{m'+n'+1}$, we form hyperspherical cylinders of the following form:

$$\mathcal{R}_P(r) = \left\{ (T; z_i; z_j^*) \left| \sum_{i=1}^{m'} \left(z_i + \sum_{j=1}^{n'} a_{ij}^{(P)} z_j^* \right)^2 + T^2 = r^2 \right. \right\}.$$

Here $\|a_{ij}^{(P)}\|$ is an $n' \times n'$ orthogonal matrix and P indexes it within the class \mathcal{P} of all orthogonal $n' \times n'$ matrices.

On the hyperspherical cylinder $\mathcal{R}_P(r)$, consider hyperplanes of the following form:

$$\mathcal{H}_P(c_1, \dots, c_{m'}; r) = \left\{ (T; z_i; z_j^*) \left| z_i + \sum_{j=1}^{n'} a_{ij}^{(P)} z_j^* = c_i, \sum_{i=1}^{m'} c_i^2 + T^2 = r^2 \right. \right\}.$$

We construct a class \mathcal{C}_α'' of regions in $R^{m'+n'+1}$. \mathcal{C}_α'' consists of all sets S having the following structure: $S = \bigcup_{r \in R^1} \bigcup_{P \in \mathcal{P}^*} S_{rP}$, where the sets S_{rP} satisfy

1. $S_{rP} \subset \mathcal{R}_P(r)$ is measurable in the space $\mathcal{R}_P(r)$.
2. $S_{rP} \cap S_{r'P'} = \emptyset$ unless $r = r', P = P'$
3. \mathcal{P}^* is a countable subset of \mathcal{P} .
4. If $\bigcup_{P \in \mathcal{P}^*} [S_{rP} \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r)] \neq \emptyset$, then using Lebesgue measure λ over the Euclidean spaces $\mathcal{H}_P(c_1, \dots, c_{m'}; r)$

$$\sum_{P \in \mathcal{P}^*} \int_{\mathcal{H}_P} p \varphi_P d\lambda = \int_{\mathcal{H}_P} p d\lambda,$$

where p is the given probability density function, and φ_P is the characteristic function of the set S_{rP} . The integrals exist for almost all values of r .

5. If m is the uniform normalized measure over the sphere $S_r = \{(c_1, \dots, c_{m'}, T) \mid \sum_{i=1}^{m'} c_i^2 + T^2 = r^2\}$ in $R^{m'+1}$, then $m(S_r) = \alpha$ for almost all values of r where

$$S_r = \{(c_1, \dots, c_{m'}; T) \mid \bigcup_{P \in \mathcal{P}^*} [S_{rP} \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r)] \neq \phi\}.$$

Let \mathcal{C}'_α be the sets $S \in \mathcal{C}''_\alpha$ for which $S_P = \bigcup_{r \in R'} S_{rP}$ is measurable for $P \in \mathcal{P}^*$, and let \mathcal{C}_α be all sets which differ from those in \mathcal{C}'_α by sets of measure zero.

4.2. *Proofs concerning similarity.* Theorem 2 in this section establishes that \mathcal{C}_α is a class of similar regions. Theorem 3 shows that any similar region satisfying Condition 4 above satisfies Condition 5. Also for any measurable region there is a region differing by a null set for which sets S_{rP} can be defined satisfying Conditions 1 and 2. Thus, if \mathcal{C}_α is the class of all similar regions of size α , a proof of this would only need to show that for each S the sets S_{rP} can be chosen to satisfy Conditions 3 and 4.

THEOREM 2. *Each region belonging to \mathcal{C}_α is similar of size α with respect to the measures of $P_1(H_1)$.*

PROOF. Since the measures of P_1 are dominated by Lebesgue measure, we need only establish that \mathcal{C}'_α is a class of similar regions.

Let $S \in \mathcal{C}'_\alpha$ and let φ_S be its characteristic function over $R^{m'+n'+1}$. Since S also belongs to \mathcal{C}''_α , let S_{rP} and \mathcal{P}^* be as given in Conditions 1–5. Define S_P by $S_P = \bigcup_{r \in R'} S_{rP}$, and let φ_P be the characteristic function of S_P . Since $S = \bigcup_{P \in \mathcal{P}^*} S_P$, we have $\varphi_S = \sum_{P \in \mathcal{P}^*} \varphi_P$. We also define a characteristic function $\varphi(c_1, \dots, c_{m'}; R)$ as follows:

$$\begin{aligned} \varphi(c_1, \dots, c_{m'}; R) &= 1 && \text{if } \bigcup_{P \in \mathcal{P}^*} [S_{rP} \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r)] \neq \phi \\ &= 0 && \text{if } \bigcup_{P \in \mathcal{P}^*} [S_{rP} \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r)] = \phi. \end{aligned}$$

