The Annals of Statistics
1973 Vol. 1, No. 4, 766–772

AN OPTIMALITY PROPERTY OF SCHEFFÉ BOUNDS[1]

BY ROBERT BOHRER

University of Illinois at Urbana

Conditions are derived which are commonly met in applications and which are sufficient for both one- and two-sided Scheffé bounds to have no greater average width than *any* other-shaped confidence bounds which have the same confidence coefficient.

1. Introduction and statement of results. For a given random vector $\mathbf{B}(n \times 1)$, we consider one-sided bounds ϕ such that

$$P\{\mathbf{x}'\mathbf{B} \leqq \phi(\mathbf{x}), \text{ all } \mathbf{x} \in E\} = \beta, \tag{1.1}$$

as well as corresponding two-sided bounds (ϕ_1, ϕ_2), such that

$$P\{\phi_1(\mathbf{x}) \leqq \mathbf{x}'\mathbf{B} \leqq \phi_2(\mathbf{x}), \text{ all } \mathbf{x} \in E\} = \beta. \tag{1.2}$$

We define the average width of such bounds by

$$W_\phi = \textstyle\int_E \phi(\mathbf{x})\, d\mathbf{x} / \int_E d\mathbf{x} \tag{1.3}$$

in the one-sided case, or

$$W_{\phi_1, \phi_2} = \textstyle\int_E [\phi_2(\mathbf{x}) - \phi_1(\mathbf{x})]\, d\mathbf{x} / \int_E d\mathbf{x} \tag{1.4}$$

in the two-sided case. For fixed values of β and fixed sets E, our goals are to choose ϕ so as to minimize (1.3) subject to (1.1) and to choose (ϕ_1, ϕ_2) to minimize (1.4) subject to (1.2).

Such bounds are useful and widely used in applications where, in terms of statistics $\hat{\boldsymbol\theta}$ and s and of unknown parameters $\boldsymbol\theta$, the random vector is $\mathbf{B} = (\boldsymbol\theta - \hat{\boldsymbol\theta})/s$. Then (1.2) gives simultaneous $100\beta\%$ confidence bounds for the linear function $\{f(\mathbf{x}) = \mathbf{x}'\boldsymbol\theta, \mathbf{x} \in E\}$, viz.,

$$\mathbf{x}'\hat{\boldsymbol\theta} + s\phi_1(\mathbf{x}) \leqq \mathbf{x}'\hat{\boldsymbol\theta} \leqq \mathbf{x}'\hat{\boldsymbol\theta} + s\phi_2(\mathbf{x}).$$

Similarly, (1.1) provides one-sided simultaneous confidence bounds for f on E.

Historically, the first such bounds over nondegenerate E were derived by Working and Hotelling [8] for the case of two-parameter linear regression over the entire real line, i.e., $f(\mathbf{x}) = x_1\theta_1 + x_2\theta_2$ on $E = \{(x_1, x_2) : x_1 = 1, -\infty < x_2 < \infty\}$. These bounds are sharp, two-sided and have property (1.5):

$$\phi_i(\mathbf{x}) \text{ is proportional to the standard deviation of } \mathbf{x}'\mathbf{B}. \tag{1.5}$$

Scheffé ([7], Section 3.5) generalized bounds with shape (1.5) to the n-parameter linear model, $f(\mathbf{x}) = \mathbf{x}'\boldsymbol\theta$, under the usual ([7], Section 2.1) analysis of variance

Received November 1971; revised December 1972.

[1] Research supported by NSF Grant 23835.

AMS 1970 *subject classifications.* Primary 62; Secondary J15, Q05, J10.

Key words and phrases. Confidence bounds, multiple comparison, simultaneous inference.

distributional assumptions. Scheffé's bounds satisfy (1.2) in case E is the whole of n-space; and in this case, (1.1) holds using the one side of Scheffé's bound (see [1]). The probability in (1.2) is, conservatively, at least β if Scheffé's bounds are used over subsets E of n-space.

