RANKING, ESTIMATION AND HYPOTHESIS TESTING IN UNBALANCED MODELS — A BAYESIAN APPROACH

by

Duncan K. H. Fong
Department of Management Science
The Pennsylvania State University
University Park, PA 16802

and

James O. Berger
Department of Statistics
Purdue University
West Lafayette, IN 47907

Technical Report #89-40

Department of Statistics
Purdue University

December 1989

*

Ranking, Estimation and Hypothesis Testing in Unbalanced Models - A Bayesian Approach

Duncan K. H. Fong

Assistant Professor of Management Science
Department of Management Science
The Pennsylvania State University
University Park, PA 16802

James O. Berger

Richard M. Brumfield Distinguished
Professor of Statistics
Department of Statistics
Purdue University
West Lafayette, IN 47907

This work was completed largely while Fong was a graduate student in the Department of Statistics, Purdue University. Support from U.S. National Science Foundation Grants DMS-8401996, DMS-8702620, and DMS-8717799 is gratefully acknowledged.

ABSTRACT

For the usual 2-factor additive model, there has been comparatively little work in the ranking and selection literature for the case of unequal sample sizes (unequal variances). The existing papers (e.g., Huang and Panchapakesan (1976), Gupta and Hsu (1980)) do not give explicit procedures unless assuming an equal number of observations and equal variances. However the case of unequal sample sizes may arise in many natural settings, say in the problem of designing an experiment for comparing treatments in the presence of blocks of different fixed sizes, where one may assign an equal number of experimental units to each treatment within the same block. A Bayesian approach to the problem is taken, leading to computation of the posterior probabilities that each treatment mean is the largest. In addition, a Bayesian version of ANOVA (including estimation and hypothesis testing) will be considered. Calculation of the quantities of interest involves, at worst, 5-dimensional numerical integration, for which an efficient Monte Carlo method of evaluation is given. An example is presented to illustrate the methodology.

KEY WORDS: Unbalanced designs; Hierarchical Bayes; Exchangeability; Monte Carlo Integration; Ranking Probabilities.

1 INTRODUCTION

Consider an experiment in which there are two factors A and B, with A having I levels and B having J levels. Suppose the observations can be modelled as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \; k = 1, \ldots, K_j; \; i = 1, \ldots, I; \; j = 1, \ldots, J; \tag{1.1}$$

here y_{ijk} is the kth observation on the combination of the ith level of A with the jth level of B, μ is to be thought of as an overall mean, α_i and β_j respectively describe the two main effects, and ϵ_{ijk} are independent and normally distributed with mean zero and variance σ^2, which can be known or unknown. The K_j may differ, allowing for analysis of unbalanced designs. We do not require $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$, but will instead consider α_i and β_j random as in a random effects model. The motivation here is to allow the modelling of exchangeability among the α_i and/or the β_j. As has been evidenced in the extensive empirical Bayesian and random effects literatures, there are many practical contexts in which such an assumption is warranted and desirable.

If one is interested in selecting the largest main effects from a Bayesian perspective, quantities of interest include (posterior) estimates of the α_i and β_j, together with their standard errors, and posterior probabilities such as

$$p_{i*} = Pr(\alpha_i \text{ is largest}|\text{data}, H_0^1 \text{ is false}), \; i = 1, \ldots, I, \tag{1.2}$$

$$p_{*j} = Pr(\beta_j \text{ is largest}|\text{data}, H_0^2 \text{ is false}),\ j = 1, \ldots, J; \tag{1.3}$$

here H_0^1 and H_0^2 are the ANOVA type null hypotheses

$$H_0^1 : \alpha_1 = \cdots = \alpha_I, \quad H_0^2 : \beta_1 = \cdots = \beta_J.$$

The inclusion of these hypotheses is allowed, but not required, and they can be tested individually or simultaneously. A very attractive feature of the unbalanced model considered here is that all the above quantities can be calculated with at most five dimensional numerical integration, no matter how large I and J are.

