Multiple Comparisons in the General Linear Model

Jason C. Hsu and Barry Nelson

Whereas multiple comparisons computations in a one-way model are well understood, multiple comparisons computations in a general linear model (GLM) are not. For models with the so-called "one-way structure," no new technique is needed beyond proper substitution of terms. Examples of designs that guarantee a one-way structure include *variance balanced* designs and *orthogonal* designs. For models without a one-way structure, more sophisticated computational techniques are needed. Approximations based on the probabilistic inequalities of Bonferroni, Šidák, and Slepian are too conservative. Even the second-order Hunter–Worsley inequality is rather conservative. The so-called *factor analytic* approximation is quite accurate for multiple comparison with a control (MCC) and multiple comparison with the best (MCB), but conditions for it to be conservative are not known. This article describes a highly accurate, deterministic, conservative approximation that is applicable to a popular class of general linear models, and a fast, stochastic, conservative approximation that is generally applicable.

Key Words: Linear programming; Quantile estimation; Variance reduction.

1. MULTIPLE COMPARISONS IN THE GENERAL LINEAR MODEL

Suppose two or more "treatments" are to be compared. A variety of multiple comparisons may be of interest, such as all-pairwise comparisons (MCA), multiple comparison with the best (MCB), multiple comparisons with a control (MCC), and multiple comparisons with the mean (MCM). As described in Hsu (1996), exact computations for MCA and MCM inferences are feasible only for variance-balanced models. However, computationally exact MCC and MCB inferences have been implemented in JMP and MINITAB for the unbalanced one-way model, and extended to variance-balanced models and models in which effects are orthogonal under the LSMEANS option of PROC GLM and PROC MIXED in SAS 6.11. This article discusses the computation of MCC and MCB inferences in the general linear model (GLM) when exact computations are impossible. We give some examples of situations in which such inferences may be of interest.

Jason C. Hsu is Professor, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247 (E-mail: jch@stat.ohio-state.edu). Barry Nelson is Associate Professor, Department of Industrial Engineering and Management Sciences, Northwestern University, 2225 North Campus Drive, Evanston, IL 60208 (E-mail: nelsonb@iems.nwu.edu).

Journal of Computational and Graphical Statistics, Volume 7, Number 1, Pages 23–41

Consider a general linear model (GLM) with fixed effects:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{Y}_{N\times 1}$ is the vector of observations, $\mathbf{X}_{N\times p}$ is a known design matrix, $\boldsymbol{\beta}_{p\times 1} = (\beta_1, \ldots, \beta_p)'$ is the vector of parameters, and $\boldsymbol{\epsilon}_{N\times 1}$ is a vector of iid normally distributed errors with mean 0 and unknown variance σ^2.

In the special case of a one-way analysis of covariance (ANCOVA) model,

$$Y_{ih} = \theta_i + \beta X_{ih} + \epsilon_{ih}, \quad i = 1, \ldots, k, \quad h = 1, \ldots, n_i, \tag{1.2}$$

the response Y not only depends on the treatment (indexed by i) but is also linearly dependent on a covariate X. If the assumed lack of interaction between the treatment effect and the covariate in (1.2) is reasonable, then one can meaningfully compare the treatments after adjusting for the covariates (which puts the treatments on an equal footing in the comparison).

Another special case is the two-way no-interaction model

$$Y_{ihr} = \mu + \tau_i + \beta_h + \epsilon_{ihr}, \quad i = 1, \ldots, k, \quad h = 1, \ldots, b, \quad r = 1, \ldots, n_{ih}, \tag{1.3}$$

in which the τ_i's denote treatment effects and the β_h's denote block effects. If the assumed lack of interaction between treatment effects and block effects in (1.3) is reasonable, then one can meaningfully compare the treatments after adjusting for block effects.

We will let $\mu_i - \mu_j$ denote the generic multiple comparison parameters of interest in a GLM. Thus, $\mu_i - \mu_j$ may denote $\theta_i - \theta_j$ in an ANCOVA model, or $\tau_i - \tau_j$ in a two-way model.

We assume that all $\mu_i - \mu_j$, $i \neq j$, are estimable. For a fixed i, if $\mathbf{C}_{-i}$ is the matrix such that $\boldsymbol{\mu}_{-i} = (\mu_j - \mu_i, \forall j \neq i)' = \mathbf{C}_{-i}\boldsymbol{\beta}$, then even if the individual μ_i are not estimable, there will be no confusion if we use the notation $\hat{\boldsymbol{\mu}}_{-i} = (\hat{\mu}_j - \hat{\mu}_i, \forall j \neq i)' = \mathbf{C}_{-i}\hat{\boldsymbol{\beta}}$. Let $\hat{\sigma}^2 = \text{MSE} = (\boldsymbol{Y} - \mathbf{X}\hat{\beta})'(\boldsymbol{Y} - \mathbf{X}\hat{\beta})/(N - \text{rank}(\mathbf{X}))$ denote the usual estimator of σ^2. Because MCC and MCB inferences are based on $\hat{\boldsymbol{\mu}}_{-i}$ and $\hat{\sigma}^2$, we consider their distributions.

Under the iid normal errors assumption of model (1.1), $\hat{\boldsymbol{\mu}}_{-i}$ is multivariate Normal$(\mathbf{C}_{-i}\boldsymbol{\beta}, \sigma^2\mathbf{V}_{-i})$, where $\mathbf{V}_{-i} = \mathbf{C}_{-i}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'_{-i}$. Also, $\nu\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with $\nu = N - \text{ rank } (\mathbf{X})$ degrees of freedom, and it is independent of $\hat{\boldsymbol{\mu}}_{-i}$. For later convenience, let $\sigma^2 v_j^i$ denote the variance of $\hat{\mu}_j - \hat{\mu}_i$—that is, v_j^i is the diagonal element of $\mathbf{V}_{-i} = \mathbf{C}_{-i}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'_{-i}$ corresponding to $\hat{\mu}_j - \hat{\mu}_i$, and let $\mathbf{R}_{-i}$ denote the correlation matrix of $\hat{\boldsymbol{\mu}}_{-i}$.

2. MULTIPLE COMPARISONS WITH A CONTROL IN GLM

Suppose treatments versus control comparisons are our primary concern, so the parameters of interest are $\mu_i - \mu_k$, $i = 1, \ldots, k-1$. Recall that $\sigma^2 v_i^k$ denotes the variance of $\hat{\mu}_i - \hat{\mu}_k$.

