The Conditional Level of the F–Test

RICHARD A. OLSHEN *

Suppose that in regression problems the simultaneous confidence intervals of the *S*-method are used only when a preliminary *F*-test rejects the null hypothesis that the regression parameters are zero. (Some of Scheffé's work emphasizes this usage [12, p. 87; 13, p. 66].) The probability of coverage should then be conditioned on the rejection. That for sufficiently large critical values and at least two degrees of freedom for error the conditional probability of simultaneous coverage is always smaller than the unconditional probability is established in this article. Also included are Monte Carlo studies of the discrepancy and interpretation.

1. INTRODUCTION

The S-method of multiple comparisons possesses a distinctive feature which has commended its use in a two-stage procedure. Specifically, the usual F-test (of the null hypothesis that all regression parameters are zero) accepts at level α if, and only if, all the $100(1-\alpha)$ percent confidence intervals formed in accordance with the method cover zero. Thus, a putatively reasonable procedure is as follows (see [12, p. 87]). Perform an F-test to see whether the vector of regression parameters is significantly different from $\mathbf{0}$ at level α. If it is not, go no further. Otherwise, form simultaneous $100(1-\alpha)$ percent confidence intervals à la Scheffé for any linear combinations of parameters deemed interesting.

If the S-intervals are reported in summarizing data *only* when a preliminary F-test rejects the null hypothesis, then the frequentistic justification for their customary confidence coefficients is not applicable, for the probability of simultaneous coverage is computed unconditionally. It is impossible to assign precisely a conditional probability of coverage, given the F-test has rejected the null hypothesis, because that conditional probability depends on unknown parameters. This difficulty would be ameliorated if the conditional probability (of simultaneous coverage) always exceeded the unconditional one, or if at least for some values of the parameters the conditional probability was the larger. The present article demonstrates that for a broad spectrum of experimental designs and critical values, the conditional probability of simultaneous coverage is less than the nominal $1-\alpha$ for all values of the unknown parameters.

Diverse results on conditional confidence coefficients were obtained by Buehler [2], Wallace [19], Stein [16], and Cornfield [5]. Two direct predecessors of the present article pertain to the t-test. In describing them I denote by $\bar{x}$ the sample mean and by s the sample standard deviation of a random sample of observations from $N(\mu, \sigma^2)$. Buehler and Fedderson [3] showed for samples of size two and all (μ, σ^2)

$$P_{\mu,\sigma^2}(|\bar{x}-\mu| < ks \,\big|\, |\bar{x}-\mu| < \tfrac{3}{2}) - P_{\mu,\sigma^2}(|\bar{x}-\mu| < ks) \geq .0181, \quad (1.1)$$

where k is the 75th percentile of the Cauchy distribution.[1] Brown [1] substantially extended their results; in particular his conclusions pertain to all sample sizes exceeding one. He fixed $k > 0$ and $0 < \alpha < 1$ to satisfy

$$P_{\mu,\sigma^2}(|\bar{x}-\mu| < ks) \equiv 1-\alpha$$

and showed that if $c > k^{-1}(\sqrt{1+k^2}+1)$, then the

$$\inf_{\mu,\sigma^2} P_{\mu,\sigma^2}(|\bar{x}-\mu| < ks \,\big|\, |\bar{x}| \leq cs) > 1-\alpha. \quad (1.2)$$

Brown notes that (1.2) is false for $c < k^{-1}$, thus for $k < 1$ if $c = k$. From (1.2) it follows that for c and k in the required relationship,

$$P_{\mu,\sigma^2}(|\bar{x}-\mu| < ks \,\big|\, |\bar{x}| > cs) < 1-\alpha \quad \text{for all } \mu \text{ and } \sigma^2; \quad (1.3)$$

the inequality is pointwise and not uniform [16]. Brown reports on the magnitude of the inequality in (1.2) for samples of size two. For example, if $\alpha = \frac{1}{2}$, $k = 1$, and $c = 1+\sqrt{2}$, the left-hand quantity is $\frac{2}{3}$. The statement (1.3) can be phrased in terms of the frequentistic confidence intervals given by the S-method ([12, 13]): when the vector of regression parameters is one-dimensional, the conditional probability of coverage, given that a preliminary "F-test" rejects the hypothesis that the parameter vector is zero, is less than the unconditional probability.

In Section 2 I specialize Brown's study to the most important special case, $c = k$, and require the degrees of freedom of the denominator of the F-statistic to be at

* Richard A. Olshen is associate professor, Departments of Statistics and Mathematics, The University of Michigan, Ann Arbor, Mich. 48104. Work on this article was supported in part by National Science Foundation Grant GP-24918 at Stanford University and in part by a grant from the Horace H. Rackham School of Graduate Studies at the University of Michigan; also, part of this work was completed at Columbia University, and part at the Hebrew University. The author is grateful to many people for helpful conversations and correspondence during the preparation of this article. Among them are L. D. Brown, Bradley Efron, Ray Faith, Ruben Gabriel, Vernon Johns, Rupert Miller, Vivian Olshen, Henry Scheffé, Stanley Sclove, Michael Woodroofe, and a very helpful referee. The programming of the Monte Carlo experiments was done by Ann Varady of the Stanford Computation Center. This article is dedicated to the memory of Vivian.

[1] Cornfield [5, p. 633] emphasized that Fisher [8, p. 81] had asserted the impossibility of Buehler and Fedderson's discovery. In fact, Fisher's argument pertains to 1-sided confidence sets, in which scenario it is also incorrect. Further, there is no evidence Fisher thought the 1- and 2-sided cases to be different.

© Journal of the American Statistical Association
September 1973, Volume 68, Number 343
Theory and Methods Section

least two; the degrees of freedom of the numerator is arbitrary. For sufficiently large critical values, results like those of Brown are obtained; my arguments are patterned after his. The proof is postponed to the appendix.

Section 3 is a summary of a Monte Carlo experiment which demonstrates how very discrepant the conditional and unconditional probabilities of coverage often can be. (The left side of (1.3) is zero if $\mu = 0$.)