There have been many Bayesian papers in the literature dealing with ANOVA or selection and ranking problems. Among them are Hill (1965), Box and Tiao (1968), Lindley and Smith (1972), Smith (1973), Lindley (1974), Goel and Rubin (1977), Dawid (1977), Ghosh and Meeden (1984), Gupta and Yang (1985) and Berger and Deely (1988). Most related to this paper is Berger and Deely (1988) in which a hierarchical Bayesian approach to ranking and selection in one-way classification was proposed. In that paper, the important point was made that, when the variances are unequal, "order reversal" can occur, i.e., the order of the sample means need not coincide with the order of the ranking probabilities. We will consider a related hierarchical Bayesian solution to the two-way classification problem. Advantages that can accrue from the Bayesian approach, besides the simple fact that answers can be obtained in difficult unbalanced problems, include:

1. A probability vector is obtained which gives the chance of each parameter being the largest. This provides a fairly complete and easily interpretable answer to the basic selection problem.

2. If hypothesis testing is desired, Bayesian measures of evidence concerning the usual null hypothesis H_0: no treatment difference, are easier to interpret than the corresponding P-values (cf. Berger and Sellke (1987)). Also, estimation, testing, and ranking can all be done simultaneously.

3. Relationships among the parameters, such as an a priori belief in exchangeability can be incorporated into the analysis.

4. The method can be readily extended to apply to three-way and higher classification problems.

Section 2 gives the classes of prior distributions that will be considered. Section 3 presents the main results. Section 4 discusses the method of computation. Section 5 considers an example.

2 THE PRIOR DISTRIBUTION

The most convenient way to model exchangeability among the factors is through a hierarchical Bayesian approach. We will use a three stage model. The first stage priors are

$$\mu \sim N(w, \sigma_\mu^2),\ \alpha_i \sim N(0, \sigma_\alpha^2),\ \beta_j \sim N(0, \sigma_\beta^2),\quad \text{all independently;} \tag{2.1}$$

where w and σ_μ^2 are known. For all three priors a "noninformative" prior option will be allowed, namely a constant prior density can be selected. (The same effect is achieved by sending the first stage variances, σ_μ^2, σ_α^2, or σ_β^2, to infinity.) The second stage is given by

$$\begin{aligned} \pi_1(\sigma_\alpha^2) &= (1-e_1)I_0(\sigma_\alpha^2) + e_1\pi_1^*(\sigma_\alpha^2), \\ \pi_2(\sigma_\beta^2) &= (1-e_2)I_0(\sigma_\beta^2) + e_2\pi_2^*(\sigma_\beta^2), \quad \text{all independently;} \end{aligned} \tag{2.2}$$

here I_0 is the degenerate distribution which gives unit mass to the point zero, and π_1^* and π_2^* are arbitrary (but must be proper if hypothesis testing is desired). The third stage is given by

$$\begin{aligned} Pr(e_1 = 0) &= \pi_1 &= 1 - Pr(e_1 = 1), \\ Pr(e_2 = 0) &= \pi_2 &= 1 - Pr(e_2 = 1), \quad \text{all independently;} \end{aligned} \tag{2.3}$$

here π_1 and π_2 are specified prior probabilities of H_0^1 and H_0^2, respectively. The third stage probabilities could have been incorporated into the second stage; having the third stage, however, makes it easier to keep track of the calculation of posterior probabilities of the null hypotheses and simplifies notation. (Note that there is a 1-1 correspondence between $\{e_* = 0\}$ and $\{H_0^* \text{ is true}\}$. If we use H_1^* to denote the alternative hypothesis, then $\{e_* = 1\}$ is equivalent to $\{H_1^* \text{ is true}\}$.)

A traditional noninformative choice for π_l (l equals 1 or 2) would be $\frac{1}{2}$, and will be used in all examples. The following noninformative and informative choices for the π_l^*, $l = 1, 2$, will be considered.