In theory, one can generalize Dunnet's (1955) MCC confidence intervals for one-way designs to the general linear model in a straightforward fashion. Suppose the constant

$|d|$ satisfies

$$P\left\{\max_{1\le i\le k-1}\hat{\sigma}^{-1}|\hat{\mu}_i-\hat{\mu}_k-(\mu_i-\mu_k)|\Big/\sqrt{v_i^k}<|d|\right\}=1-\alpha, \tag{2.1}$$

then

$$P\{\hat{\mu}_i-\hat{\mu}_k-|d|\hat{\sigma}\sqrt{v_i^k}<\mu_i-\mu_k<\hat{\mu}_i-\hat{\mu}_k+|d|\hat{\sigma}\sqrt{v_i^k}\quad\text{for}\quad i=1,\ldots,k-1\}=1-\alpha.$$

Thus, for two-sided MCC, one infers,

$$\hat{\mu}_i-\hat{\mu}_k-|d|\hat{\sigma}\sqrt{v_i^k}<\mu_i-\mu_k<\hat{\mu}_i-\hat{\mu}_k+|d|\hat{\sigma}\sqrt{v_i^k},\quad i=1,\ldots,k-1.$$

Suppose the constant d satisfies

$$P\left\{\max_{1\le i\le k-1}\hat{\sigma}^{-1}(\hat{\mu}_i-\hat{\mu}_k-(\mu_i-\mu_k))\Big/\sqrt{v_i^k}<d\right\}=1-\alpha, \tag{2.2}$$

then

$$P\{\hat{\mu}_i-\hat{\mu}_k-d\hat{\sigma}\sqrt{v_i^k}<\mu_i-\mu_k\quad\text{for}\quad i=1,\ldots,k-1\}=1-\alpha$$

or

$$P\{\mu_i-\mu_k<\hat{\mu}_i-\hat{\mu}_k+d\hat{\sigma}\sqrt{v_i^k}\quad\text{for}\quad i=1,\ldots,k-1\}=1-\alpha.$$

So for one-sided MCC, one infers,

$$\mu_i-\mu_k>\hat{\mu}_i-\hat{\mu}_k-d\hat{\sigma}\sqrt{v_i^k}\quad\text{for}\quad i=1,\ldots,k-1$$

or

$$\mu_i-\mu_k<\hat{\mu}_i-\hat{\mu}_k+d\hat{\sigma}\sqrt{v_i^k}\quad\text{for}\quad i=1,\ldots,k-1.$$

Computing the probabilities (2.1) and (2.2) as k dimensional integrals in order to solve for the critical values $|d|$ and d is only feasible for very small k, as the computational effort increases exponentially as k increases. One can simulate the probabilities (2.1) and (2.2) as described in Edwards and Berry (1987) or Somerville (1995). However, deterministic answers are often desirable. (For example, it is difficult to imagine FDA approving a drug based on analysis which changes each time it is repeated.) We describe in the following a condition under which $|d|$ and d can be computed by solving two-dimensional integral equations, so that the computing time is not impacted significantly by the dimension k.

Let ρ_{ij}^k denote $\text{corr}(\hat{\mu}_i-\hat{\mu}_k,\hat{\mu}_j-\hat{\mu}_k)$. Suppose there exist constants $\lambda_1,\ldots,\lambda_{k-1}$ such that $\rho_{ij}^k=\lambda_i\lambda_j$ for all $i<k$ or, equivalently,

$$\mathbf{R}_{-k}=\text{diag}(1-\lambda_1^2,\ldots,1-\lambda_{k-1}^2)+(\lambda_1,\ldots,\lambda_{k-1})'(\lambda_1,\ldots,\lambda_{k-1}). \tag{2.3}$$

Then $\sigma^{-1}((\hat{\mu}_i-\hat{\mu}_k-(\mu_i-\mu_k))/\sqrt{v_i^k},\ i<k)$ has the same distribution as

$$\left(\sqrt{1-\lambda_1^2}Z_1+\lambda_1Z_0,\ldots,\sqrt{1-\lambda_{k-1}^2}Z_{k-1}+\lambda_{k-1}Z_0\right),$$

where $Z_0, Z_1, \ldots, Z_{k-1}$ are iid standard normal random variables. By conditioning on $\hat{\sigma}$ and Z_0, $|d|$ can be written as the solution to

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \left[\Phi\left(\frac{\lambda_i z + |d|u}{\sqrt{1-\lambda_i^2}} \right) - \Phi\left(\frac{\lambda_i z - |d|u}{\sqrt{1-\lambda_i^2}} \right) \right] d\Phi(z)\gamma(u)du = 1 - \alpha$$

and d can be written as the solution to

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \Phi\left(\frac{\lambda_i z + du}{\sqrt{1-\lambda_i^2}} \right) d\Phi(z)\gamma(u)du = 1 - \alpha,$$

where Φ is the standard normal distribution, and γ is the density of $\hat{\sigma}/\sigma$. With efficient numerical integration and root finding algorithms, $|d|$ and d can be obtained quickly enough for interactive data analysis.

Hsu (1996) calls a model for which there exists a set of positive constants $a_1, \ldots, a_k$, such that

$$v_j^i = a_i + a_j \tag{2.4}$$

for all $k(k-1)/2$ pairs of $(i, j), i \neq j$, a model with *one-way structure*, and proves that such a model satisfies (2.3) with $\lambda_i = (1 + a_i/a_k)^{-1/2}$, $i < k$.

At least two possible characteristics of an experimental design are individually sufficient to ensure a one-way structure for the model: *Balance* in the assignment of treatments to blocks ensures an equal covariance one-way structure for the model; and *orthogonality* of effects ensures a one-way structure of the model, although not necessarily an equal covariance one-way structure.

A design is *variance balanced* if all elementary contrasts $\mu_i - \mu_j$ are estimated with the same precision: $\text{var}(\hat{\mu}_i - \hat{\mu}_j) = 2\sigma^2/e$ for all $i \neq j$, where e is a design dependent but known constant. Examples of variance-balanced designs include balanced incomplete block designs, Latin squares, and Youden square designs. Clearly variance-balanced designs have one-way structure.

Latin square designs are not only variance-balanced but are also *orthogonal* in the sense that the blocking effects are orthogonal to the treatment effect. Therefore, estimates of treatment effect parameters are based on treatment means and no adjustment for blocking effects is necessary except in the estimation of experiment error (John 1987, p. 101 and p. 103). An example of a linear model that is not variance-balanced, but is orthogonal, is the two-way proportionate cell frequencies model.

However, models with covariates or missing observations typically lead to $\mathbf{R}_{-k}$ not satisfying (2.3). In the following, we describe methods of approximating $|d|$ and d with computing time not impacted significantly by the dimension k.

Let $t_{\alpha,\nu}$ denote the upper αth quantile of the univariate t distribution with ν degrees of freedom. By the Bonferroni inequality, a conservative approximation to $|d|$ is $t_{\alpha/(2(k-1)),\nu}$, and a conservative approximation to d is $t_{\alpha/(k-1),\nu}$. By Šidák's inequality, a less conservative approximation to $|d|$ is $|d|_{\text{ind.}} = t_{(1-(1-\alpha)^{k-1})/2,\nu}$, but the improvement is very slight. If the correlations among the treatment versus control estimators are all positive, then by Slepian's inequality a less conservative approximation to d is $d_{\text{ind.}} = t_{1-(1-\alpha)^{k-1},\nu}$, but again the improvement is very slight.

If, instead of approximating the multivariate-t probability by functions of univariate-t probabilities, one approximates them with functions of bivariate-t probabilities, then a better approximation results. In two-sided MCC, let

$$E_i = \left\{ |\hat{\mu}_i - \hat{\mu}_k - (\mu_i - \mu_k)| / \hat{\sigma}\sqrt{v_i^k} < |d| \right\}.$$

In one-sided MCC, let

$$E_i = \left\{ (\hat{\mu}_i - \hat{\mu}_k - (\mu_i - \mu_k)) / \hat{\sigma}\sqrt{v_i^k} > -d \right\}$$

or

$$E_i = \left\{ (\hat{\mu}_i - \hat{\mu}_k - (\mu_i - \mu_k)) / \hat{\sigma}\sqrt{v_i^k} < d \right\}.$$

The second-order inequality of Hunter–Worsley (Hunter 1976; Worsley 1982), in the context of MCC, states that if the e_{ij} represent the edges in a spanning tree $\mathcal{T}$ of the nodes $\{1, 2, \ldots, k-1\}$, then

$$P(\cup E_i^c) \leq \sum_{i=1}^{k-1} P(E_i^c) - \sum_{e_{ij} \in \mathcal{T}} P(E_i^c \cap E_j^c).$$

Clearly, the best Hunter–Worsley approximation is obtained by finding the *optimal* spanning tree, the tree $\mathcal{T}$ which maximizes the second term on the right-hand side. This best Hunter–Worsley approximation, though typically significantly better than the Bonferroni, Šidák, or Slepian approximations, is still quite conservative.