The fourth section contains tentative explanations and further questions pertaining to the material of Sections 2 and 3.

2. *F* DISTRIBUTIONS AND THE S-METHOD

In what follows $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ refers to the (possibly multivariate) normal distribution with expectation $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; $F(a, b)$ refers to the F distribution with a and b degrees of freedom. The symbol $\|\cdot\|$ will be used for Euclidean length; the dimension of the space in which a given vector sits is implicit, but obvious. A tilde which is not a subscript means "is distributed as."

Now assume $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is a known $n \times m$ matrix, $n \geq m$. For simplicity, assume $\mathbf{X}$ is of rank m. Entries of the vector $\boldsymbol{\beta}$ are parameters of interest, and inferences about them are to be made from $\mathbf{Y}$. Put $\mathbf{D} = (\mathbf{X}'\mathbf{X})^{-1}$, where the prime denotes transpose. The least squares and Gauss-Markov estimate of $\boldsymbol{\beta}$ calculated from $\mathbf{Y}$ is $\boldsymbol{\beta} = \mathbf{D}\mathbf{X}'\mathbf{Y}$, while the usual residual mean square is $s^2 = (1/n - m)\mathbf{Y}'(\mathbf{I} - \mathbf{X}\mathbf{D}\mathbf{X}')\mathbf{Y}$; $\hat{\boldsymbol{\beta}}$ and s^2 are independent (see [11]), so that

$$\mathfrak{F} \overset{\text{def}}{=} \frac{(\boldsymbol{\beta} - \boldsymbol{\beta})'\mathbf{D}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta})/m}{s^2} \sim F(m, n - m).$$

A test of the null hypothesis $\boldsymbol{\beta} = \mathbf{0}$ is commonly performed as follows. The upper α-point F_α of $F(m, n - m)$ is found from tables. And the null hypothesis is rejected at level α if $\hat{\boldsymbol{\beta}}'\mathbf{D}^{-1}\hat{\boldsymbol{\beta}}/(ms^2) > F_\alpha$.

In this paragraph $\mathbf{U}$ denotes an m-dimensional vector. Scheffé established [12] that (in the notation of [11])

$$\mathfrak{F} \leq F_\alpha \Longleftrightarrow \forall \mathbf{U},\ \mathbf{U}'\boldsymbol{\beta} \in [\mathbf{U}'\hat{\boldsymbol{\beta}} - s\sqrt{mF_\alpha \mathbf{U}'\mathbf{D}\mathbf{U}},\ \mathbf{U}'\hat{\boldsymbol{\beta}} + s\sqrt{mF_\alpha \mathbf{U}'\mathbf{D}\mathbf{U}}]$$

(see also [10]) from which it follows that

$$P_{\beta,\sigma^2}(\forall \mathbf{U},\ \mathbf{U}'\hat{\boldsymbol{\beta}} - s\sqrt{mF_\alpha \mathbf{U}'\mathbf{D}\mathbf{U}} \leq \mathbf{U}'\boldsymbol{\beta} \leq \mathbf{U}'\hat{\boldsymbol{\beta}} + s\sqrt{mF_\alpha \mathbf{U}'\mathbf{D}\mathbf{U}}) \equiv 1 - \alpha. \quad (2.1)$$

This enables a statistician to form frequentistic confidence intervals for each linear combination of the parameters, whose validity is not impaired if the data were examined to find interesting linear combinations. These intervals are notoriously conservative for any such combination, or estimable function, which was fixed in advance. In terms of the foregoing definitions and summaries it is easy to state my main result as follows.

Theorem 1: If $(m/n - m)F_\alpha$ is at least three and $n - m$ is at least two, then

$$P_{\beta,\sigma^2}\left(\mathfrak{F} \leq F_\alpha \,\middle|\, \frac{\hat{\boldsymbol{\beta}}\mathbf{D}^{-1}\hat{\boldsymbol{\beta}}/m}{s^2} > F_\alpha\right) < 1 - \alpha \quad \text{for all } \boldsymbol{\beta} \text{ and } \sigma^2. \quad (2.2)$$

To prove Theorem 1 and to study the magnitudes of the inequalities (2.2) for various $\boldsymbol{\beta}$ and σ^2, it is helpful to state the theorem differently. To that end, let $\mathbf{D}^{\frac{1}{2}}$ be any symmetric square root of $\mathbf{D}$, and $\mathbf{Z} \sim N[(1/\sigma)\mathbf{D}^{-\frac{1}{2}}\boldsymbol{\beta}, \mathbf{I}]$. Then $\boldsymbol{\beta}'\mathbf{D}^{-1}\boldsymbol{\beta} \sim \sigma^2\|\mathbf{Z}\|^2$, and $\|\mathbf{Z}\|^2$ has a non-central chi-square distribution with m degrees of freedom and non-centrality parameter $\boldsymbol{\beta}'\mathbf{D}^{-1}\boldsymbol{\beta}/\sigma^2$; define the latter parameter to be G^2. Thus $\boldsymbol{\beta}'\mathbf{D}^{-1}\hat{\boldsymbol{\beta}} \sim \sigma^2\{x_1^2 + \cdots + x_m^2\}$, where $x_1, \cdots, x_m$ are independent normal variables with variance 1, $E(x_1) = G$, and $E(x_i) = 0$ for $i \geq 2$. Also, $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{D}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \sigma^2\{(x_1 - G)^2 + x_2^2 + \cdots + x_m^2\}$, and $(n - m)s^2 \sim \sigma^2 \sum_{j=m+1}^n x_j^2$, where $x_{m+1}, \cdots, x_n$ are independent, standard normal variables, independent of $x_1, \cdots, x_m$. It follows that the left-hand expression in (2.2) can be written

$$P_G\left(\frac{(x_1 - G)^2 + x_2^2 + \cdots + x_m^2}{x_{m+1}^2 + \cdots + x_n^2} \leq k^2 \,\middle|\, \frac{x_1^2 + \cdots + x_m^2}{x_{m+1}^2 + \cdots + x_n^2} > k^2\right), \quad (2.3)$$

where the x's and G are as above, and $k^2 = (m/n - m)F_\alpha$; $P_G(\cdot)$ means that the probability or conditional probability in the parentheses is calculated supposing $E(x_1) = G$, $E(x_i) = 0$ for $i \geq 2$. Now Theorem 1 can be rephrased in an equivalent form.