For example, the probability in (1.2) exceeds β in the case of linear regression over a bounded interval $[a, b]$, i.e., $E = \{(1, x) : a \leqq x \leqq b\}$. Thus the $\phi_i(\mathbf{x})$, and hence the average width W_{ϕ_1,ϕ_2}, are conservatively too large. To overcome this conservatism, Gafarian [4] derived bounds satisfying (1.2) and introduced the average width measure for comparing bounds. Gafarian's bounds have $\phi_i(\mathbf{x})$ constant and thus are different in shape from Scheffé's, as are the trapezoidal bounds of Bowden and Graybill [3], [5]. Average width comparisons of these differently shaped bounds for linear regression over intervals have been made in [2], [3], [4], [5], and [6].

A problem of even more practical interest is to find, for general linear models, bounds which have minimum average width among *all* bounds. This problem does not appear to have been considered heretofore. The present paper treats this problem for the general linear model and establishes conditions under which bounds of the Scheffé shape (1.5) minimize the average width over elipsoidal sets E.

The distributional assumptions of the paper are somewhat general. They include the general linear model ([7], Section 2.1) with variance known or unknown. In such cases, Scheffé bounds are proved optimal over ellipsoids for all sufficiently large "coverage" probabilities β. The tables in Section 4 show that "usual" β-values are in the range of optimality.

In practice, the ellipsoidal shape of E in the optimality result means that Scheffé optimality is proved for bounds over relatively large sets in n-space, i.e., sets of positive volume which can, by taking a sufficiently large elipsoid, include arbitrarily large subsets of n-space. On the other hand, the theorem herein does not prove optimality in the linear regression case, where the set E of interest is a line segment in two-space. For one-way analysis of variance, as another example, the theorem says that Scheffé bounds ((3.4.1) in [7]) are optimal for bounding all (or, at least an arbitrarily large ellipsoid full of) contrasts, but not for bounding all pairwise contrasts, for which case they are known ([7], Section 3.7) to be sub-optimal.

The following theorem is the main result of the paper and is proved in Section 3. Section 2 gives the particularly simple yet crucial building block for the proof. Section 4 concerns a result and tables useful in applying the theorem in practice.

THEOREM. *Suppose, for $M(n \times n)$ nonsingular, that $\Sigma = MM'$ and that $\mathbf{B}(n \times 1)$ is a random vector with density $f_{\mathbf{B}}(\mathbf{b}) = g((\mathbf{b}'\Sigma^{-1}\mathbf{b})^{\frac{1}{2}})$. Suppose $R = ||M^{-1}\mathbf{B}||$ is a continuous random variable and that its density f_R is unimodal. Define r^* as the 100 β percentile of R, i.e. $P\{R \leqq r^*\} = \beta$, and define*

$$I_{r^*} = \inf_{r \in (0, r^*)} \textstyle\int_r^{r^*} f_R(x)\,dx/(r^* - r)\,.$$

Suppose that the average density on (r, r^*) *always exceeds* $f_R(r^*)$, *whence*

$$I_{r^*} = f_R(r^*) \,. \tag{1.6}$$

Define $E = \{\mathbf{x}: \mathbf{x}'\Sigma\mathbf{x} \leqq c^2\}$. *Then* $\phi^*(\mathbf{x}) = r^*(\mathbf{x}'\Sigma\mathbf{x})^{\frac{1}{2}}$ *satisfies* (1.1); *and if* ϕ *is any bound which satisfies* $P\{\mathbf{x}'\mathbf{B} \leqq \phi(\mathbf{x}), \text{all } \mathbf{x} \in E\} \geqq \beta$, *then* $W_\phi \geqq W_{\phi^*}$. *Moreover, with* $\phi_2^* = -\phi_1^* = \phi^*$, (1.2) *is satisfied; and if* (ϕ_1, ϕ_2) *satisfy* $P\{\phi_1(\mathbf{x}) \leqq \mathbf{x}'\mathbf{B} \leqq \phi_2(\mathbf{x}), \text{all } \mathbf{x} \in E\} \geqq \beta$, *then* $W_{\phi_1,\phi_2} \geqq W_{\phi_1^*,\phi_2^*}$.

2. The case of $n = 1$.

LEMMA 1. *Suppose E is a bounded subset of* $(0, \infty)$ *and B is a random variable with* $P\{B \leqq b^*\} = \beta$. *With* $\phi^*(x) = b^*x$,

$$P\{Bx \leqq \phi^*(x), \text{ all } x \in E\} = \beta \,; \tag{2.1}$$

and if ϕ *satisfies*

$$P\{Bx \leqq \phi(x), \text{ all } x \in E\} \geqq \beta \,, \tag{2.2}$$

then $W_\phi \geqq W_{\phi^*}$.