If I and J are both greater than 3, a reasonable noninformative choice for $\pi_l^*(\sigma_\tau^2)$ is $\pi_l^*(\sigma_\tau^2) \equiv 1$, where τ can be α or β. However if $I \leq 3, J \leq 3$ or both are less than or equal to 3, then this choice may result in improper posteriors. Following Berger and Deely (1988) the noninformative priors given below (having a form typical of noninformative priors for variance components) will be used to insure that the posterior density is proper:

$$\pi_1^*(\sigma_\alpha^2) = \frac{1}{\sigma^2 / \sum_{j=1}^{J} K_j + \sigma_\alpha^2}, \tag{2.4}$$

$$\pi_2^*(\sigma_\beta^2) = \prod_{j=1}^{J} (\sigma^2 / I K_j + \sigma_\beta^2)^{\frac{-1}{J}}. \tag{2.5}$$

Two different classes of subjective proper priors will be considered, so as to allow investigation of robustness with respect to the prior.

(I) The first class of priors to be considered is

$$\pi_l^*(\sigma_\tau^2) = \frac{(m-1)c}{(1+c\sigma_\tau^2)^m} \tag{2.6}$$

where $m > 1$ and $c > 0$ (cf. Berger and Deely (1988)). If one were to subjectively elicit the median and third quartile of this prior, to be denoted, respectively, by $\rho_{.5}$ and $\rho_{.75}$, then c and m would be

$$c = \frac{\rho_{.75} - 2\rho_{.5}}{(\rho_{.5})^2}, \quad m = 1 + \frac{\log 2}{\log([\rho_{.75} - \rho_{.5}]/\rho_{.5})}.$$

(It is assumed that the elicited quartiles satisfy $\rho_{.5} < \rho_{.75}/2$; if not, a different functional form should be used.) It is interesting to note that, if $c = \sum_{j=1}^J K_j/\sigma^2$ and $\tau = \alpha$, the noninformative prior in (2.4) is the renormalized limit of these proper priors as $m \to 1$.

(II) The second class of proper priors that will be considered (cf. Deely and Zimmer (1987)) is

$$\pi_l^*(\sigma_\tau^2) = \frac{r}{d} I_{\{\sigma_\tau^2 \le d\}}(\sigma_\tau^2) + \frac{r}{d}\left(\frac{d}{\sigma_\tau^2}\right)^{1/(1-r)} I_{\{\sigma_\tau^2 > d\}}(\sigma_\tau^2) \tag{2.7}$$

where $I_{\{\cdot\}}(\cdot)$ is the usual indicator function; thus we assume that σ_τ^2 has a constant density between 0 and d, the region having probability r, with the density tailing off at a polynomial rate. If the mean of the prior exists, then d is $\frac{2(2r-1)}{r}$ times the mean. Sometimes it is easy to specify r; then by varying d to reflect the uncertainty in eliciting the prior mean, one may investigate the influence of the prior on the final answers.

Finally, if σ^2 is unknown and an accurate estimate of it is not available, then it will be given a noninformative prior; both $\pi^*(\sigma^2) = 1$ and $\pi^*(\sigma^2) = \sigma^{-2}$ are reasonable.

3 THE MAIN RESULTS

3.1 Known σ^2

3.1.1 The Posterior Distribution of $\sigma_\alpha^2, \sigma_\beta^2$

We start with the calculation of $\pi(\sigma_\alpha^2, \sigma_\beta^2|\mathbf{y}, \mathbf{e})$, where $\mathbf{e}= (e_1, e_2)$, $\mathbf{y}= (y_{111}, \ldots, y_{IJK_J})^T$. We shall consider the cases (1) $\mathbf{e}$ = (1, 1), (2) $\mathbf{e}$ = (1, 0), and (3) $\mathbf{e}$ = (0, 1). To simplify the notation, $\pi_{11}(\sigma_\alpha^2, \sigma_\beta^2|\mathbf{y}), \pi_{10}(\sigma_\alpha^2|\mathbf{y})$ and $\pi_{01}(\sigma_\beta^2|\mathbf{y})$ will be used to denote the densities in these three cases.