Theorem 2: Assume that $1 \leq m < n$ are fixed, $n - m \geq 2$, and that $x_1, \cdots, x_n$ are as in the previous paragraph. Put $S^2 = x_{m+1}^2 + \cdots + x_n^2$. Let $k^2 \geq 3$ and $0 < \alpha < 1$ satisfy

$$P_G\left(\frac{(x_1 - G)^2 + x_2^2 + \cdots + x_m^2}{S^2} > k^2\right) = \alpha.$$

Then

$$P_G\left(\frac{(x_1 - G)^2 + x_2^2 + \cdots + x_m^2}{S^2} \leq k^2 \,\middle|\, \frac{x_1 + \cdots + x_m^2}{S^2} \leq k^2\right) > 1 - \alpha. \quad (2.4)$$

The proof of Theorem 2 is postponed to the appendix. Section 4 contains plausibility arguments for Theorem 1 (and therefore Theorem 2) and for related but as yet unproved results. It is consistent with those arguments and easily demonstrated that the theorems are false when there are infinitely many degrees of freedom for the denominator of the F-statistic. In that case the F-test is a χ^2 test.

A COMPARISON OF THE CONDITIONAL AND UNCONDITIONAL PROBABILITIES OF COVERAGE

Experiment number	Degrees of freedom Numerator (m)	Degrees of freedom Denominator (n-m)	k^2	Non-centrality parameter (G^2)	Empirical probability of coverage[c]	Empirical conditional probability of coverage given F rejects[d]
1	1	5	1.3200	1	.950	.764
2	1	5	1.3200	9	.950	.950
3[a]	1	5	3.2520	1	.992	.863
4[a]	1	5	3.2520	1	.990	.843
5[a]	1	5	3.2520	25	.989	.987
6	1	10	.4960	1	.951	.819
7	1	20	.2175	1	.950	.829
8[b]	1	20	.2175	9	.950	.965
9[a]	5	5	10.9700	9	.991	.892
10	5	20	1.0250	9	.990	.963
11[a]	10	5	9.4800	1	.951	.242
12[a]	10	5	9.4800	9	.952	.705
13[a]	10	5	9.4800	25	.949	.871
14[a]	10	5	20.2000	9	.991	.746
15[a]	20	5	18.2400	1	.950	.128
16	20	20	2.1200	9	.952	.768

[a] Covered by the theorem.

[b] This experiment is unique in that the empirical conditional probability is higher than the empirical unconditional probability; note the low value of k^2.

[c] In those experiments for which the empirical probability of coverage is .95 (.99) to two decimal places, k^2 was chosen to correspond to the 95% (99%) point of the appropriate F-distribution.

[d] For fixed m, $n - m$, and k^2 the conditional probability appears to be an increasing function of G. As G tends to ∞ the conditional probability of coverage tends to the unconditional probability, for then the power of the F test tends to 1.

3. MONTE CARLO STUDIES

To assess the magnitude of the inequality in (2.4), 16 Monte Carlo experiments were performed. The experiments are described in the notation of (2.3) and (2.4). Let

$$\frac{(x_1 - G)^2 + x_2^2 + \cdots + x_m^2}{x_{m+1}^2 + \cdots + x_n^2} = F_1$$

and

$$\frac{x_1^2 + \cdots + x_m^2}{x_{m+1}^2 + \cdots + x_n^2} = F_2.$$

In each experiment m, n, G, and k^2 were fixed; $x_1, \cdots, x_n$ were generated. Then it was determined whether or not $F_1 > k^2$ and whether or not $F_2 \leq k^2$. There were 10,000 independent replications (of each experiment). None of the pseudo-random normal deviates was used twice. For each experiment the numbers N_1, N_2, and N_{12} were computed, where N_1 is the number of replications for which $F_1 \leq k^2$; N_2 is the number of replications for which $F_2 \leq k^2$; N_{12} is the number of replications for which both $F_1 \leq k^2$ and $F_2 > k^2$. The empirical probability of coverage of the S-intervals is $N_1/10{,}000$. The empirical conditional probability of coverage given the F-test rejects the null hypothesis is $N_{12}/(10{,}000 - N_2)$. The estimates of the unconditional probabilities, compared with the known probabilities of coverage, served as checks on the calculation. Two of the experiments have identical values of m, n, G^2, and k^2. It is trivial and uninteresting that both theoretical and empirical conditional probabilities of coverage given F rejects are zero if $G = 0$. Bear in mind that Theorem 2 has been proved only for $k^2 \geq 3$ (and $n - m \geq 2$) and that its conclusions are false for $m = 1$ and $k < 1$ (simultaneously).

4. CONCLUSIONS AND SPECULATIONS

The inequality (2.2) is easy to intuit. F rejects the hypothesis $\|\boldsymbol{\beta}\| = 0$ when the mean square for error is sufficiently small relative to the mean square for regression. In that circumstance the S-intervals are short, and so they are not as likely to cover their true values as they are unconditionally.

The conditions of Theorem 1 are more easily satisfied the larger the ratio of degrees of freedom: numerator to denominator. Naturally, whether or not the S-intervals cover their true values is a function of their centers and their lengths. The centers are determined by the Gauss-Markov estimates, and the lengths by the residual mean square. The estimates are stochastically farther (for fixed numerator degrees of freedom) from their true values the smaller the denominator degrees of freedom. But then, plausibly, the phenomenon of the preceding paragraph is more pronounced.