PROOF.

$$\begin{aligned} P\{Bx \leqq \phi(x), \text{ all } x \in E\} &= P\{B \leqq \phi(x)/x, \text{ all } x \in E\} \\ &= P\{B \leqq \inf_E [\phi(x)/x]\} \,, \end{aligned}$$

which is no less than β only in case $\inf [\phi(x)/x] \geqq b^*$, with equality if $\phi = \phi^*$. Hence, if ϕ satisfies (2.2),

$$\begin{aligned} \int_E \phi(x)\, dx = \int_E \frac{\phi(x)}{x} x\, dx &\geqq \inf_E \frac{\phi(x)}{x} \int_E x\, dx \\ &\geqq b^* \int_E x\, dx = \int_E \phi^*(x)\, dx \,. \end{aligned}$$

REMARK. Geometrically the result is not surprising, as one shows by graphing any coefficient β bound ϕ and the shorter bound $\phi^*(y) = \inf_E [\phi(x)/x]y$; ϕ lies above a given line only in case ϕ^* does.

LEMMA 2. *If E and the distribution of B are symmetric about zero, if* $P\{|B| < b^*\} = \beta$, *and if* $\phi^*(x) = b^*x$ *then*

$$P\{Bx \leqq \phi^*(x), \text{ all } x \in E\} = \beta$$
$$P\{-\phi^*(x) \leqq Bx \leqq \phi^*(x), \text{ all } x \in E\} = \beta \,;$$

and if

$$P\{Bx \leqq \phi(x), \text{ all } x \in E\} \geqq \beta \,,$$

then $W_\phi \geqq W_{\phi^*}$, *or if*

$$P\{\phi_1(x) \leqq Bx \leqq \phi_2(x), \text{ all } x \in E\} \geqq \beta \,,$$

then,

$$W_{\phi_1,\phi_2} \geqq W_{-\phi^*,\phi^*} \,.$$

Proof is as in Lemma 1. Note that $\pm Bx \leqq \phi^*(\pm x)$ only in case $|B|\,|x| \leqq \phi^*(|x|)$.

3. Proof of the Theorem. Parts 1–3 derive the one-sided case, and part 4 extends this to the two-sided case.

1. (a simplifying transformation). Define $\mathbf{T} = M^{-1}\mathbf{B}$, which has density $f_{\mathbf{T}}(\mathbf{t}) = f_{\mathbf{B}}(M\mathbf{t})\,|\mathrm{Det}\,(M)| = g(||\mathbf{t}||)\,|\mathrm{Det}\,(M)|$. This is circularly symmetric, i.e., $||\mathbf{t}|| = ||\mathbf{t}^*||$ implies $f_{\mathbf{T}}(\mathbf{t}) = f_{\mathbf{T}}(\mathbf{t}^*)$.

We restate the problem in terms of $\mathbf{T}$. Define $S = \{\mathbf{y} : ||\mathbf{y}|| \leqq c\}$ and $P_\psi = P\{\mathbf{T}'\mathbf{y} \leqq \psi(\mathbf{y}), \text{all } \mathbf{y} \in S\}$. If, among bounds ψ with $P_\psi \geqq \beta$, ψ^* minimizes $\int_S \psi(\mathbf{y})\,d\mathbf{y}$, then $\phi^*(\mathbf{x}) = \psi^*(M'\mathbf{x})$ minimizes W_ϕ among coefficient β bounds in the original problem, since

$$\int_S \psi(\mathbf{y})\,d\mathbf{y} = \int_E \psi(M'\mathbf{x})\,d\mathbf{x}\,|\mathrm{Det}\,(M)| = \int_E \phi(\mathbf{x})\,d\mathbf{x}\,|\mathrm{Det}\,(M)|\,.$$

It thus suffices to prove that

(3.1) $$\psi^*(\mathbf{y}) = \phi^*(M'^{-1}\mathbf{y}) = r^*(\mathbf{y}'M^{-1}\Sigma\mathbf{M}'^{-1}\mathbf{y})^{\frac{1}{2}} = r^*||\mathbf{y}||$$

minimizes $\int_S \psi(\mathbf{y})\,d\mathbf{y}$.