We calculate the probability measure of the set S , letting p stand for the p.d.f. under $P_1(H_1)$.

$$\begin{aligned} \mu(S) &= \int_{R^{m'+n'+1}} \varphi_S p \, dT \prod_i dz_i \prod_j dz_j^* = \int_{R^{m'+n'+1}} \left(\sum_P \varphi_P \right) p \, dT \prod_i dz_i \prod_j dz_j^* \\ &= \sum_P \int_{R^{m'+n'+1}} \varphi_P p \, dT \prod_i dz_i \prod_j dz_j^* \\ &= \sum_P \int_{R'} dr \int_{S_r} A(r) \, dm \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \varphi_P \, d\lambda \\ &= \int_{R'} dr \int_{S_r} A(r) \, dm \sum_P \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \varphi_P \, d\lambda \\ &= \int_{R'} dr \int_{S_r} A(r) \, dm \varphi(c_1, \dots, c_{m'}; r) \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \, d\lambda \\ &= \int_{R'} dr \int_{S_r} A(r) \varphi(c_1, \dots, c_{m'}; r) p_{c_1, \dots, c_{m'}; r} \, d\lambda, \end{aligned}$$

where $A(r)$ is the area of S_r and where we let

$$p_{c_1, \dots, c_{m'}; r} = \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \, d\lambda,$$

that is, the marginal probability density of $c_1, \dots, c_{m'}, T$. This density is easily seen to be a spherically symmetric normal density centered on the origin with variance $\sigma^2 + \sigma^{*2}$. Since it is constant valued over spheres centered at the origin, let $\Gamma(r) = A(r)p_{c_1, \dots, c_{m'}; r}$ and we obtain

$$\begin{aligned} \mu(S) &= \int_{R'} \Gamma(r) \, dr \int_{S_r} \varphi(c_1, \dots, c_{m'}; r) \, dm \\ &= \int_{R'} \Gamma(r) m(S_r) \, dr = \int_{R'} \Gamma(r) \alpha \, dr = \alpha. \end{aligned}$$

Hence the probability measure of S is α , and is independent of the parameters σ^2 and σ^{*2} of the distribution. This completes the proof.

THEOREM 3. *If the region S is similar of size α and satisfies Condition 4, then S satisfies Condition 5.*

PROOF. Using the notion of Theorem 2, we have

$$\alpha = \mu(S) = \int_{R^{m'+n'+1}} \varphi_S p \, dT \prod_i dz_i \prod_j dz_j^*.$$

All steps in the previous theorem remain valid under the assumptions of this theorem except the last two. The value of the integral $\int_{S_r} \varphi(c_1, \dots, c_{m'}; r) \, dm$

may now be a function of r , say $v(r)$. We have therefore $\alpha = \int_{R'} \Gamma(r)v(r) \, dr$.

Since the Gamma densities form a complete system of functions, with parameter $\sigma^2 + \sigma^{*2}$, $v(r)$ must be equal to α almost everywhere. But since this implies the fulfillment of Condition 5, the theorem is proved.

4.3. Choice of the test. From the class \mathcal{C}_α of similar regions, we wish to choose an optimum test of the hypothesis H_1 . For this we require the following theorem.

THEOREM 4. *For any region of the class \mathcal{C}_α , there corresponds a region in \mathcal{C}_α with the same power function and having the structure*

$$S = \bigcup_{r \in R'} \bigcup_{(c_1, \dots, c_{m'}; T) \in C_r} \mathcal{H}_P(c_1, \dots, c_{m'}; r)$$

where C_r is a subset of S_r for each r .

PROOF. Let $S' \in \mathcal{C}_\alpha$. From the definition of \mathcal{C}_α , there exists a measurable set $S'' \in \mathcal{C}_\alpha$ which is essentially equivalent to S' . S'' satisfies Condition 4, which is a relation to be satisfied for each $c_1, \dots, c_{m'}, r$. But $c_1, \dots, c_{m'}$, being fixed implies T is fixed.

Under an alternative hypothesis the p.d.f. " p " is changed by a factor depending only on $\sigma^2 + \sigma^{*2}$, T , and d . This factor is a constant over all points in the ranges of integration in Condition 4. It follows then that Condition 4 being

true on the hypothesis H_1 implies that it is true under any alternative hypothesis.