Noting that correlation matrices with the product structure (2.3) correspond to correlation matrices with a single *factor* in multivariate factor analysis, Hsu (1992) proposed using existing factor analysis algorithms to approximate $\mathbf{R}_{-k}$ by the "closest" $\mathbf{R}_{-k}^{FA}$ satisfying (2.3), and then computing $|d|$ or d from $\mathbf{R}_{-k}^{FA}$. In contrast to methods based on probabilistic inequalities, the *factor analytic* approximation will automatically produce exact MCC inference whenever possible. Hsu (1992) presented evidence that the factor analytic approximation typically far outperforms the Hunter–Worsley approximation when exact MCC inference is impossible. Based on this evidence, as well as their own studies, SAS Institute has implemented the factor analytic approximation as the default computational method for MCC inference under the LSMEANS option in PROC GLM and PROC MIXED of SAS Version 6.11.

However, the conditions under which the factor analytic approximation will be conservative (when exact MCC inference is impossible) is not known, and this may bother some users. For such users, we describe in Section 3 a highly accurate, deterministic, conservative approximation that is applicable to a popular class of general linear models, and in Section 4 a fast, stochastic, conservative approximation that is generally applicable.

3. THE LINEAR PROGRAMMING METHOD

According to Slepian's inequality, if one approximates $\mathbf{R}_{-k}$ by a correlation matrix of the form (2.3), where $\lambda_i \lambda_j \leq \rho_{ij}^k, \forall i \neq j$, then the computed critical value is conservative.

When all ρ_{ij}^k are positive, one can approximate $\mathbf{R}_{-k}$ by a correlation matrix $\mathbf{R}_{-k}^{LP}$ of the form (2.3) with not only $\lambda_i\lambda_j \leq \rho_{ij}^k, \forall i \neq j$, but also $0 < \lambda_i \leq 1$, $i = 1, \ldots, k-1$. This improves upon the approximation $d_{\text{ind.}}$ which, by Kimball's inequality, is slightly more conservative than the approximation obtained by replacing $\mathbf{R}_{-k}$ by the identity matrix.

An example of a GLM for which all ρ_{ij}^k in $\mathbf{R}_{-k}$ are positive is the two-way no-interaction model (1.3). Let $\mathbf{N}$ denote the *incidence* matrix whose (i,j)th entry is the number of times the ith treatment occurs in the jth block. Then the *inverse* of the variance-covariance matrix of $\hat{\tau}_1 - \hat{\tau}_k, \ldots, \hat{\tau}_{k-1} - \hat{\tau}_k$ is

$$\mathbf{K} = [\text{diag}(\boldsymbol{n}_{*+}) - \mathbf{N}\left[\text{diag}(\boldsymbol{n}_{+*})\right]^{-1}\mathbf{N}']_{k-}$$

(see Searle 1971, p. 267). Here $\text{diag}(\boldsymbol{n}_{*+})$ is the diagonal matrix with the row sums of $\mathbf{N}$ as its diagonal elements, $\text{diag}(\boldsymbol{n}_{+*})$ is the diagonal matrix with the column sums of $\mathbf{N}$ as its diagonal elements, and for any matrix $\mathbf{L}$, $\mathbf{L}_{k-}$ denotes that matrix with its kth row and column deleted. Because the off-diagonal elements of $\mathbf{K}$ are all negative, it is a so-called *M-matrix*. Consequently, all covariances of $\hat{\tau}_1 - \hat{\tau}_k, \ldots, \hat{\tau}_{k-1} - \hat{\tau}_k$ are positive (Graybill 1983, chap. 11). In fact, this shows the joint distribution of $\hat{\tau}_1 - \hat{\tau}_k, \ldots, \hat{\tau}_{k-1} - \hat{\tau}_k$ has the stronger property of MTP_2; see Karlin and Rinnot (1980).

The motivation for using linear programming algorithms to approximate $\mathbf{R}_{-k}$ is that the closer $\lambda_i\lambda_j$ are to ρ_{ij}^k the less conservative the approximate critical value. Thus, we want to ensure that

$$\lambda_i\lambda_j \leq \rho_{ij}^k, \forall j > i, \tag{3.1}$$

and

$$0 < \lambda_i \leq 1, \forall i \tag{3.2}$$

while keeping $\lambda_i\lambda_j$ close to ρ_{ij}^k. One measure of closeness is the ratio $\rho_{ij}^k/\lambda_i\lambda_j$, and one overall objective is to

$$\text{minimize} \prod_{j>i} \frac{\rho_{ij}^k}{\lambda_i\lambda_j}. \tag{3.3}$$

The objective function (3.3) combined with the constraints (3.1)–(3.2) form a nonlinear program. By taking logarithms of (3.1)–(3.3) and letting $x_i = -\log\lambda_i$ we obtain the linear program

$$\begin{aligned} \text{minimize} \sum_{j>i}(x_i + x_j) \quad &+ \quad \text{constant} \\ x_i + x_j \quad &\geq \quad -\log\rho_{ij}^k, \forall j > i \\ x_i \quad &\geq \quad 0, \forall i. \end{aligned} \tag{3.4}$$

This linear program has $k-1$ variables and $(k-1)(k-2)/2$ constraints, excluding the nonnegativity constraints. We call the approximate quantile obtained using formulation (3.4) the MinAve approximation.

Formulation (3.4) attempts to minimize the average ratio $\rho_{ij}^k/(\lambda_i\lambda_j)$. Another reasonable objective is to minimize the largest ratio. To formulate this objective we add the constraints

$$\frac{\rho_{ij}^k}{\lambda_i\lambda_j} \leq w, \forall j > i \tag{3.5}$$

and replace the objective function (3.3) with

$$\text{minimize } w.$$

After the log transformation and the substitution $x = \log w$ we obtain the linear program

$$\begin{aligned} &\text{minimize } x \\ x_i + x_j &\geq -\log \rho_{ij}^k, \forall j > i \\ x - x_i - x_j &\geq \log \rho_{ij}^k, \forall j > i \\ x_i &\geq 0, \forall i \\ x &\geq 0. \end{aligned} \tag{3.6}$$

This linear program has k variables and $(k-1)(k-2)$ constraints, excluding the nonnegativity constraints. We call the approximate quantile obtained using formulation (3.6) the MinMax approximation.

When $\rho_{ij}^k > 0, \forall i \neq j$, both linear programs are guaranteed to have a feasible solution. When some $\rho_{ij}^k < 0$, as can occur in an ANCOVA model (1.2) with common slope (see Fleiss 1986, p. 195, for example), or a three-way no-interaction model

$$Y_{ihr} = \mu + \tau_i + \beta_h + \gamma_r + \epsilon_{ihr},$$

there may not be a feasible solution. For example, if

$$\mathbf{R}_{-k} = \begin{pmatrix} 1 & -\frac{1}{4} & -\frac{1}{4} \\ & 1 & -\frac{1}{4} \\ & & 1 \end{pmatrix},$$

then there are no feasible values of $\lambda_1, \lambda_2, \lambda_3$ such that $\lambda_i\lambda_j \leq \rho_{ij}^k, \forall i \neq j$.