2. (linearity along rays). Define $U = \{\mathbf{u} : ||\mathbf{u}|| = 1\}$, and write

(3.2) $$\int_S \psi(\mathbf{y})\,d\mathbf{y} = \int_U \int_0^c \psi(k\mathbf{u})k^{n-1}\,dk\,d\mathbf{u} \geqq \int_U \int_0^c \inf_{k' \in (0,c]} \frac{\psi(k'\mathbf{u})}{k'}\,k^n\,dk\,d\mathbf{u}\,,$$

with equality if ψ has the form $\psi(\mathbf{y}) = ||\mathbf{y}||\lambda(\mathbf{y}/||\mathbf{y}||)$. Note that

(3.3) $$P_\psi = P\{\mathbf{T}'k\mathbf{u} \leqq \psi(k\mathbf{u}),\ \text{all}\ \mathbf{u} \in U\ \text{and}\ 0 < k < c\} = P\left\{\mathbf{T}'\mathbf{u} \leqq \inf_{k \in (0,C]} \frac{\psi(k\mathbf{u})}{k},\ \text{all}\ \mathbf{u} \in U\right\}.$$

Given any ψ with $P_\psi = \beta$, the function $\psi'(\mathbf{y}) = ||\mathbf{y}||\inf_{k \in (0,c]}[\psi(k\mathbf{y}/||\mathbf{y}||)/k]$ also attains $P_{\psi'} = \beta$ with no increase in the width (3.2). Hence the ψ search can be restricted to those of the form

(3.4) $$\psi(\mathbf{y}) = ||\mathbf{y}||\lambda(\mathbf{y}/||\mathbf{y}||)\,,$$

where λ is a function defined for $\mathbf{u} \in U$. Note that the width of such a bound is

(3.5) $$\int_S \psi(\mathbf{y})\,d\mathbf{y} = \int_U \lambda(\mathbf{u}) \int_0^c k^n\,dk\,d\mathbf{u} = c^{n+1}(n+1)^{-1} \int_U \lambda(\mathbf{u})\,d\mathbf{u}\,.$$

3. (optimality of $\lambda(\mathbf{u}) \equiv r^*$). For ψ as in (3.4), $P_\psi = P\{\mathbf{T}'\mathbf{u} \leqq \lambda(\mathbf{u}), \text{all } \mathbf{u} \in U\}$. For $\mathbf{u} = \mathbf{T}/||\mathbf{T}|| \in U$, $\mathbf{T}'\mathbf{u} = ||\mathbf{T}||$ and $P_\psi \leqq P\{||\mathbf{T}|| \leqq \lambda(\mathbf{T}/||\mathbf{T}||)\}$. A geometric argument, or its corresponding analytics, based on circular symmetry of $f_{\mathbf{T}}$, shows that $\mathbf{T}/||\mathbf{T}||$ is uniformly distributed on U independent of $R = ||\mathbf{T}||$, and hence that

(3.6) $$P_\psi \leqq \int_U \int_0^{\lambda(\mathbf{u})} f_R(r)\,dr\,d\mathbf{u}/A\,,$$

where A is the area of U. From the definitions (3.1) of ψ^* and (3.3) of P_{ψ^*}, note that ψ^* satisfies (3.6) with equality.

Finally consider any ψ as in (3.4) with $P_\psi \geqq \beta$. Define $U^+ = \{\mathbf{u} : r^* \geqq \lambda(\mathbf{u})\}$

and $U^- = \{\mathbf{u}: r^* \leqq \lambda(\mathbf{u})\}$. Using (1.6), (3.5), and (3.6) note that

$$(3.7)\qquad \begin{aligned} 0 &\geqq P_{\psi^*} - P_\psi \geqq A^{-1}[\textstyle\int_{U^+}\int_{\lambda(\mathbf{u})}^{r^*} - \int_{U^-}\int_{r^*}^{\lambda(\mathbf{u})}] f_R(r)\,dr\,d\mathbf{u} \\ &\geqq f_R(r^*)A^{-1}\textstyle\int_U [r^* - \lambda(\mathbf{u})]\,d\mathbf{u} \\ &= f_R(r^*)A^{-1}(n+1)c^{-(n+1)}[\textstyle\int_S \psi^*(\mathbf{u})\,d\mathbf{u} - \int_S \psi(\mathbf{u})\,d\mathbf{u}]\,. \end{aligned}$$