To construct S we arbitrarily choose a $P \in \mathcal{P}^*$ and using the structure of S'' given by Conditions 1-5, we choose

$$C_r = \left\{ (c_1, \dots, c_{m'}; T) \mid \bigcup_{P \in \mathcal{P}^*} [S''_P \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r)] \neq \emptyset \right\},$$

$$S = \bigcup_{r \in R'} \bigcup_{(c_1, \dots, c_{m'}; T) \in C_r} \mathcal{H}_P(c_1, \dots, c_{m'}; r).$$

It follows easily that Conditions 1-5 are fulfilled for the set S . Therefore $S \in \mathcal{C}''_\alpha$.

To show that $S \in \mathcal{C}_\alpha$ we must prove that S is a measurable set. S is a cylindrical set with a cross section or base set \tilde{S} defined in $R^{m'+1}$, the space of $c_1, \dots, c_{m'}, T$.

$$\tilde{S} = \bigcup_{r \in R'} S_r$$

$$S_r = \{ (c_1, \dots, c_{m'}; T) \mid S \cap \mathcal{H}_P(c_1, \dots, c_{m'}; r) \neq \emptyset \}.$$

It therefore suffices to show that this base set \tilde{S} is measurable. This can be established by the following simple argument.

For the region S'' , define as in Theorem 3 regions S''_P and characteristic functions φ'' and φ''_P . Then

$$\begin{aligned} \mu(S'') &= \int_{R^{m'+n'+1}} \varphi'' p \, dT \prod_i dz_i \prod_j dz_j^* \\ &= \sum_{P \in \mathcal{P}^*} \int_{R^{m'+n'+1}} \varphi''_P p \, dT \prod_i dz_i \prod_j dz_j^* \\ &= \int_{r \in R'} dr \int_{S_r} A(r) \, dm \sum_{P \in \mathcal{P}^*} \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} \varphi''_P p \, d\lambda. \end{aligned}$$

Then applying Condition 4 we have

$$\begin{aligned} \mu(S'') &= \int_{r \in R'} dr \int_{S_r} A(r) \, dm \varphi(c_1, \dots, c_{m'}; r) \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \, d\lambda \\ &= \int_{r \in R'} dr \int_{S_r} A(r) \varphi(c_1, \dots, c_{m'}; r) p_{c_1, \dots, c_{m'}; r} \, dm \\ &= \int_{R^{m'+1}} \varphi(c_1, \dots, c_{m'}, r) p_{c_1, \dots, c_{m'}; r} \, dT \prod_1^m dc_i. \end{aligned}$$

But since $\mu(S'')$ is equal to this last expression, it follows that $\varphi_{(c_1, \dots, c_{m'}; r)}$ is measurable. However, $\varphi(c_1, \dots, c_{m'}; r)$ is the characteristic function of the set \tilde{S} and therefore S is measurable and $S \in \mathcal{C}$.

We have now only to establish that the power function for the region S is identical to the power function of S'' . We shall need to know that Condition 4

is fulfilled under alternative hypotheses; this we have already shown. Letting P be the p.d.f. of $P_1(H_1 \text{ or } \bar{H}_1)$, we have

$$\begin{aligned}\mu(S'') &= \int_{R^{m'+n'+1}} \varphi'' p \, dT \prod_i dz_i \prod_j dz_j^* \\ &= \int_R dr \int_{S_r} A(r) \, dm \sum_{P \in Q^*} \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} \varphi'' p \, d\lambda \\ &= \int_R dr \int_{S_r} A(r) \, dm \varphi(c_1, \dots, c_{m'}; r) \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} p \, d\lambda \\ &= \int_R dr \int_{S_r} A(r) \, dm \int_{\mathcal{H}_P(c_1, \dots, c_{m'}; r)} \varphi_P(S) p \, d\lambda \\ &= \int_{R^{m'+n'+1}} \varphi_P(S) p \, dT \prod_i dz_i \prod_j dz_j^* = \mu(S).\end{aligned}$$

This completes the proof.

Thus for tests of size α with critical regions chosen from the class \mathcal{C}_α , we can get as good tests by confining our attention to critical regions of the form

$$S = \bigcup_{r \in R'} \bigcup_{(c_1, \dots, c_{m'}, T) \in C_r} \mathcal{H}_P(c_1, \dots, c_{m'}; r).$$

Regions of this type are cylindrical with base in the space $R^{m'+1}$ of $c_1, \dots, c_{m'}, T$ and axes the coordinate axes of the remaining variables. Whether a sample point in $R^{m'+n'+1}$ belongs to S can be ascertained by observing whether

$$c_1, \dots, c_{m'}, T$$

falls in the base set in $R^{m'+1}$. It is worth noticing that the size and power of regions of this type are independent of the particular P used in the defining equation above.