That the numerator of F_1 is a chi-square variable is not important to either of the previous points. So, e.g., I expect that conclusions like those of Theorem 1 extend to Tukey's T-method as well (see [10, 13]). Also, many questions pertaining to the S-method remain. For reasonable values of m, n, k^2, and G, how much longer must be the S-intervals (conditional on rejecting the hypothesis $\|\boldsymbol{\beta}\| = 0$) to increase the probability of simultaneous coverage to the unconditional probability? How many intervals must be used in order that the conditional probability of simultaneous coverage by those intervals actually be lower than the overall unconditional probability? The S-intervals are notoriously conservative for fixed estimable functions. How conservative are they for fixed functions, conditional on F rejecting the hypothesis $\|\boldsymbol{\beta}\| = 0$?

The inequality (2.2) is not limited to the preliminary test $\boldsymbol{\beta} = 0$, but holds for any fixed, preliminary test $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. I believe interpretation of the S-intervals is difficult when one preliminary F-test of the hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_1$ accepts while another preliminary F-test of the hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$ rejects; this is what happens for *every* datum vector $\mathbf{Y}$.

The notion of preliminary testing is familiar in the study of estimation as well as in the study of confidence intervals. Positive part Stein estimates, which improve on the Gauss-Markov estimate of $\boldsymbol{\beta}$ when $m \geq 3$ (see, e.g., [18]), involve an implicit preliminary test of the hypothesis $\boldsymbol{\beta} = \mathbf{0}$ (or any other fixed vector). A similar remark pertains to his estimate of σ^2 which improves on the best invariant estimate when $\boldsymbol{\beta}$ is unknown [17]. See also Cohen's recent work [4] on confidence sets for σ^2.

In the article by Sclove, Morris, and Radhakrishnan [15], Stein's estimates are shown to be better than naive preliminary-test estimates in regression problems for which loss is quadratic.

The conclusion for Theorem 1 is related to previous studies of coherent betting behavior. For example, suppose a master of ceremonies picks $\boldsymbol{\beta}$ and σ^2 and an observation from $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. And suppose a bookie offers odds $\alpha/1-\alpha$ against the event $\mathfrak{F} \le F_\alpha$. Then you may put $\$\alpha D$ in escrow, after which the master of ceremonies will reveal whether $\mathfrak{F} \le F_\alpha$. If $\mathfrak{F} \le F_\alpha$ the bookie gets your money, whereas if $\mathfrak{F} > F_\alpha$ you receive $\$D$ from the bookie. It is clear that if the conditions of Theorem 1 are met, and if the master of ceremonies makes public in advance of your betting whether or not the α-level F-test rejects H_0: $\boldsymbol{\beta} = \mathbf{0}$, and if you bet only when the F-test rejects H_0, then the bet is favorable to you no matter what $\boldsymbol{\beta}$ and σ^2 are.

No bettor can make favorable bets (for all parameter values) against a bookie who offers posterior odds of $\alpha/1-\alpha$ against credible sets [6, p. 213] of posterior probability $1-\alpha$ (see [2, 5, 19]). The bookie must use a proper prior distribution, but the bettor is allowed any preliminary test whatsoever.

A detailed study of coherent betting behavior in case both parameter space and sample space are finite has been made by Freedman and Purves [9]. The first theorem of their paper is equivalent to the result of Cornfield [5, Sect. 5]. L.D. Brown (in private communication) has extended the scope of the Freedman-Purves results.

APPENDIX

Proof of Theorem 2: It is enough to study (2.4) for a fixed $G > 0$. Define B^2, R, and, $\underset{\sim}{\theta}$ as follows:

$$B^2 = x_2^2 + \cdots + x_m^2$$

$$R = \sqrt{(x_1 - G)^2 + S^2}$$

$$\underset{\sim}{\theta} = \arctan \frac{S}{x_1 - G}$$

The square root is non-negative, and $0 \le \underset{\sim}{\theta} \le \pi$. R is a function of the length of the random vector $(x_1 - G, x_{m+1}, \cdots, x_n)$, and $\underset{\sim}{\theta}$ is a function of its direction, so B^2, R, and $\underset{\sim}{\theta}$ are independent. In the notation of indicator variables (2.4) says

$$E(I_{[x_1^2+B^2/S^2\le k^2]})E(I_{[(x_1-G)^2+B^2/S^2\le k^2]}) < E(I_{[x_1^2+B^2/S^2\le k^2]}I_{[(x_1-G)^2+B^2/S^2\le k^2]}). \quad \text{(A.1)}$$

Toward establishing (A.1) let, for fixed $b^2 \ge 0$,

$$\phi_1(b^2) = E(I_{[x_1^2+b^2/S^2\le k^2]}) \qquad \phi_2(b^2) = E(I_{[(x_1-G)^2+b^2/S^2\le k^2]}).$$

Then ϕ_1 and ϕ_2 are both decreasing in b^2, but are not proportional. So

$$\operatorname{Cov}[\phi_1(B^2), \phi_2(B^2)] > 0. \quad \text{(A.2)}$$

Because x_1, B^2, S^2 are independent

$$\phi_1(b^2) = E(I_{[x_1^2+B^2/S^2\le k^2]}|B^2 = b^2)$$
$$\phi_2(b^2) = E(I_{[(x_1-G)^2+B^2/S^2\le k^2]}|B^2 = b^2).$$

The conclusion of (A.2) is

$$E[E(I_{[(x_1-G)^2+B^2/S^2\le k^2]}|B^2)E(I_{[x_1^2+B^2/S^2\le k^2]}|B^2)] > P\left(\frac{(x_1-G)^2+B^2}{S^2} \le k^2\right)P\left(\frac{x_1^2+B^2}{S^2} \le k^2\right). \quad \text{(A.3)}$$

To prove the theorem it is now enough that the left side of (A.3) is not more than the right side of (A.1), which can be expressed

$$E[E(I_{[x_1^2+B^2/S^2\le k^2]}I_{[(x_1-G)^2+B^2/S^2\le k^2]}|B^2)].$$

The conclusion will follow if the two integrands are in the relationship wished for their expectations, i.e., if $\phi_1(b^2)\phi_2(b^2) \le \phi_3(b^2)$, where

$$\phi_3(b^2) = E(I_{[x_1^2+b^2/S^2\le k^2]}I_{[(x_1-G)^2+b^2/S^2\le k^2]}).$$

Brown's result [1] implies $\phi_1(0)\phi_2(0) < \phi_3(0)$. Moreover, ϕ_1, ϕ_2, ϕ_3 are all decreasing in b^2, and all continuous in b^2. If $\phi_i(\infty) = \lim_{b^2\to\infty}\phi_i(b^2)$, then $\phi_1(\infty) = \phi_2(\infty) = \phi_3(\infty) = 0$.