4. (the two-sided case). As in part 2, it can be shown that it suffices to consider $\phi_i(M'^{-1}\mathbf{y}) = \psi_i(\mathbf{y}) = ||\mathbf{y}||\lambda_i(\mathbf{y}/||\mathbf{y}||)$. For such bounds, in a hopefully obvious notation,

$$\begin{aligned} P_\psi &= P\{\lambda_1(\mathbf{u}) \leqq \mathbf{T}'\mathbf{u} \leqq \lambda_2(\mathbf{u}),\ \text{all}\ \ \mathbf{u}\in U\} \\ &\leqq P\{\lambda_1(\mathbf{u}) \leqq \mathbf{T}'\mathbf{u} \leqq \lambda_2(\mathbf{u}),\ \mathbf{u} = \pm\mathbf{T}/||\mathbf{T}||\} \\ &= P\{||\mathbf{T}|| \leqq \min(-\lambda_1(\mathbf{T}/||\mathbf{T}||), \lambda_2(\mathbf{T}/||\mathbf{T}||))\}\,, \end{aligned}$$

with equality for $\phi_i^*(\mathbf{y}) = \phi_i^*(M'^{-1}\mathbf{y})$. The analogs of (3.5)—(3.7) follow, to establish the two-sided result.

4. On condition (1.6). Lemma 3 and its corollary give properties of I_{r^*}, which are useful in checking condition (1.6) over the range of possible r^* values in which Scheffé bounds are optimal for normal distribution linear models.

LEMMA 3. *Suppose f_R is continuous and strictly unimodal, i.e., strictly increasing for $r < r_M$ and strictly decreasing for $r > r_M$.*

(A) *If $r_M = 0$, then $I_{r^*} = f_R(r^*)$ for all r^*.*

(B) *If $0 < r < r^* \leqq r_M$, then $\int_r^{r^*} f_R(x)\,dx < f_R(r^*)(r^* - r)$, so $I_{r^*} < f_R(r^*)$.*

(C) *If $r^* > r_M$ and $I_{r^*} < f_R(r^*)$ then*

$$I_{r^*} = \textstyle\int_0^{r^*} f_R(x)\,dx/r^* = P\{R \leqq r^*\}/r^*\,.$$

(D) *If $r^* < r^{**}$ and $I_{r^*} = f_R(r^*)$, then $I_{r^{**}} = f_R(r^{**})$.*

PROOF. (A) and (B) follow immediately. To prove (C), denote $I(a, b) = \int_a^b f_R(x)\,dx$ and define $r' < r^*$ by $f_R(r') = f_R(r^*)$. Note that the inf in defining I_{r^*} can be attained at r only if $r < r'$, in which case $f_R(r) = \inf_{x\in[r,r^*]} f_R(x)$. Also, f_R is increasing on $[0, r']$. From this and from

$$\Delta^{-1}\left[\frac{I(r+\Delta, r^*)}{r^* - r - \Delta} - \frac{I(r, r^*)}{r^* - r}\right] = \frac{I(r, r^*) - ((r^* - r)/\Delta)I(r, r+\Delta)}{(r^* - r)(r^* - r - \Delta)}$$

it follows that

$$\frac{\partial}{\partial r}\frac{I(r, r^*)}{r^* - r} = \frac{[I(r, r^*) - (r^* - r)f_R(r)]}{(r^* - r)^2} > 0\,.$$

Thus, the function to be minimized increases with r and hence is minimum at $r = 0$. (D) follows, similarly, by proving that, if $I_{r^*} = f_R(r^*)$, then $\partial[I(r, r^*)/(r^* - r)]/\partial r \geqq 0$.

COROLLARY. *Condition* (1.6) *obtains for all $r^* \geqq r_0$ if and only if*

$$(4.1)\qquad I(0, r_0) \geqq r_0 f_R(r_0)\,.$$

Proof of "only if" is immediate. To show that (4.1) suffices, note that $r_0 \geqq r_M$

by Part (B) of Lemma 3. Hence (C) says that, if (1.6) fails, then $I_{r_0} = I(0, r_0)/r_0 \geqq f_R(r_0)$, to contradict (4.1), i.e., (4.1) implies that (1.6) holds at r_0, and hence, by (D), for each $r^* \geqq r_0$.