Our problem has thus been reduced to the following: $\bar{P}_1 T, c_1, \dots, c_{m'}$ are normally and independently distributed with means $d, 0, \dots, 0$ and common variance $\sigma^2 + \sigma^{*2}$. The hypothesis is $H_1 = \{d = 0\}$. This is the simplest 'Student' problem.

The choice of a similar test for this problem \bar{P}_1 depends on whether the alternatives are one- or two-sided and will be the usual t -test, one- or two-sided. Thus among similar tests of H_1 with critical region restricted to the class \mathcal{C}_α , we have found a most powerful or most powerful unbiased test. This test is identical to that proposed by Barankin [1].

5. Calculation of the Test Criterion for P_1 . Rather than give an explicit expression for the test criterion, we describe in terms usual to the analysis of variance the procedure for determining it. This will avoid errors in substituting in an unwieldy expression and the possibility of typographical errors in such expressions (for example, in formulas (37) and (39) in [1] the sign of the second term in the denominator should be negative).

We need first the linear function contained in the numerator of the t -criterion. This is to within a constant factor the difference between the regression coefficient for \bar{y} in the first sample and the corresponding coefficient in the second sample. These coefficients are independent of whether the \bar{y} vectors have been orthogonalized to the corresponding \bar{x} vectors: we need only the coefficients in the joint regression. Calculate the variance of the difference between these two coefficients: it will be of the form $a\sigma^2 + b\sigma^{*2}$.

For the calculation of the denominator, we distinguish two cases. In each delete $n - m$ members of the second sample.

I. The vectors $\bar{y}, \bar{x}_1, \dots, \bar{x}_p$ generate the same space as $\bar{y}^*, \bar{x}_1^*, \dots, \bar{x}_p^*$. In the two samples of size m fit by regression the vectors $\bar{y}, \bar{x}_1, \dots, \bar{x}_p$ (or equivalently $\bar{y}^*, \bar{x}_1^*, \dots, \bar{x}_p^*$). Proceed as in the analysis of covariance; calculate the sum of squares (SS) of the U 's, the sum of products (SP) of the U 's and V 's, and the sum of squares of the V 's. Then using the regression coefficients as obtained from the samples of size m , calculate SS_U, SP_{UV} , and SS_V for residuals. The denominator of the test criterion will be:

$$(5.1) \quad \sqrt{\frac{aSS_U + 2(ab)^{\frac{1}{2}}SP_{UV} + bSS_V}{m - p - 1}}.$$

II. The vectors $\bar{y}, \bar{x}_1, \dots, \bar{x}_p$ do not generate the same space as $\bar{y}^*, \bar{x}_1^*, \dots, \bar{x}_p^*$ do. Choose a linearly independent set of vectors $\bar{w}_1, \dots, \bar{w}_t$ which generate the space spanned by the combined set of vectors $\bar{y}, \bar{y}^*, \bar{x}_1, \dots, \bar{x}_p, \bar{x}_p^*$ and which in addition satisfy the conditions:

$\bar{w}_1, \dots, \bar{w}_{p+1}$ generate the space spanned by $\bar{y}, \bar{x}_1, \dots, \bar{x}_p$,

$\bar{w}_{t-p}, \dots, \bar{w}_t$ generate the space spanned by $\bar{y}^*, \bar{x}_1^*, \dots, \bar{x}_p^*$.

As a consequence of these conditions $\bar{w}_{t-p}, \dots, \bar{w}_{p+1}$ generate the intersection of the spaces mentioned in the two conditions above. Calculate the regression coefficients obtained from fitting the vectors $\bar{w}_1, \dots, \bar{w}_t$ to the first sample. In doing so, fit the first $(p + 1)$ w 's and then record $l_1^2, \dots, l_{t-p-1}^2$, the successive amounts by which the sums of squares of residuals is reduced by fitting in addition $\bar{w}_{p+2}, \dots, \bar{w}_t$ one by one. Let SS_U be the final sum of squares of residuals. Repeating the above procedure for the second sample of m , but using the \bar{w} vectors in the order $\bar{w}_t, \bar{w}_{t-1}, \dots, \bar{w}_1$, obtain $l_1^{*2}, \dots, l_{t-p-1}^{*2}, SS_V$. Calculate the overall sum of products and using the two sets of regression coefficients on all the \bar{w} 's, obtain SP_{UV} the sum of products of residuals. The denominator of the test criterion will be

$$(5.2) \quad \sqrt{\frac{aSS_U + 2(ab)^{\frac{1}{2}}SP_{UV} + bSS_V + \sum_{i=1}^{t-p-1} (a^{\frac{1}{2}}l_i + b^{\frac{1}{2}}l_i^*)^2}{m - p - 1}}$$

where l_i, l_i^* are given the signs of the corresponding regression coefficients.