Fix an arbitrary $b^2 > 0$, and put $K' = \{S^2 > k^{-2}[(x_1-G)^2 + b^2]\}$, $C = \{S^2 > k^{-2}(x_1^2 + b^2)\}$, C is the "acceptance region of the α-level F test" when $B^2 = b^2$. Now and for the remainder of the proof refer to Figure A, in which the abscissa is x_1 and ordinate is S. C_1 is the intersection of the two graphs; it has coordinates $[G/2, k^{-1}\sqrt{(G^2/4)+b^2}]$. A_0 is the point $(G, 0)$. C_2 is the intersection of $\{(x_1, k^{-1}\sqrt{x_1^2+b^2})\}$ and $\{x_1 = G\}$. It has coordinates $(G, k^{-1}\sqrt{G^2+b^2})$. P is the point in $\{(x_1, k^{-1}\sqrt{x_1^2+b^2})\}$ closest to A_0; P has coordinates

$$\left(\frac{k^2G}{1+k^2}, k^{-1}\sqrt{\frac{k^4G^2}{(1+k^2)^2}+b^2}\right).$$

Because $k \ge \sqrt{3}$, $\overline{A_0C_1} \ge \overline{A_0C_2}$.

Let

$$p_1 = p_1(r, b, G) = P\left(K' \cap C \cap \left\{\underset{\sim}{\theta} \le \frac{\pi}{2}\right\}|R = r, B^2 = b^2\right),$$

$$p_2 = p_2(r, b, G) = P\left(C \cap \left\{\underset{\sim}{\theta} \le \frac{\pi}{2}\right\}|R = r, B^2 = b^2\right),$$

$$p_3 = p_3(r, b, G) = P\left(K' \cap C \cap \left\{\underset{\sim}{\theta} > \frac{\pi}{2}\right\}|R = r, B^2 = b^2\right),$$

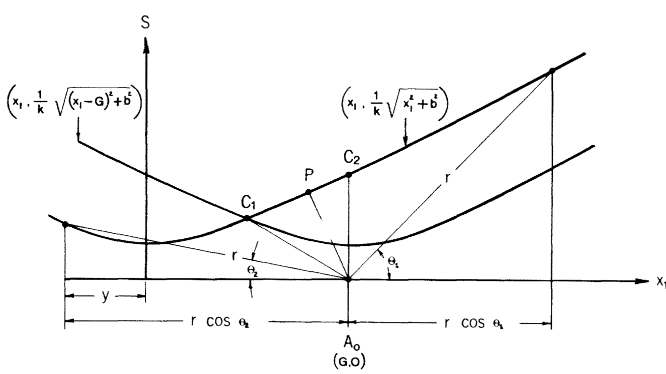
$$p_4 = p_4(r, b, G) = P\left(C \cap \left\{\underset{\sim}{\theta} > \frac{\pi}{2}\right\}|R = r, B^2 = b^2\right),$$

and $\rho = p_1 + p_3/p_2 + p_4$ when the ratio makes sense. Also, let $1 - \alpha_{r,b}$ be the conditional probability of K' given $R = r$ (and $B^2 = b^2$). Finally, let $g(r)$ be the density of R at r (that R^2 has a χ^2_{n-m+1} distribution is not pertinent to the calculation).

Now fix $b^2 > 0$; because B^2 and R are independent $\phi_1(b^2)\phi_2(b^2) \le \phi_3(b^2)$ is equivalent to this condition

$$\int [p_2(r, b, G) + p_4(r, b, G)]g(r)dr \int (1 - \alpha_{r,b})g(r)dr \le \int [p_1(r, b, G) + p_3(r, b, G)]g(r)dr. \quad \text{(A.4)}$$

A. GEOMETRY OF THEOREM 2



Suppose it could be shown that the left side of (A.4) is not more than

$$\int [p_2(r, b, G) + p_4(r, b, G)](1 - \alpha_{r,b})g(r)dr \qquad (A.5)$$

and that

$$1 - \alpha_{r,b} < \rho \qquad (A.6)$$

where $p_4 + p_4 > 0$. Then the expression of (A.5) is less than

$$\int [p_1(r, b, G) + p_3(r, b, G)]g(r)dr,$$

and under the stated contingencies the theorem is proven.

If $r \leq \overline{A_0P}$, $p_i = 0$, $i = 1, \cdots, 4$, and those values of r do not enter into the study of (A.6). Note that $p_1 = p_2$ for all r, b, G, and the common value is 0 for $r \leq \overline{A_0C_2}$, positive for $r > \overline{A_0C_2}$; it will be shown to be strictly increasing in r for $r > \overline{A_0C_2}$. Also, $p_3 = p_4$ for $r \leq \overline{A_0C_1}$. It follows that $\rho = 1$ for $r \leq \overline{A_0C_1}$. For $r > \overline{A_0C_1}$, $p_3 = (1 - \alpha_{r,b})/2$, and $p_4 \leq \frac{1}{2}$, as the distribution of $\underset{\sim}{\theta}$ is symmetric about $\pi/2$. These observations show that $\rho > 1 - \alpha_{r,b}$ for all $r > \overline{A_0P}$; and only one contingency remains.