Theorem 3.1

$$\begin{aligned}\pi_{11}(\sigma_\alpha^2, \sigma_\beta^2|\mathbf{y}) &= K_{11}^{-1}\sigma_\mu(\sigma^2/N_K + \sigma_\alpha^2)^{-(I-1)/2}[\prod_{j=1}^{J}(\sigma^2/K_j + I\sigma_\beta^2)]^{-1/2} \\ &\times[1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2]^{-1/2}\exp\{\frac{-1}{2}[\frac{S_1}{\sigma^2/N_K + \sigma_\alpha^2}\end{aligned}$$

$$+I\sum_{j=1}^{J}\frac{(y_{\cdot j\cdot}-\tilde{y}_{\dots})^2}{\sigma^2/K_j+I\sigma_\beta^2}+\frac{I(\tilde{y}_{\dots}-w)^2}{\sigma_b^2+\sigma_\alpha^2+I\sigma_\mu^2}]\}\pi_1^*(\sigma_\alpha^2)\pi_2^*(\sigma_\beta^2), \quad (3.1)$$

where $y_{\cdot j\cdot}=\frac{\sum_{i=1}^{I}\sum_{k=1}^{K_j}y_{ijk}}{IK_j}$, $\tilde{y}_{\dots}=\frac{\sigma_b^2}{I}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K_j}\frac{y_{ijk}}{\sigma^2+IK_j\sigma_\beta^2}$,

$\sigma_b^2=(\sum_{j=1}^{J}\frac{K_j}{\sigma^2+IK_j\sigma_\beta^2})^{-1}$, $N_K=\sum_{j=1}^{J}K_j$,

$S_1=\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\dots})^2$, $\bar{y}_{i\cdot\cdot}=\frac{\sum_{j=1}^{J}\sum_{k=1}^{K_j}y_{ijk}}{N_K}$, $\bar{y}_{\dots}=\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K_j}y_{ijk}}{IN_K}$,

and K_{11} *is the appropriate normalizing constant.*

Proof. Given in the appendix. □

Since $\mathbf{e}=(1,\ 0)$ and $\mathbf{e}=(0,\ 1)$ implies $\sigma_\beta^2=0$ and $\sigma_\alpha^2=0$ with probability 1, respectively, the following corollary is immediate.

Corollary 3.2

$$\begin{aligned}\pi_{10}(\sigma_\alpha^2|\mathbf{y}) &= K_{10}^{-1}\sigma_\mu(\sigma^2/N_K+\sigma_\alpha^2)^{-(I-1)/2}[1+N_K(\sigma_\alpha^2+I\sigma_\mu^2)/\sigma^2]^{-1/2}\\ &\quad\times\exp\{\frac{-1}{2}[\frac{S_1}{\sigma^2/N_K+\sigma_\alpha^2}+\frac{I(\bar{y}_{\dots}-w)^2}{\sigma^2/N_K+\sigma_\alpha^2+I\sigma_\mu^2}]\}\pi_1^*(\sigma_\alpha^2), \quad (3.2)\\ \pi_{01}(\sigma_\beta^2|\mathbf{y}) &= K_{01}^{-1}\sigma_\mu[\prod_{j=1}^{J}(\sigma^2/K_j+I\sigma_\beta^2)]^{-1/2}[1+(I\sigma_\mu^2)/\sigma_b^2]^{-1/2}\\ &\quad\times\exp\{\frac{-I}{2}[\sum_{j=1}^{J}\frac{(y_{\cdot j\cdot}-\tilde{y}_{\dots})^2}{\sigma^2/K_j+I\sigma_\beta^2}+\frac{(\tilde{y}_{\dots}-w)^2}{\sigma_b^2+I\sigma_\mu^2}]\}\pi_2^*(\sigma_\beta^2), \quad (3.3)\end{aligned}$$

where K_{10} *and* K_{01} *are the appropriate normalizing constants.*

It is interesting to note that if we send σ_μ^2 to infinity, representing vague prior information about μ, the density in (3.1) is just the product of the densities in (3.2) and (3.3) (as $\sigma_\mu^2 \to \infty$). This observation is important in later consideration of "collapsibility". We shall formally state this as a lemma for the usual noninformative prior on μ.