The unimodality assumption of Lemma 3 is satisfied, as noted in Section 1, in the usual normal distribution analyses of variance for general linear model. Then $R = (\chi^2(n))^{\frac{1}{2}}$ for known variance and $(nF(n, v))^{\frac{1}{2}}$ if variance is estimated with v degrees of freedom. The corollary, together with F and χ^2 tables, can be used to find minimal coverage probabilities β^* for which Scheffé bounds are optimal; by the corollary they are optimal also for all $\beta \geqq \beta^*$. Some values are tabled.

Formulas (4.2) and (4.3) provide the general rules for checking whether Scheffé bounds minimize average width for coefficient β, n variables, and v degrees of freedom. If $v < \infty$ and **B** is n-variate Student t with v degrees of freedom, then R^2/n is $F(n, v)$, so $r^* = (nF_\beta(n, v))^{\frac{1}{2}}$, where $F_\beta(n, v)$ is the 100 β percentile of the $F(n, v)$ distribution. Hence, Scheffé bounds are optimal if

$$\frac{2\Gamma((n+v)/2)}{\Gamma(n/2)\Gamma(v/2)} \frac{v^{u/2} r^{*n}}{(v + r^{*2})(n+v)/2} \leqq \beta . \tag{4.2}$$

If **B** is n-variate standard normal, as in the limit of the previous case as $v \to \infty$ or in the case of known variance analysis of variance, then R^2 is $\chi^2(n)$, so $r^* = (\chi_\beta^{\,2}(n))^{\frac{1}{2}}$, and Scheffé bounds are optimal if

$$r^{*n} e^{-r^{*2}/2} 2^{1-n/2}/\Gamma(n/2) \leqq \beta . \tag{4.3}$$

TABLE 1

β^* such that known variance, normal distribution Scheffé bounds for $\sum_{i=1}^n \beta_i x_i$ are optimal if coverage probability is $\beta \geqq \beta^*$. Thus β^* is the smallest β for which (4.3) *obtains*

$n = 1$	2	4	6	8	10	20
$\beta^* = 0$	.715	.830	.869	.890	.905	.93

TABLE 2

Bounds on β^*. Scheffé bounds for $\sum_{i=1}^n \beta_i x_i$, variance unknown and estimated with ν degrees of freedom, are optimal if coverage probability exceeds β^*. Thus β^* is the smallest β for which (4.2) *obtains*

	$.5 \leqq \beta^* \leqq .75$	$.75 \leqq \beta^* \leqq .90$	$.90 \leqq \beta^* \leqq .95$
$n = 2$	$2 \leqq \nu$		
$n = 10$	$2 \leqq \nu \leqq 4$	$5 \leqq \nu$	
$n = 20$	$\nu = 2$	$2 \leqq \nu \leqq 30$	$60 \leqq \nu$

REFERENCES

[1] Bohrer, R. (1969). On one-sided confidence bounds for response surfaces. *Bull. Inst. Internat. Statist.* **43** 255–257.

[2] Bohrer, R. and Francis, G. K. (1972). Sharp one-sided confidence bounds for linear regression over intervals. *Biometrika* **59** 99–107.

[3] Bowden, D. C. and Graybill, F. A. (1966). Confidence bands of uniform and proportional width for linear models. *J. Amer. Statist. Assoc.* **61** 182–190.

[4] GAFARIAN, A. V. (1964). Confidence bands for straight line regression. *J. Amer. Statist. Assoc.* **59** 182–213.
[5] GRAYBILL, F. A. and BOWDEN, D. C. (1967). Linear segment confidence bands for simple linear models. *J. Amer. Statist. Assoc.* **62** 403–408.
[6] HALPERIN, M. and GURIAN, J. (1968). Confidence bands in linear regression with constraints. *J. Amer. Statist. Assoc.* **63** 1020–1027.
[7] SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.
[8] WORKING, H. and HOTELLING, H. (1929). Application of the theory of error to the interpretation of trends. *J. Amer. Statist. Assoc. Suppl.* (*Proc.*) **24** 73–85.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801