When there are negative correlations and a feasible solution exists, then it is possible to reformulate (3.4) and (3.6) by separating the problem of assigning values and assigning signs to the λ_i as follows:

1. Determine a feasible sign assignment for the λ_i so that $\text{sign}(\lambda_i\lambda_j) = \text{sign}(\rho_{ij}^k), \forall i \neq j$ (there will be either no or two such sign assignments).
2. For each $\rho_{ij}^k < 0$, revise the objective function term and constraints associated with ρ_{ij}^k as follows:
   (a) Replace $\lambda_i\lambda_j \leq \rho_{ij}^k$ with $\lambda_i\lambda_j \geq |\rho_{ij}^k|$ in (3.1).
   (b) Replace the ratio $\rho_{ij}^k/(\lambda_i\lambda_j)$ with $\lambda_i\lambda_j/|\rho_{ij}^k|$ in (3.3).
   (c) Replace $\rho_{ij}^k/(\lambda_i\lambda_j) \leq w$ with $\lambda_i\lambda_j/|\rho_{ij}^k| \leq w$ in (3.5).

3. Transform the problem using logarithms and solve the linear program as if all $\rho_{ij}^k > 0$. Assign the signs determined in Step 1 to the resulting λ_i to form the solution.

The remaining outstanding case is when one or more $\rho_{ij}^k = 0$. This is a particularly difficult case because the log transformation used to obtain a linear program prohibits $\lambda_i = 0$. In fact, this is a difficult case for a one-factor approximation determined by any method because if $\lambda_\ell = 0$, then $\lambda_\ell \lambda_j = 0$ for all $j \neq \ell$, meaning that all correlations $\rho_{\ell j}$ and $\rho_{j\ell}$ are approximated as 0.

In principle we could address the problem of $\rho_{ij}^k = 0$ by solving the nonlinear program (3.1)–(3.3) directly. A second approach is to set $\rho_{ij}^k = -\epsilon$, where ϵ is a small positive number, which guarantees that our linear programming solution will be conservative (provided there is a feasible sign assignment).

The following is a list of properties of the linear programs (3.4) and (3.6), which we refer to as the MinAve and MinMax formulations, respectively. In all cases we assume $|\rho_{ij}^k| < 1$ and $\rho_{ij}^k \neq 0, \forall i \neq j$.

1. When $\mathbf{R}_{-k}$ has a one-factor structure then $\mathbf{R}_{-k}^{LP} = \mathbf{R}_{-k}$ is the unique optimal solution of both linear programs.
2. Even when a feasible sign assignment exists, a feasible solution may not exist when some $\rho_{ij}^k < 0$. For example, if

$$\mathbf{R}_{-k} = \begin{pmatrix} 1 & e^{-3} & -e^{-1} \\ & 1 & -e^{-1} \\ & & 1 \end{pmatrix}$$

a positive definite correlation matrix, then a feasible sign assignment is λ_1 and λ_2 positive, and λ_3 negative. However, after adjusting for the negative correlations, the constraints in (3.4) become

$$\begin{aligned} x_1 + x_2 &\geq 3 \\ x_1 + x_3 &\leq 1 \\ x_2 + x_3 &\leq 1 \\ x_i &\geq 0. \end{aligned}$$

Clearly there is no feasible solution.

3. If $\mathbf{R}_{-k}$ does not have a one-factor structure, then the optimal solution to both linear programs may not be unique. For example, if

$$\mathbf{R}_{-k} = \begin{pmatrix} 1 & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ & 1 & \frac{1}{2} & \frac{1}{10} \\ & & 1 & \frac{1}{10} \\ & & & 1 \end{pmatrix}$$

Then $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \sqrt{\frac{1}{10}}$ and $\lambda_2 = \lambda_3 = \sqrt{\frac{1}{2}}$, $\lambda_1 = \lambda_4 = \sqrt{\frac{1}{50}}$ are both optimal solutions to MinAve.

4. In the optimal solution of either linear program, at most one $\lambda_\ell = \pm 1$. In addition, $\lambda_\ell = \pm 1$ if and only if $\rho^k_{ij} = \rho^k_{i\ell}\rho^k_{\ell j}, \forall i \neq j$.

It is important to note that if $\mathbf{R}_{-k}$ has a one-factor structure (2.3) then $\mathbf{R}^{LP}_{-k} = \mathbf{R}_{-k}$ is the unique optimal solution of both linear programs. Thus, in contrast to the probabilistic inequality methods, the LP methods will automatically produce exact simultaneous MCC confidence intervals whenever possible.

3.1 SIMULATION STUDIES

Simulations were performed to see how closely the coverage probability provided by d_{LP}, the linear programming approximation to d, comes to $1 - \alpha$. In these simulation studies the conservatism of the coverage probability caused by using the MinMax, MinAve, and Hunter–Worsley approximations,

$$\text{conservatism} \equiv (\text{true coverage probability}) - (1 - \alpha),$$

was estimated as described in Appendix B of Hsu (1992) (except the number of simulations per design was increased to 100,000). Boxplots of estimated conservatism are then given to compare the MinMax (labeled $LP1$), MinAve (labeled $LP2$), and Hunter–Worsley (labeled HW) approximations. Note that these are boxplots of *point estimates* of conservatism for *different* designs.

To see how closely the coverage probability provided by d_{LP} comes to $1 - \alpha$ in two-way models (1.3), 100 random designs were generated under the missing-completely-at-random model (Little and Rubin 1987, p. 14) for $k = b = 10$, $n_{1,1} = \cdots = n_{10,10} = 10$, where independently each observation had a .2 probability of being missing. The top half of Figure 1 gives side-by-side boxplots of point estimates of conservatism for these 100 models, for $\alpha = .10,\ .05,\ .01$. For this particular setting, both the MinAve approximation and the MinMax approximation are extremely accurate, much more so than the Hunter–Worsley approximation. The conservatism of the MinAve approximation seems slightly more variable than the conservatism of the MinMax approximation. Some of the point estimates of *conservatism* of the MinMax and MinAve approximations were negative, but not significantly so when randomness in the simulation is taken into account, the most negative ratio of estimated conservatism over its estimated standard deviation being -2.

To see how closely the coverage probability provided by d_{LP} come to $1 - \alpha$ in one-way analysis of covariance models (1.2) with a common slope, 100 random designs were generated for $k = 10$, $n_1 = \cdots = n_{10} = 10$, taking $X_1, \ldots, X_{10}$ to be iid standard normal random variables. The bottom half of Figure 1 gives side-by-side boxplots of point estimates of conservatism for these 100 models for $\alpha = .10,\ .05,\ .01$. For this particular setting, the MinAve approximation seems better than the MinMax approximation, which in turn is much better than the Hunter–Worsley approximation. A few of the point estimates of *conservatism* of the MinMax and MinAve approximations were negative, but not significantly so when randomness in the simulation is taken into account, the most negative ratio of estimated conservatism over its estimated standard deviation being about -1.57.