Lemma 3.3 *If $\pi(\mu) = 1$, then*

$$\pi_{11}(\sigma_\alpha^2, \sigma_\beta^2|\mathbf{y}) = \pi_{10}(\sigma_\alpha^2|\mathbf{y})\pi_{01}(\sigma_\beta^2|\mathbf{y}) \tag{3.4}$$

where

$$\pi_{10}(\sigma_\alpha^2|\mathbf{y}) \propto (\sigma^2/N_K + \sigma_\alpha^2)^{-(I-1)/2} \exp\{\frac{-S_1}{2(\sigma^2/N_K + \sigma_\alpha^2)}\}\pi_1^*(\sigma_\alpha^2),$$

$$\pi_{01}(\sigma_\beta^2|\mathbf{y}) \propto \sigma_b[\prod_{j=1}^{J}(\sigma^2/K_j + I\sigma_\beta^2)]^{-1/2} \exp\{\frac{-I}{2}\sum_{j=1}^{J}\frac{(y_{\cdot j\cdot} - \tilde{y}_{\dots})^2}{\sigma^2/K_j + I\sigma_\beta^2}\}\pi_2^*(\sigma_\beta^2).$$

Also, π_{10} is the posterior density of σ_α^2 if the noninformative prior $\pi(\beta_j) \equiv 1$ is used while π_{01} is the posterior density of σ_β^2 when $\pi(\alpha_i) \equiv 1$ is used.

Proof. Given in the appendix. □

3.1.2 Hypothesis Testing

Consider testing the null hypotheses $H_0^1 : \alpha_1 = \cdots = \alpha_I$ and $H_0^2 : \beta_1 = \cdots = \beta_J$. Define

$$L_1 = (\sigma^2/N_K)^{\frac{-(I-1)}{2}} \exp\{\frac{-S_1}{2\sigma^2/N_K}\}, \tag{3.5}$$

$$L_2 = [\prod_{j=1}^{J}(\sigma^2/K_j)]^{-1/2}\exp\{\frac{-I}{2}\sum_{j=1}^{J}\frac{(y_{.j.}-\bar{y}_{...})^2}{\sigma^2/K_j}\}, \quad (3.6)$$

$$L_3 = [1+IN_K\sigma_\mu^2/\sigma^2]^{-1/2}\exp\{\frac{-I(\bar{y}_{...}-w)^2}{2(\sigma^2/N_K+I\sigma_\mu^2)}\}. \quad (3.7)$$

Let

$$K'_{00} = L_1L_2L_3,\ K'_{10} = K_{10}L_2,\ K'_{01} = K_{01}L_1,\ K'_{11} = K_{11}.$$

Then, if π_1 and π_2 are the prior probabilities of H_0^1 and H_0^2, respectively, and $H_1^l, l = 1, 2$, are the alternative hypotheses (complements of the null H_0^l), it is straightforward to show, for $i = 1, 2$, that

$$\begin{aligned} P_{H_0^i|H_{e_{3-i}}^{3-i}} &= Pr(H_0^i \text{ is true}|\mathbf{y}, H_{e_{3-i}}^{3-i} \text{ is true}) \\ &= [1+\frac{(1-\pi_i)}{\pi_i}\frac{K'_{e'_1e'_2}}{K'_{e_1e_2}}]^{-1}, \quad (3.8) \\ P_{H_0^i} &= Pr(H_0^i \text{ is true}|\mathbf{y}) \\ &= [1+\frac{(1-\pi_i)}{\pi_i}\frac{(K'_{e_1^ie_2^i}\pi_{3-i}+K'_{11}(1-\pi_{3-i}))}{(K'_{00}\pi_{3-i}+K'_{e_2^ie_1^i}(1-\pi_{3-i}))}]^{-1}, \quad (3.9) \end{aligned}$$

where $e_l = 0$ or 1; $e'_l = 1$ if $l = i$, $e'_l = e_l$ if $l \neq i$; and $e_l^i = 1$ if $l = i$, and 0 otherwise.

The joint posterior probabilities of the null hypotheses and their alternatives are, in tabular form,

	H_0^1	H_1^1
H_0^2	$P_{H_0^1\|H_0^2}P_{H_0^2}$	$(1-P_{H_0^1\|H_0^2})P_{H_0^2}$
H_1^2	$P_{H_0^1\|H_1^2}(1-P_{H_0^2})$	$(1-P_{H_0^1\|H_1^2})(1-P_{H_0^2})$

e.g. $Pr(H_0^1$ and H_0^2 both true$|\mathbf{y}) = P_{H_0^1|H_0^2} P_{H_0^2}$, etc.