Now suppose it could be shown that (for fixed $G \geq 0$) $p_2(r, b, G) + p_4(r, b, G)$ is non-decreasing in r. Then also $1 - \alpha_{r,b}$ is non-decreasing in r, for that is the case $G = 0$. But then the left side of (A.4) is not more than the expression of (A.5) because the covariance of two non-decreasing functions of R is non-negative. The remainder of the proof is devoted to showing that $p_2 + p_4$ is non-decreasing in r.

The two summands are studied separately, first p_4; p_4 is 0 for $r \leq \overline{A_0P}$, and so it is non-decreasing in r there. Next suppose $\overline{A_0P} < r \leq \overline{A_0C_2}$. There are two points $[x^{(1)}(r), S^{(1)}(r)]$ and $[x^{(2)}(r), S^{(2)}(r)]$, in $\{(x_1, k^{-1}\sqrt{x_1^2 + b^2})\} \cap \{x_1 < G\}$ at distance r from A_0. They satisfy $k^2(x^{(i)} - G)^2 + (x^{(i)})^2 + b^2 = k^2r^2$, $i = 1, 2$. Without loss assume that $x^{(1)} < x^{(2)}$. In the region being studied (see Figure B)

$$p_4(r, b, G)$$
$$= P\left(\frac{\pi}{2} + \arcsin\frac{G - x^{(2)}}{r} \leq \theta \leq \frac{\pi}{2} + \arcsin\frac{G - x^{(1)}}{r}\right). \qquad (A.7)$$

Algebra shows that $x^{(2)}$ and $x^{(1)}$ are, respectively,

$$G\frac{k^2}{1 + k^2} \pm \frac{k}{1 + k^2}\sqrt{r^2(1 + k^2) - \{b^2 + k^{-2}b^2 + G^2\}}.$$

Moreover,

$$\frac{G - x^{(2)}}{r} = \left(\frac{G}{1 + k^2}\right)\frac{1}{r} - \frac{k}{1 + k^2}\sqrt{(1 + k^2) - r^{-2}\{b^2 + k^{-2}b^2 + G^2\}} \qquad (A.8)$$

which patently is decreasing as r increases. Also,

$$\frac{G - x^{(1)}}{r} = \left(\frac{G}{1 + k^2}\right)\frac{1}{r} + \frac{k}{1 + k^2}\sqrt{(1 + k^2) - r^{-2}(b^2 + k^{-2}b^2 + G^2)}.$$

Therefore,

$$\frac{d}{dr}\left(\frac{G - x^{(1)}}{r}\right) = \left(\frac{1}{1 + k^2}\right)\left(\frac{1}{r^2}\right) \times \left\{\frac{k(b^2 + k^{-2}b^2 + G^2)}{r\sqrt{(1 + k^2) - r^{-2}(b^2 + k^{-2}b^2 + G^2)}} - G\right\}. \qquad (A.9)$$

But $\overline{A_0C_2} = k^{-1}\sqrt{G^2 + B^2}$ and $r \leq \overline{A_0C_2}$ in the region under study presently, so

$$\frac{1}{r} \geq \frac{k}{\sqrt{G^2 + b^2}}. \qquad (A1.0)$$

Consequently,

$$1 + k^2 - r^{-2}(b^2 + k^{-2}b^2 + G^2) \leq \frac{G^2}{b^2 + G^2}. \qquad (A.11)$$

Plug (A.10) and (A.11) into (A.9) to see that the expression (A.9) is bounded below by

$$\left(\frac{1}{1 + k^2}\right)\left(\frac{1}{r^2}\right)(k^2G - G) \geq 0$$

because $k^2 \geq 3$.

In summary, $(d/dr)(G - x^{(1)}/r) > 0$, which together with (A.8) and (A.7) (and the fact that both angles appearing in (A.7) are between 0 and $\pi/2$) implies that if $\overline{A_0P} < r \leq \overline{A_0C_2}$, then p_4 is increasing.

In order to continue the study of p_4, assume now that $r > \overline{A_0C_2}$ For $r > \overline{A_0C_2}$ the angle $\theta_2 = \theta_2(r)$ is defined as in Figure A: $\theta_2(0 \leq \theta_2 \leq \pi/2)$ is the angle formed by the x_1 axis and a line from A_0 to the unique point in $\{x_1 < G\} \cap \{(x_1, 1/k\sqrt{x_1^2 + b^2})\}$ at distance r from A_0. In order to study dp_4/dr for $r > \overline{A_0C_2}$ it is convenient to study first $d\theta_2/d(r\cos\theta_2)$, where as in Figure A, $r\cos\theta_2$ is called $G + y$. From the stated definitions,

$$\theta_2 = \arctan\left[\left(\frac{1}{k}\right)\sqrt{y^2 + b^2}\Big/(y + G)\right].$$

$$\frac{d\theta_2}{d(r\cos\theta_2)} = \frac{d\theta_2}{dy}, \qquad (A.12)$$

which is easily seen to be

$$\frac{k(Gy - b^2)}{[k^2(y + G)^2 + y^2 + b^2]\sqrt{y^2 + b^2}}. \qquad (A.13)$$

Also,

$$\frac{d\theta_2}{dr} = \frac{d\theta_2}{d(r\cos\theta_2)}\cdot\frac{d(r\cos\theta_2)}{dr} = \frac{d\theta_2}{d(r\cos\theta_2)}\left(-r\sin\theta_2\frac{d\theta_2}{dr} + \cos\theta_2\right). \qquad (A.14)$$

In view of (A.12) and (A.14)

$$\frac{d\theta_2}{dr} = \frac{d\theta_2}{dy}(\cos\theta_2)\left\{\frac{1}{1 + \frac{d\theta_2}{dy}(r\sin\theta_2)}\right\}. \qquad (A.15)$$