It is easily seen that formulas (5.1) and (5.2) give the denominators for the optimum test criterion as described in Section 4.

6. The problem P_q , $q > 1$. No attempt has been made to obtain a general class of similar regions for $P_q(q > 1)$ except in a very particular case described below. The asymptotically best test in the sense that the F ratio is best in the analysis of variances (see Wald [6]) is, however, immediately available. To obtain this, calculate a t -criterion for each of the q regression coefficients, in each case putting the remaining $q - 1$ vectors with the \bar{x} vectors and proceeding as in Section 5. Each t^2 has asymptotically the χ^2_1 distribution and is asymptotically independent of the other t^2 's. Combine the q t^2 's to obtain a criterion having asymptotically the χ^2_q distribution under H_q . Use large values of χ^2 to form the size α critical region.

In small samples a similar procedure is available if each numerator for the t criteria has the same ratio $b/a = k$ where a and b are given in Section 5. Divide each numerator by the corresponding $a^{\frac{1}{2}}$, and square and add to obtain the numerator of an F -ratio. For the denominator use the square of the t -ratio denominator where a and b are given the values 1 and k . This F -ratio will have the $F_{q, m-p-q}$ distribution under H_q and the usual power function under the alternatives. The procedure in Section 4 extends immediately; a general class of similar regions \mathcal{C}_α exists and its construction is almost identical to the P_1 construction. Among tests having regions belonging to \mathcal{C}_α , a best test in the Wald sense [6] exists and is the one described at the beginning of this paragraph.

If the ratios b/a are not constant corresponding to the q pairs of regression coefficients, the following procedure gives an exact test for small samples. Calculate the numerators for the t -ratios as described in Section 5. Also calculate the sums of squares and products as used in formulas (5.1) and (5.2). Partition these sums of squares and products into q sets having n_1, \dots, n_q degrees of freedom ($\sum n_i = m - p - q$). Next evaluate q t -ratios using for the i th denominator the expression

$$\sqrt{\frac{a_i SS_U(n_i) + 2(a_i b_i)^{\frac{1}{2}} SP_{UV}(n_i) + b_i SS_V(n_i)}{n_i}}.$$

From each t_i^2 using F tables with 1 and n_i degrees of freedom, calculate p_i the probability of a larger value of t_i^2 under H_q . Under H_q each p_i will be uniformly and independently distributed $[0, 1]$. $\chi^2 = \sum_1^q -2 \ln p_i$ can then be used as a test criterion; under H_q χ^2 has a χ^2 distribution with $2q$ degrees of freedom and under alternative hypothesis large values of χ^2 become relatively more likely. See Fisher [5]. For tests of this type the power curve along the axes of d_1, \dots, d_q in the parameter space depends respectively and only on n_1, \dots, n_q (see Barankin [1] p. 435). Using considerations similar to those used to justify the principle of invariance, it seems reasonable to maximize the lowest curve; that is, we choose the smallest value of n_i as large as possible. We therefore choose $n_i =$

$[m - p - q/q]$ for as many i as possible subject to the remaining n_i satisfying $n_i = [m - p - q/q] + 1$.

Other possible ways of combining individual tests are discussed in [7].

REFERENCES

- [1] E. W. BARANKIN, "Extension of the Romanovsky-Bartlett-Scheffé test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949.
- [2] H. SCHEFFÉ, "On the solutions of the Behrens-Fisher problem based on the t -distribution," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 35-44.
- [3] W. V. BEHRENS, "Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen," *Landw. jrb.*, Vol. 68 (1929), pp. 807-837.
- [4] UTTAM CHAND, "Distributions related to comparison of two means and regression coefficients," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 507-521.
- [5] R. A. FISHER, *Statistical Methods for Research Workers*, Oliver and Boyd, 1948.
- [6] ABRAHAM WALD, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 434-439.
- [7] ALLEN WALLIS, "Compounding probabilities from independent significance tests," *Econometrica*, Vol. 10 (1942), pp. 229-248.