3.1.3 Ranking and Selection

Define

$$p_{i*|H_1^1,H_{e_2}^2} = Pr(\alpha_i \text{ is largest}|\mathbf{y}, H_1^1, H_{e_2}^2 \text{ are true})$$

and

$$p_{*j|H_{e_1}^1,H_1^2} = Pr(\beta_j \text{ is largest}|\mathbf{y}, H_{e_1}^1, H_1^2 \text{ are true}).$$

Lemma 3.4

$$p_{i*|H_1^1,H_0^2} = E^{\pi_{10}(\sigma_\alpha^2|\mathbf{y})}[\eta_i(\sigma_\alpha^2)], \quad (3.10)$$

$$p_{i*|H_1^1,H_1^2} = E^{\pi_{11}(\sigma_\alpha^2,\sigma_\beta^2|\mathbf{y})}[\eta_i(\sigma_\alpha^2)], \quad (3.11)$$

$$p_{*j|H_0^1,H_1^2} = E^{\pi_{01}(\sigma_\beta^2|\mathbf{y})}[\psi_j(0,\sigma_\beta^2)], \quad (3.12)$$

$$p_{*j|H_1^1,H_1^2} = E^{\pi_{11}(\sigma_\alpha^2,\sigma_\beta^2|\mathbf{y})}[\psi_j(\sigma_\alpha^2,\sigma_\beta^2)], \quad (3.13)$$

where $\eta_i(\sigma_\alpha^2) = E^Z \prod_{r\neq i} \Phi(Z + \frac{N_K \sigma_\alpha (\bar{y}_{i..} - \bar{y}_{r..})}{\sigma\sqrt{\sigma^2 + N_K \sigma_\alpha^2}})$;

$\psi_j(\sigma_\alpha^2,\sigma_\beta^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{\prod_{s\neq j}\Phi(\frac{\phi_j - u_s}{\sqrt{V_s}})\}\pi_{1,1}^*(\phi_j)\pi_{2,1}^*(\beta)d\phi_j d\beta$;

$u_s = y_{.s.} - \frac{\sigma^2}{\sigma^2 + IK_s\sigma_\beta^2}(y_{.s.} - \beta),\ V_s = \frac{\sigma^2\sigma_\beta^2}{\sigma^2 + IK_s\sigma_\beta^2}$;

$\pi_{1,1}^*(\phi_j)$ *is a* $N(u_j, V_j)$ *density;*

$\pi^*_{2,1}(\beta)$ *is a* $N(u^*, V^*)$ *density, where*

$$u^* = \tilde{y}_{...} - [1 + (\sigma_\alpha^2 + I\sigma_\mu^2)\sigma_b^{-2}]^{-1}(\tilde{y}_{...} - w) \text{ and}$$

$$V^* = [I((\sigma_\alpha^2 + I\sigma_\mu^2)^{-1} + \sigma_b^{-2})]^{-1};$$

Z is a standard normal random variable and

Φ is the standard normal c.d.f.

Proof. Given in the appendix. □

Theorem 3.5

$$\begin{aligned} p_{i*} &= Pr(\alpha_i \text{ is largest}|\mathbf{y}, H_1^1 \text{ is true}) \\ &= p_{i*|H_1^1,H_0^2} P_{H_0^2|H_1^1} + p_{i*|H_1^1,H_1^2}(1 - P_{H_0^2|H_1^1}), \quad (3.14) \\ p_{*j} &= Pr(\beta_j \text{ is largest}|\mathbf{y}, H_1^2 \text{ is true}) \\ &= p_{*j|H_0^1,H_1^2} P_{H_0^1|H_1^2} + p_{*j|H_1^1,H_1^2}(1 - P_{H_0^1|H_1^2}), \quad (3.15) \end{aligned}$$

where $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

Proof. Follows directly from Bayes' theorem. □

To compute p_{i*} and p_{*j}, 3- and 4-dimensional numerical integration, respectively, is needed. However, if μ is given a vague prior, then one level of integration can be eliminated.