From Figure A and the definition of θ_2 it follows that $r\sin\theta_2 = k^{-1}\sqrt{y^2 + b^2}$ and that

$$\cos\theta_2 = \frac{k(G + y)}{\sqrt{k^2(y + G)^2 + y^2 + b^2}}.$$

Plug these and (A.13) into (A.15) to see that

$$\frac{d\theta_2}{dr} = \frac{k(Gy - b^2)}{[k^2(y + G)^2 + y^2 + b^2]\sqrt{y^2 + b^2}}\frac{k(y + G)}{\sqrt{k^2(y + G)^2 + y^2 + b^2}} \times 1\Big/\left[1 + \frac{k(Gy - b^2)}{[k^2(y + G)^2 + y^2 + b^2]\sqrt{y^2 + b^2}}\frac{\sqrt{y^2 + b^2}}{k}\right]$$
$$= \frac{k(Gy - b^2)(y + G)}{r\sqrt{y^2 + b^2}\{k^2(y + G)^2 + y^2 + Gy\}}. \qquad (A.16)$$

The denominator of the final expression in (A.16) is non-positive if, and only if, (iff) the term in braces is (remember $b^2 > 0$). That term is non-positive iff $0 \geq y^2(1 + k^2) + y(2k^2G + G) + k^2G^2$, i.e., iff $-G \leq y \leq -k^2G/(1 + k^2)$. In the region under study presently, $r > \overline{A_0C_2} = k^{-1}\sqrt{G^2 + b^2}$. Let $\bar{y}$ be the value of y for $r = \overline{A_0C_2}$. Refer to Figure A to see that $\bar{y} > -G$ and that for all other $y = y(r)$, for $r > \overline{A_0C_2}$, $y > \bar{y}$. Now $\bar{y} > -G$ satisfies

$$k^{-2}(G^2 + b^2) = (G + \bar{y})^2 + k^{-2}(\bar{y}^2 + b^2),$$

so $\bar{y} = -k^2G/(1 + k^2) + G/(1 + k^2) > -k^2G/(1 + k^2)$. Therefore, for $r > \overline{A_0C_2}$ the denominator of (A.16) is positive.

Let $q = n - m - 1$, and let $f(\theta)$ be the value of the density of $\underset{\sim}{\theta}$

at θ. Brown calculated (no change of scale needed here) $f(\theta) = \ell_q(\sin\theta)^q$, where ℓ_q is a constant.

$$p_4(r, b, G) = \int_{\pi/2}^{\pi-\theta_2} f(\theta)d\theta. \quad \frac{dp_4}{dr} = f(\pi - \theta_2)\,\frac{d(\pi - \theta_2)}{dr}.$$

Because $\sin(\pi - \theta_2) = \sin\theta_2$ and $d(\pi - \theta_2)/dr = -d\theta_2/dr$, $dp_4/dr = -\ell_q(d\theta_2/dr)(\sin\theta_2)^q$. From the definition of θ_2 it follows that $\sin\theta_2 = \sqrt{y^2 + b^2}/\sqrt{k^2(y + G)^2 + y^2 + b^2} = \sqrt{y^2 + b^2}/kr$. Therefore,

$$\frac{dp_4}{dr} = \frac{-\ell_q(Gy - b^2)(y + G)(y^2 + b^2)^{(q-1)/2}}{k^{q-1}r^{q+1}\{k^2(y + G)^2 + y^2 + Gy\}} \tag{A.17}$$

if $y < b^2/G$, $dp_4/dr > 0$. (From Figure A $dp_4/dr > 0$ if $y < 0$.)

Now p_2 can be studied in the same manner as was p_4; p_2 is 0 for $r \leq \overline{A_0C_2}$, and so it is non-decreasing in r there. For $r > \overline{A_0C_2}$ define the angle $\theta_1 = \theta_1(r)$ as in Figure A: $\theta_1(0 \leq \theta_1 < \pi/2)$ is the angle formed by the x_1 axis and a line from A_0 to the unique point in $\{x_1 > G\} \cap \{(x_1, (1/k)\sqrt{x_1^2 + b^2})\}$ at distance r from A_0. Put $r\cos\theta_1 = G + x$. Then reasoning as before, one finds that

$$\frac{d\theta_1}{dx} = \frac{k(-Gx - 2G^2 - b^2)}{[k^2(x + G)^2 + (x + 2G)^2 + b^2]\sqrt{(x + 2G)^2 + b^2}}. \tag{A.18}$$

Also,

$$\begin{aligned}\frac{d\theta_1}{dr} &= \frac{d\theta_1}{dx}\,\frac{\cos\theta_1}{1 + r\sin\theta_1(d\theta_1/dx)} \\ &= \frac{k^2(-Gx - 2G^2 - b^2)(x + G)}{[k^2(x + G)^2 + (x + 2G)^2 + b^2]^{\frac{1}{2}}\sqrt{(x + 2G)^2 + b^2}} \\ &\quad \times \frac{1}{\left\{1 - \dfrac{Gx + 2G^2 + b^2}{k^2(x + G)^2 + (x + 2G)^2 + b^2}\right\}} \\ &= \frac{k(-Gx - 2G^2 - b^2)(x + G)}{r\sqrt{(x + 2G)^2 + b^2}\{k^2(x + G)^2 + (x + G)(x + 2G)\}}.\end{aligned} \tag{A.19}$$

The denominator of (A.19) is always positive because

$$x + G = r\cos\theta_1 > 0.$$

Also,

$$p_2(r, b, G) = \int_{\theta_1}^{\pi/2} f(\theta)d\theta \text{ implies}$$

$$\begin{aligned}\frac{dp_2}{dr} &= -\ell_q(\sin\theta_1)^q\frac{d\theta_1}{dr} \\ &= \frac{\ell_q(Gx + 2G^2 + b^2)(x + G)[(x + 2G)^2 + b^2]^{(q-1)/2}}{k^{q-1}r^{q+1}\{k^2(x + G)^2 + (x + G)(x + 2G)\}}.\end{aligned} \tag{A.20}$$