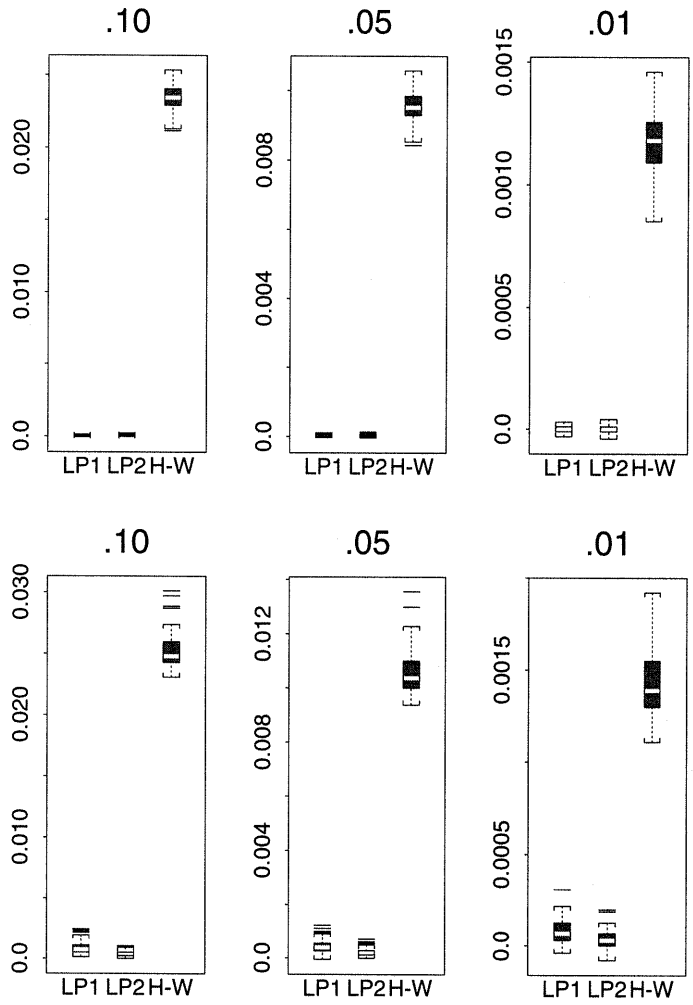

Figure 1. Conservatism of Approximations for Two-Way (top) and ANCOVA (bottom) Designs.

4. SIMULATED QUANTILE METHODS

When the LP method is not applicable, one can obtain an upper confidence bound for the desired quantile via simulation. For notational simplicity, we shall discuss the case of determining an upper confidence bound on d; the technique for $|d|$ is analogous.

Let

$$D = \max_{1 \le i \le k-1} \hat{\sigma}^{-1}(\hat{\mu}_i - \hat{\mu}_k - (\mu_i - \mu_k))/\sqrt{v_i^k}$$

and let $D_{(1)} < \cdots < D_{(M)}$ be the order statistics of a random sample $D_1, \cdots, D_M$ of D.

The approach of Edwards and Berry (1987) uses $D_{(r)}$ as a point estimate of d, assuming $r = (1-\alpha)(m+1)$ is an integer. This approach has been implemented as the

`ADJUST = SIMULATE` option for multiple comparisons under the LSMEANS option in PROC GLM and PROC MIXED of SAS Version 6.11.

In order to achieve conservatism with high confidence, instead of a point estimate of d, our approach obtains an upper confidence bound U for d. A crude level $1-\gamma$ upper confidence bound for d is $D_{(m^c)}$, where m^c is the smallest integer such that the binomial(M,α) probability of at most m^c-1 successes is at least $1-\gamma$. This upper confidence bound is obtained by inverting the usual level-α test for

$$H_0 : P\{D > d_0\} = \alpha. \tag{4.1}$$

We can obtain a sharper bound by also generating a *control variate* D^{cv} with *known* $(1-\alpha)$th quantile d_{cv} along with each D, and inverting McNemar's conditional test for

$$H_0 : P\{D > d_0\} = P\{D^{cv} > d_{cv}\}\ (=\alpha). \tag{4.2}$$

This will yield a *conditional* upper confidence bound for d. If one can generate D^{cv} to be highly correlated with D, then the test for (4.2) will be substantially more powerful than the test for (4.1) and the corresponding confidence bound will be more accurate.

We first derive the conditional upper confidence bound then provide a practical way of computing it. Let $(D_m^{cv}, D_m), m = 1, 2, \ldots, M$ be a random sample of (D^{cv}, D). For a candidate $(1-\alpha)$th quantile d_0 of D, let $n^{+-}(d_0)$ be the number of (D_m^{cv}, D_m) such that

$$D_m^{cv} > d_{cv} \quad \text{and} \quad D_m < d_0$$

and let $n^{-+}(d_0)$ be the number of (D_m^{cv}, D_m) such that

$$D_m^{cv} < d_{cv} \quad \text{and} \quad D_m > d_0.$$

Let $t(d_0) = n^{+-}(d_0) + n^{-+}(d_0)$. Under the null hypothesis

$$H_0 : d = d_0, \tag{4.3}$$

which is equivalent to (4.1) and (4.2) by the definition of d. The conditional distribution of $n^{-+}(d_0)$ given $t(d_0)$ is binomial$(t(d_0), 1/2)$, provided $t(d_0) > 0$ (see Lehmann 1986, p. 169). Suppose we reject (4.3) in favor of $H_a : d < d_0$ (which is equivalent to $H_0 : P\{D > d_0\} < \alpha$) if

$$\text{p value}(d_0) = \sum_{m=0}^{n^{-+}(d_0)} \binom{t(d_0)}{m} (1/2)^{t(d_0)} < \gamma$$

(with the understanding that $\gamma \leq .5$ and p value$(d_0) = .5$ when $t(d_0) = 0$), then by the correspondence between tests and confidence sets (see Lehmann 1986, p. 90), a $100(1-\gamma)\%$ confidence set for d is $C = \{d_0 : \text{p value}(d_0) \geq \gamma\}$. Note that setting $\gamma = .5$ results in essentially the median unbiased point estimate of d proposed by Hsu and Nelson (1990) (with a slight positive bias induced by discreteness). The computation of the upper confidence bound $U = \sup\{d_0 : d_0 \in C\}$ is simplified by the following theorem.

Theorem 1. *p value*(d_0) *is nonincreasing in* d_0.

Proof: Let $D_{(1)} < \cdots < D_{(M)}$ be the order statistics of $D_1, \ldots, D_M$. Suppose

$$D_{(m-1)} < d_0 < D_{(m)} < d_0^* < D_{(m+1)}$$

(with the understanding $D_{(0)} = -\infty$ and $D_{(M+1)} = \infty$). Then either