Corollary 3.6 *If $\pi(\mu) = 1$, then*

$$p_{i*} = E^{\pi_{10}(\sigma_\alpha^2|\mathbf{y})}[\eta_i(\sigma_\alpha^2)], \tag{3.16}$$

$$p_{*j} = E^{\pi_{01}(\sigma_\beta^2|\mathbf{y})}[\psi_j(\sigma_\beta^2)], \tag{3.17}$$

where $\psi_j(\sigma_\beta^2)$ is given by the expression for $\psi_j(\sigma_\alpha^2, \sigma_\beta^2)$ in Lemma 3.4 with the changes $u^* = \tilde{y}_{...}$ and $V^* = \sigma_b^2/I$. The quantities will now involve 2- and 3-dimensional integration, respectively, and both are independent of π_1 and π_2. Equations (3.16) and (3.17) also hold if the noninformative priors $\pi(\beta_j) \equiv 1$ and $\pi(\alpha_i) \equiv 1$ are used, respectively.

Proof. It follows from Lemma 3.3, Lemma 3.4 and Theorem 3.5. □

3.1.4 Estimation

We begin by establishing the following lemma.

Lemma 3.7

$$E(\alpha_i|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) = \frac{\sigma_\alpha^2(\bar{y}_{i..} - \bar{y}_{...})}{\sigma^2/N_K + \sigma_\alpha^2} + \frac{\sigma_\alpha^2(\tilde{y}_{...} - w)}{\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2}; \tag{3.18}$$

$$var(\alpha_i|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) = \frac{\sigma_\alpha^2\sigma^2}{\sigma^2 + N_K\sigma_\alpha^2}(\frac{I-1}{I}) + \frac{\sigma_\alpha^2(\sigma_b^2 + I\sigma_\mu^2)}{I(\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2)}; \tag{3.19}$$

$$cov(\alpha_j, \alpha_s|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) = \frac{-\sigma_\alpha^2\sigma^2}{I(\sigma^2 + N_K\sigma_\alpha^2)} + \frac{\sigma_\alpha^2(\sigma_b^2 + I\sigma_\mu^2)}{I(\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2)}; \tag{3.20}$$

$$E(\beta_j|\mathbf{y},\sigma_\alpha^2,\sigma_\beta^2) = \frac{I\sigma_\beta^2\{y_{.j.}-\tilde{y}_{...}+[1+(\sigma_\alpha^2+I\sigma_\mu^2)/\sigma_b^2]^{-1}(\tilde{y}_{...}-w)\}}{\sigma^2/K_j+I\sigma_\beta^2}; \quad (3.21)$$

$$var(\beta_j|\mathbf{y},\sigma_\alpha^2,\sigma_\beta^2) = \frac{\sigma^2\sigma_\beta^2}{\sigma^2+IK_j\sigma_\beta^2}+\frac{IK_j^2\sigma_\beta^4}{(\sigma^2+IK_j\sigma_\beta^2)^2}\frac{(\sigma_\alpha^2+I\sigma_\mu^2)}{[1+(\sigma_\alpha^2+I\sigma_\mu^2)/\sigma_b^2]}; \quad (3.22)$$

$$cov(\beta_j,\beta_s|\mathbf{y},\sigma_\alpha^2,\sigma_\beta^2) = \frac{IK_jK_s\sigma_\beta^4}{(\sigma^2+IK_j\sigma_\beta^2)(\sigma^2+IK_s\sigma_\beta^2)}\frac{(\sigma_\alpha^2+I\sigma_\mu^2)}{[1+(\sigma_\alpha^2+I\sigma_\mu^2)/\sigma_b^2]}, \quad (3.23)$$

where $j \neq s$.

Proof. Given in the appendix. □

Using Lemma 3.7, one may obtain estimates and variances of α_i and β_j. All contrasts $\alpha_i - \alpha_r$ and $\beta_j - \beta_s$ can then be analyzed.