Recall that the goal of the computation is the conclusion: $p_2 + p_4$ is non-decreasing in r. In the interval for which $y \leq b^2/G$ both summands are non-decreasing. On $\{r: y > b^2/G\}$, $dp_4/dr < 0$. The unique $r = r_0$ for which $y = b^2/G$ is larger than $\overline{A_0C_2}$. For all $r > \overline{A_0C_2}$, $dp_2/dr > 0$. Therefore, the proof will be complete if for each fixed $r > r_0$, $dp_2/dr/|dp_4/dr| > 1$. To that end, fix $r > r_0$. From (A.17) and (A.20)

$$\begin{aligned}\frac{dp_2}{dr}\Big/\left|\frac{dp_4}{dr}\right| &= \left[\frac{(Gx + 2G^2 + b^2)(x + G)}{(Gy - b^2)(y + G)}\right] \\ &\quad \cdot\left[\frac{k^2(y + G)^2 + y^2 + Gy}{k^2(x + G)^2 + (x + G)(x + 2G)}\right] \\ &\quad \cdot\left(\frac{(x + 2G)^2 + b^2}{y^2 + b^2}\right)^{(q-1)/2},\end{aligned} \tag{A.21}$$

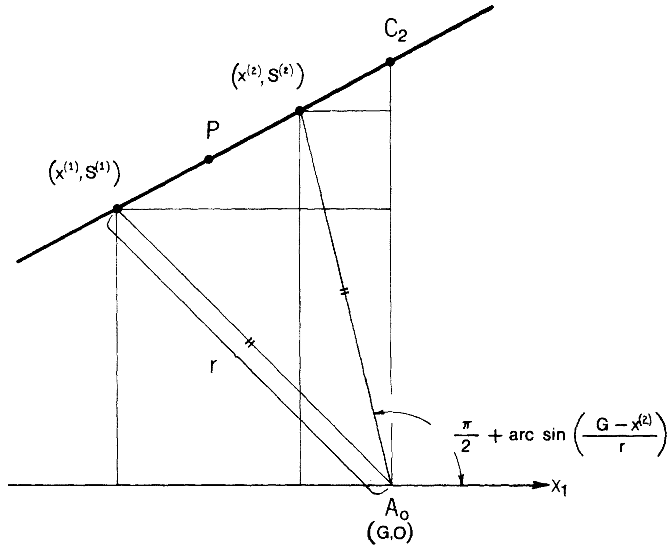
where x and y satisfy

$$k^2(y + G)^2 + y^2 = k^2(x + G)^2 + (x + 2G)^2. \tag{A.22}$$

From (A.22) and the fact that $r > r_0$ it follows that

$$-G < x < y < x + G. \tag{A.23}$$

B. FURTHER GEOMETRY OF THEOREM 2



In view of (A.22) and the fact that $Gy > 0$, the second term of (A.21) is more than one. Also, (A.22) and $q \geq 1$ imply the third term of (A.21) is more than one, so (A.21) is more than one if the first term in the product is more than one. But that term is at least

$$\frac{(x + 2G)(x + G)}{y(y + G)} > \frac{(x + 2G)(x + G)}{(x + G)(x + 2G)} = 1.$$

[Received October 1971. Revised February 1973.]

REFERENCES

[1] Brown, L., "The Conditional Level of the t Test," *Annals of Mathematical Statistics*, 38 (August 1967), 1068–71.

[2] Buehler, R.J., "Some Validity Criteria for Statistical Inferences," *Annals of Mathematical Statistics*, 30 (December 1959), 845–63.

[3] ——— and Fedderson, A.P., "Note on a Conditional Property of Student's t," *Annals of Mathematical Statistics*, 34 (September 1963), 1098–100.

[4] Cohen, A., "Improved Confidence Intervals for the Variance of a Normal Distribution," *Journal of the American Statistical Association*, 67 (June 1972), 382–7.

[5] Cornfield, J. "The Bayesian Outlook and Its Application," *Biometrics*, 24 (December 1969), 617–42.

[6] Edwards, W., Lindman, H. and Savage, L.J., "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, 70 (May 1963), 193–242.

[7] Fisher, R.A., "On a Test of Significance in Pearson's *Biometrika Tables (No. 11)*," *Journal of the Royal Statistical Society*, Ser. B. 18 (1956), 56–60.

[8] ———, *Statistical Methods and Scientific Inference*, Second ed., New York: Hafner Publishing Co., 1959.

[9] Freedman, D.A. and Purves, R.A., "Bayes' Method for Bookies," *Annals of Mathematical Statistics*, 40 (August 1969), 1177–86.

[10] Miller, R.G., Jr., *Simultaneous Statistical Inference*, New York: McGraw Hill Book Co., 1966.

[11] Rao, C.R., *Linear Statistical Inference and Its Applications*, New York: John Wiley and Sons, Inc., 1965.

[12] Scheffé, H., "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika*, 40 (June 1953), 87–104; Corrigenda, *Biometrika*, 56 (March 1969), 229.

[13] ———, *The Analysis of Variance*, New York: John Wiley and Sons, Inc., 1959.

[14] ———, "Multiple Testing Versus Multiple Estimation. Improper Confidence Sets. Estimation of Directions and Ratios," *Annals of Mathematical Statistics*, 41 (February 1970), 1–29.

[15] Sclove, S.L., Morris, C. and Radhakrishnan, R., "Non-Optimality of Preliminary-Test Estimators for the Mean of a Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 43 (October 1972), 1481–90.

[16] Stein, C. "Approximation of Prior Measures by Probability Measures," (Multilithed) Notes of Lecture 2 of the Wald Lectures, 1961.

[17] ———, "Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution with Unknown Mean," *Annals of the Institute of Statistical Mathematics*, 16 (1964), 155–60.

[18] ———, "An Approach to the Recovery of Inter-Block Information in Balanced Incomplete Block Designs," in F.N. David, ed., *Festschrift for J. Neyman*, New York: John Wiley and Sons, Inc., 1966.

[19] Wallace, D.L., "Conditional Confidence Level Properties," *Annals of Mathematical Statistics*, 38 (December 1959), 864–76 (corrected in [16, p. 8]).