$$n^{-+}(d_0^*) = n^{-+}(d_0) - 1 \quad \text{and} \quad t(d_0^*) = t(d_0) - 1$$

or

$$n^{-+}(d_0^*) = n^{-+}(d_0) \quad \text{and} \quad t(d_0^*) = t(d_0) + 1.$$

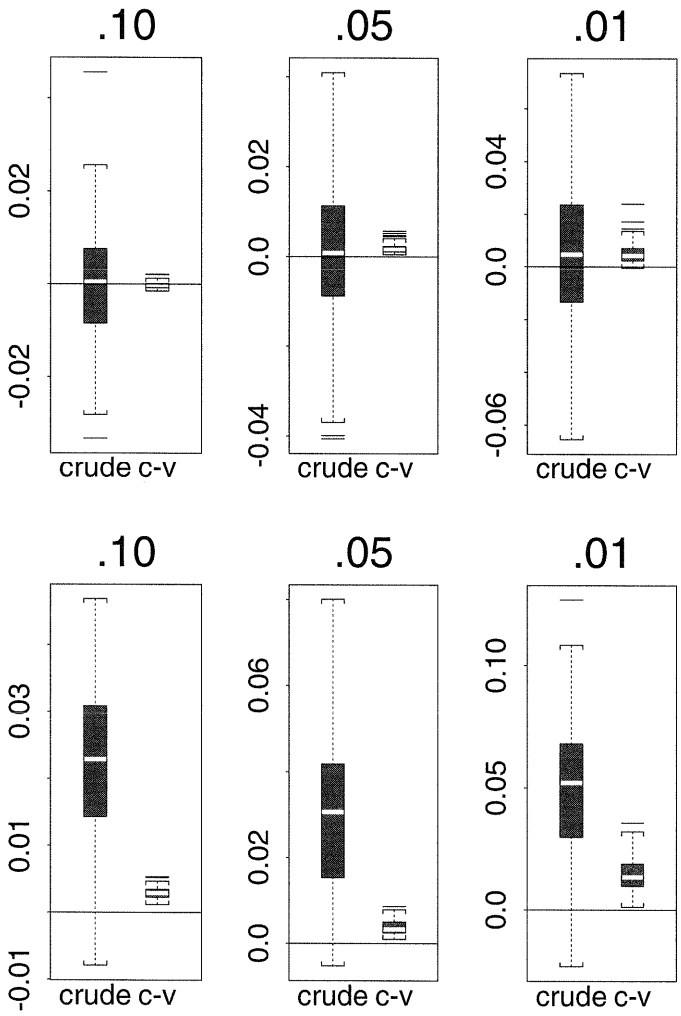

Figure 2. Bias of Simulated Quantile Estimates for a Two-Way Design (top γ = .50, *bottom* γ = .05).

By the correspondence between binomial tail probabilities and the incomplete beta function,

$$\text{p value}(d_0) = I_{.5}(t(d_0) - n^{-+}(d_0), n^{-+}(d_0) + 1), \tag{4.4}$$

and

$$\text{p value}(d_0^*) = I_{.5}(t(d_0) - n^{-+}(d_0), n^{-+}(d_0)), \tag{4.5}$$

or

$$\text{p value}(d_0^*) = I_{.5}(t(d_0) - n^{-+}(d_0) + 1, n^{-+}(d_0) + 1), \tag{4.6}$$

where

$$I_c(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^c x^{a-1}(1-x)^{b-1} dx$$

(with the understanding that p value$(d_0^*) = .5$ when $t(d_0^*) = 0$). In either case, by monotonicity and symmetry of the incomplete beta function (Abramowitz and Stegun 1972, eqns. 26.5.16 and 26.5.2), (4.4) > (4.5) and (4.4) > (4.6). □

Therefore, $U = D_{(m^{cv})}$ where m^{cv} is the smallest integer m such that p value$(D_{(m)}) < \gamma$.

To generate (D^{cv}, D) with high correlation, we propose the following strategy: Recall that if $\mathbf{R}_{-k}$ has the one-factor structure (2.3), then $|d|$ and d can be computed exactly. But a factor-analytic approximation can be used to find a correlation matrix $\mathbf{R}_{-k}^{FA}$ with one-factor structure that is close to $\mathbf{R}_{-k}$. We therefore proceed as follows:

Generate a vector of iid standard normal random variates and a scalar chi-squared variate, and transform them into a multivariate t vector with correlation $\mathbf{R}_{-k}$, and also into a multivariate t vector with correlation matrix $\mathbf{R}_{-k}^{FA}$. Because we transform common variates, it can be shown that monotonic functions of the t vectors, such as D^{cv} and D, obtain the maximum possible correlation. And since one can usually find an excellent factor-analytic approximation to $\mathbf{R}_{-k}$, as verified in Hsu (1992), we expect the correlation to be quite large.

Specifically, we use the following algorithm:

1. Use a *factor analytic* algorithm to find the $\mathbf{R}_{-k}^{FA}$ satisfying (2.3) *closest* to $\mathbf{R}_{-k}$. Let $\mathbf{C}_{-k}$ and $\mathbf{C}_{-k}^{FA}$ represent the Cholesky decompositions of $\mathbf{R}_{-k}$ and $\mathbf{R}_{-k}^{FA}$, respectively.
2. Repeat
   (a) Generate $Z_1, \ldots, Z_{k-1} \overset{\text{iid}}{\sim} \text{Normal}(0, 1)$. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_{k-1})'$.
   (b) Generate $S^2 \sim \chi_\nu^2/\nu$.
   (c) Let

$$\begin{aligned}(V_1, \ldots, V_{k-1})' &= \mathbf{C}_{-k}\boldsymbol{Z} \\ (W_1, \ldots, W_{k-1})' &= \mathbf{C}_{-k}^{FA}\boldsymbol{Z}.\end{aligned}$$

Return

$$\begin{aligned}D^{cv} &= \max\{V_1/S, \ldots, V_{k-1}/S\} \\ D &= \max\{W_1/S, \ldots, W_{k-1}/S\}.\end{aligned}$$

4.1 SIMULATION STUDIES

With the obvious change of notation for the two-sided case, we studied the stochastic behavior of the bias $|D|_{(m^c)} - |d|$ of the crude quantile estimate and the bias $|D|_{(m^{cv})} - |d|$ of the control variate quantile estimate.

We first considered the two-way model (1.3) with $k = b = 10$, $n_{1,1} = \cdots = n_{10,10} = 10$, and generated a design that is neither variance balanced nor orthogonal by giving each observation a probability of .2 of being missing, under the missing-completely-at-random model. We took the sample $1 - \alpha$ quantile from 1,000,000 simulations to be the true quantile $|d|$, and then observed *bias* of 100 pairs of quantile estimates based on 10,000 simulations of $|D|$ each using the crude simulation technique and the control variate technique. The results are summarized in Figure 2. The top three plots in

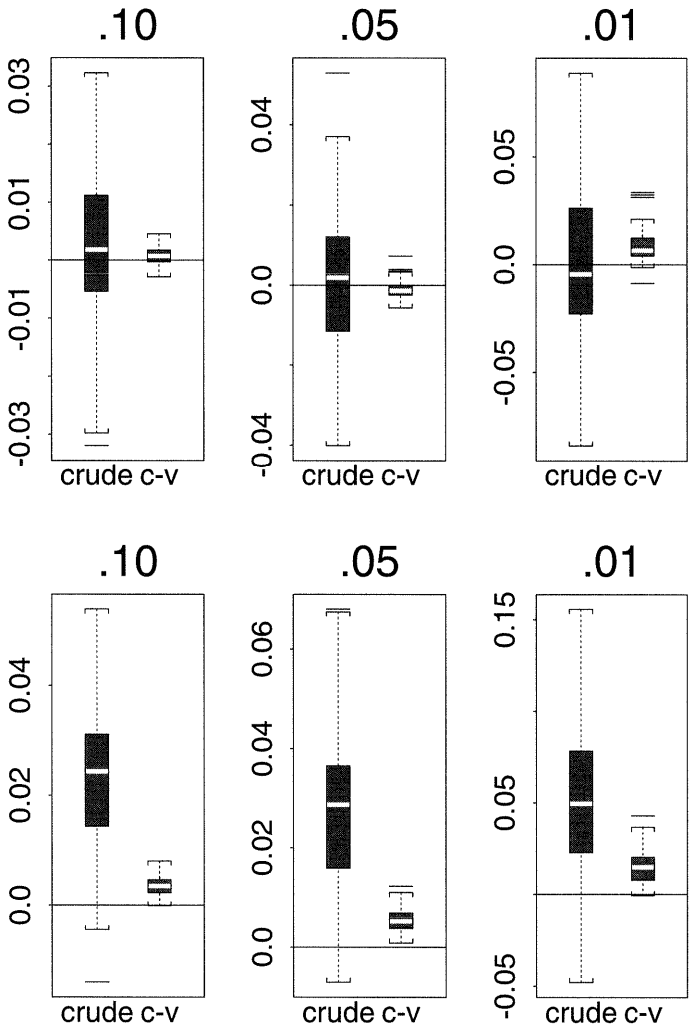

Figure 3. Bias of Simulated Quantile Estimates for an ANCOVA Design (top γ = .50, bottom γ = .05)

Figure 2 correspond to $\gamma = .50$, while the bottom three correspond to $\gamma = .05$. Within each boxplot, `crude` denotes quantile estimates obtained by crude simulation, while `c-v` denotes quantile estimates obtained using the control variate technique. Labeling in the top margin of each side-by-side boxplot denotes α.

We then considered the one-way ANCOVA model (1.2) with a common slope, $k = 10$, $n_1 = \cdots = n_{10} = 10$, and generated a design which is neither variance balanced nor orthogonal by taking $X_1, \ldots, X_{10}$ to be realizations of iid standard normal random variables. We took the sample $1 - \alpha$ quantile from 1,000,000 simulations to be the true quantile $|d|$, and then observed *bias* of 100 pairs of quantile estimates based on 10,000 simulations of $|D|$ each using the crude simulation technique and the control variate technique. The results are summarized by the side-by-side boxplots in Figure 3, which are arranged as in Figure 2.

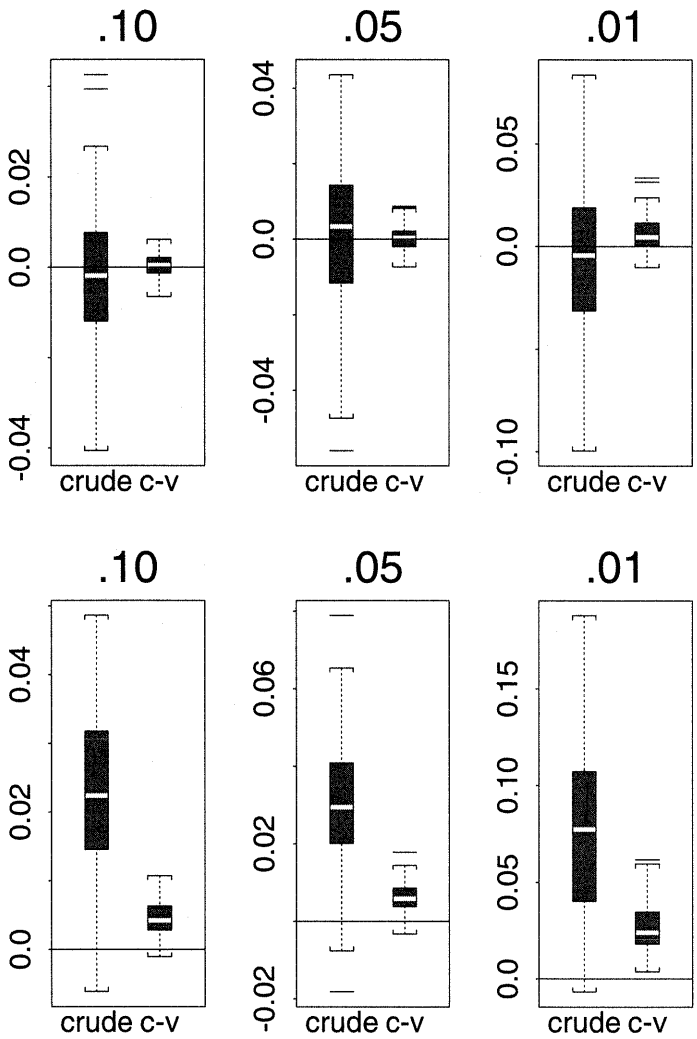

Figure 4. Bias of Simulated Quantile Estimates for a 2^5 Design (top γ = .50, bottom γ = .05).

Table 1. Timing and MSE Comparisons of Quantile Estimates

Design	α	γ	*c-v time/crude time*	*crude MSE/c-v MSE*
two-way	.10	.50	1.42	271
two-way	.10	.05	1.42	68
two-way	.05	.50	1.42	57
two-way	.05	.05	1.42	70
two-way	.01	.50	1.42	18
two-way	.01	.05	1.45	12
ANCOVA	.10	.50	1.41	62
ANCOVA	.10	.05	1.44	47
ANCOVA	.05	.50	1.44	53
ANCOVA	.05	.05	1.41	27
ANCOVA	.01	.50	1.41	10
ANCOVA	.01	.05	1.41	13
2^5	.10	.50	1.47	40
2^5	.10	.05	1.47	26
2^5	.05	.50	1.53	36
2^5	.05	.05	1.47	24
2^5	.01	.50	1.47	14
2^5	.01	.05	1.53	10

Finally, we considered the 2^k factorial model with $k = 5$, and generated a design which is neither variance balanced nor orthogonal by giving each observation a probability of .2 of being missing, under the missing-completely-at-random model. We took the sample $1 - \alpha$ quantile from 1,000,000 simulations to be the true quantile $|d|$, and then observed *bias* of 100 pairs of quantile estimates based on $10,000$ simulations of $|D|$ each using the crude simulation technique and the control variate technique. The results are summarized by the side-by-side boxplots in Figure 4, which are arranged as in Figure 2.

It is clear from Figures 2–4 that the control variate quantile estimate is significantly more accurate than the crude simulation estimate for these models, and estimates with $\gamma = .05$ are typically more conservative than estimates with $\gamma = .50$.

Table 1 compares crude simulation with the control variate technique in terms of CPU time and MSE of the quantile estimate. The column labeled "c-v time/crude time" is the estimated ratio of CPU time required by the control variate technique and crude simulation. The column labeled "crude MSE/c-v MSE" is the estimated ratio of the MSE of quantile estimates using crude simulation and the control variate technique. Clearly, the control variate technique is worth implementing.

5. EXAMPLES

The following examples illustrate the techniques discussed previously.

The first data set has an unbalanced two-way design (1.3). It is popular for illustrating unbalanced ANOVA (Fleiss 1986, p. 166; SAS 1989, p. 972). Of interest is the increase in systolic blood pressure in dogs after treatment, with disease as a blocking factor. The sample means and sample sizes (in parentheses) are given in Table 2.

Suppose 1-sided confidence bounds on $\tau_1 - \tau_4, \tau_2 - \tau_4, \tau_3 - \tau_4$ are of interest. The

Table 2. Mean Systolic Pressure and Sample Sizes (in parentheses)

Disease	*Treatment 1*	*Treatment 2*	*Treatment 3*	*Treatment 4*
1	29.333 (6)	28.000 (8)	16.333 (3)	13.600 (5)
2	28.250 (4)	33.500 (4)	4.400 (5)	12.833 (6)
3	20.400 (5)	18.167 (6)	8.500 (4)	14.200 (5)

correlation matrix $\mathbf{R}_{-4}$ is

$$\begin{pmatrix} 1 & - & - \\ .4863 & 1 & - \\ .4493 & .4515 & 1 \end{pmatrix},$$

which can be conveniently obtained from a package such as SAS with the $\tau_k = 0$ parametrization. Because $k = 4$ and all the correlations are positive, $\mathbf{R}_{-4}$ satisfies (2.3), and the LP method finds $\lambda_1 = .6957, \lambda_2 = .6990, \lambda_3 = .6458$. With error degrees of freedom $\nu = 52$, the critical values d given by the various techniques are given in Table 3 (simulated quantile estimates based on 10,000 simulations).

For this example, the LP technique is ideal but the control variate quantile estimates with $\gamma = .50$ seem acceptable as well.

The second data set comes from Scheffé (1959, p. 216), which gives breaking strength (Y) in grams and thickness (X) in 10^{-4} inch from tests on seven types of starch film (starch 1 = canna, 2 = sweet potato, 3 = corn, 4 = rice, 5 = dasheen, 6 = wheat, and 7 = potato). It is assumed that the regression coefficient of Y on X is the same for all starches.

Suppose two-sided treatments versus control confidence intervals are of interest, with potato as the control. Then the correlation matrix $\mathbf{R}_{-7}$ of $\theta_1, \ldots, \theta_6$ is

$$\begin{pmatrix} 1 & - & - & - & - & - \\ .3958 & 1 & - & - & - & - \\ .5677 & .4936 & 1 & - & - & - \\ .5468 & .4621 & .7598 & 1 & - & - \\ .5140 & .4488 & .7675 & .6930 & 1 & - \\ .5505 & .4922 & .8651 & .7738 & .7915 & 1 \end{pmatrix}.$$

The LP technique is not applicable to two-sided inference. With error degrees of freedom $\nu = 86$, the critical values $|d|$ given by the other techniques are given in Table 4 (simulated quantile estimates based on 10,000 simulations).

Table 3. Critical Values for Two-Way Example

α	*.10*	*.05*	*.01*
Bonferroni	1.873	2.186	2.826
Slepian	1.857	2.179	2.825
Hunter–Worsley	1.800	2.137	2.804
LP (exact)	1.774	2.119	2.795
Crude simulation ($\gamma = .50$)	1.764	2.090	2.768
Control variate ($\gamma = .50$)	1.776	2.121	2.804
Crude simulation ($\gamma = .05$)	1.798	2.167	2.889
Control variate ($\gamma = .05$)	1.777	2.127	2.816

Table 4. Critical Values for ANCOVA Example

α	.10	.05	.01
Bonferroni	2.442	2.701	3.246
Šidák	2.425	2.693	3.245
Hunter–Worsley	2.324	2.606	3.185
Crude simulation ($\gamma = .50$)	2.282	2.573	3.124
Control variate ($\gamma = .50$)	2.272	2.565	3.165
Crude simulation ($\gamma = .05$)	2.303	2.586	3.140
Control variate ($\gamma = .05$)	2.271	2.572	3.196
Quantile from 10^6 samples	2.265	2.561	3.159

For this example, the control variate quantile estimate with $\gamma = .50$ gave the most accurate estimates overall.

In general, we recommend the LP method when it is applicable. When the LP method is not applicable, we recommend using the control variate method with $\gamma = .05$ when a conservative quantile estimate is desired, and the control variate method with $\gamma = .50$ otherwise.

6. CONCLUSIONS AND EXTENSIONS

We have presented computationally efficient, conservative approximations for the critical values required for MCC inference in the GLM. Although these approximations are conservative, they are much less conservative than standard approximations based on Bonferroni, Sidak, Slepian, or Hunter–Worsley inequalities. Fast approximations are essential for interactive data analysis, since it is impossible to develop tables for the (uncountably infinite) number of cases that may arise.

Another type of multiple comparisons to which the techniques discussed in this article apply is multiple comparisons with the best (MCB). The parameters of interest in MCB are $\mu_i - \max_{j \neq i} \mu_j$, assuming a larger treatment effect is better. This is because if $\mu_i - \max_{j \neq i} \mu_j > 0$, then treatment i is the best treatment, while if $\mu_i - \max_{j \neq i} \mu_j < 0$, then treatment i is not the best treatment. Furthermore, even if treatment i is not the best, if $\mu_i - \max_{j \neq i} \mu_j > -\delta$, where δ is a small positive number, then it is close to the best.

Hsu (1981, 1984a, 1984b) derived simultaneous confidence intervals for $\mu_i - \max_{j \neq i} \mu_j$ in balanced and unbalanced one-way models, which have been implemented in JMP and MINITAB. Chang and Hsu (1992) showed that analogous MCB simultaneous confidence intervals can be obtained in the GLM setting, provided one can compute the k critical values for one-sided MCC when one takes each of the k treatments as the control in turn. They did not, however, address the computational problem. The techniques for computing MCC critical values discussed in the previous sections thus facilitate MCB computations.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Grants DDM-8922721 and DMI-9622065. and by Grant No. 5 R01 CA41168, awarded by the National Cancer Institute. We thank Tim Costigan for discussions on the MTP_2 property, and Randy Tobias for discussions on the implementation of multiple comparisons in the GLM.

[Received June 1996. Revised March 1997.]

REFERENCES

Abramowitz, M., and Stegun, I. A. (eds.) (1972), *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

Chang, J. Y., and Hsu, J. C. (1992), "Optimal Designs for Multiple Comparisons With the Best," *Journal of Statistical Planning and Inference*, 30, 45–62.

Dunnett, C. W. (1955), "A Multiple Comparison Procedure for Comparing Several Treatments With a Control," *Journal of the American Statistical Association*, 50, 1096–1121.

Edwards, D. G., and Berry, J. J. (1987), "The Efficiency of Simulation-Based Multiple Comparisons," *Biometrics*, 43, 913–928.

Fleiss, J. L. (1986), *The Design and Analysis of Clinical Experiments*, New York: Wiley.

Graybill, F. A. (1983), *Matrices with Applications in Statistics*, (2nd ed.), Belmont, CA: Wadsworth.

Hsu, J. C. (1981), "Simultaneous Confidence Intervals for all Distances From the 'Best'," *The Annals of Statistics*, 9, 1026–1034.

——— (1984a), "Constrained Two-Sided Simultaneous Confidence Intervals for Multiple Comparisons With the Best," *The Annals of Statistics*, 12, 1136–1144.

——— (1984b), "Ranking and Selection and Multiple Comparisons With the Best," in *Design of Experiments: Ranking and Selection*, eds. T. J. Santner, and A. C. Tamhane, New York: Marcel Dekker.

——— (1992), "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model," *Journal of Graphical and Computational Statistics*, 1, 151–168.

——— (1996), *Multiple Comparisons: Theory and Methods*, London: Chapman & Hall.

Hsu, J. C., and Nelson, B. L. (1990), "Control Variates for Quantile Estimation," *Management Science*, 36, 835–951.

Hunter, D. (1976), "An Upper Bound for the Probability of a Union," *Journal of Applied Probability*, 13, 597–603.

John, J. A. (1987), *Cyclic Designs*, London: Chapman & Hall.

Karlin, S., and Rinott, Y. (1980), "Classes of Ordering of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distribution," *Journal of Multivariate Analysis*, 10, 467–498.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, (2nd ed.), New York: Wiley.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, New York: Wiley.

SAS (1989), *SAS/STAT User's Guide, Release 6,* (4th ed., vol. 2), Cary, NC: SAS Institute, Inc.

Scheffé, H. (1959), *Analysis of Variance*, New York: Wiley.

Somerville, P. N. (1995), "Calculations of Constants for One-Step Multiple Comparison Procedures," in *Proceedings of the 25th Symposium on the Interface: Computing Science and Statistics.*

Worsley, K. J. (1982), "An Improved Bonferroni Inequality and Applications," *Biometrika*, 69, 297–302.