Theorem 3.8 *Letting τ_i denote either α_i or β_i, and recalling that e_1 and e_2 are either 0 or 1 (with $e_1 = 0$ and $e_2 = 0$ corresponding to H_0^1 and H_0^2 being true, respectively, and $e_l = 1$ corresponding to the relevant hypothesis being false),*

$$E(\tau_i|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true) = E^{\pi_{e_1e_2}(\sigma_\alpha^2,\sigma_\beta^2|\mathbf{y})}E(\tau_i|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2), \quad (3.24)$$

$$var(\tau_i|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true) = E^{\pi_{e_1e_2}(\sigma_\alpha^2,\sigma_\beta^2|\mathbf{y})}\{var(\tau_i|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2)$$

$$+[E(\tau_i|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2)-E(\tau_i|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true)]^2\}, \quad (3.25)$$

$$cov(\tau_i,\tau_r|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true) = E^{\pi_{e_1e_2}(\sigma_\alpha^2,\sigma_\beta^2|\mathbf{y})}\{cov(\tau_i,\tau_r|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2)$$

$$+[E(\tau_i|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2)-E(\tau_i|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true)]$$

$$\times[E(\tau_r|\mathbf{y},e_1\sigma_\alpha^2,e_2\sigma_\beta^2)-E(\tau_r|\mathbf{y},H_{e_1}^1,H_{e_2}^2 \ true)]\}; \quad (3.26)$$

here $E(\tau_i|\mathbf{y}, e_1\sigma_\alpha^2, e_2\sigma_\beta^2)$, $var(\tau_i|\mathbf{y}, e_1\sigma_\alpha^2, e_2\sigma_\beta^2)$, *and* $cov(\tau_i, \tau_r|\mathbf{y}, e_1\sigma_\alpha^2, e_2\sigma_\beta^2)$ *represent the mean and the variance of* τ_i *and the covariance of* (τ_i, τ_r) *conditional on* $(\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2)$, *calculated using Lemma 3.7. (Note that, when* $e_l = 0$, *one simply replaces the corresponding* σ_α^2 *or* σ_β^2 *by zero in (3.18) through (3.23).) Also,*

$$
\begin{aligned}
E(\tau_i|\mathbf{y}, H_1^l \text{ is true}) = {} & P_{H_0^{3-l}|H_1^l} E(\tau_i|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& + (1 - P_{H_0^{3-l}|H_1^l}) E(\tau_i|\mathbf{y}, H_1^1, H_1^2 \text{ true}), \qquad (3.27)
\end{aligned}
$$

$$
\begin{aligned}
var(\tau_i|\mathbf{y}, H_1^l \text{ is true}) = {} & P_{H_0^{3-l}|H_1^l}\, var(\tau_i|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& + (1 - P_{H_0^{3-l}|H_1^l}) var(\tau_i|\mathbf{y}, H_1^1, H_1^2 \text{ true}) \\
& + P_{H_0^{3-l}|H_1^l}(1 - P_{H_0^{3-l}|H_1^l})[E(\tau_i|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& - E(\tau_i|\mathbf{y}, H_1^1, H_1^2 \text{ true})]^2, \qquad (3.28)
\end{aligned}
$$

$$
\begin{aligned}
cov(\tau_i, \tau_r|\mathbf{y}, H_1^l \text{ is true}) = {} & P_{H_0^{3-l}|H_1^l}\, cov(\tau_i, \tau_r|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& + (1 - P_{H_0^{3-l}|H_1^l}) cov(\tau_i, \tau_r|\mathbf{y}, H_1^1, H_1^2 \text{ true}) \\
& + P_{H_0^{3-l}|H_1^l}(1 - P_{H_0^{3-l}|H_1^l})[E(\tau_i|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& - E(\tau_i|\mathbf{y}, H_1^1, H_1^2 \text{ true})][E(\tau_r|\mathbf{y}, H_{e_l^1}^1, H_{e_l^2}^2 \text{ true}) \\
& - E(\tau_r|\mathbf{y}, H_1^1, H_1^2 \text{ true})], \qquad (3.29)
\end{aligned}
$$

where $e_l^m = 1$ *if* $l = m$, *and 0 otherwise, and* $l = 1, 2$.

<u>Proof</u>. Follows from Bayes' Theorem. □

In the case of a vague prior on μ, we have the following corollary.

Corollary 3.9 *If $\pi(\mu) = 1$, then*

$$
\begin{aligned}
E(\alpha_i|\mathbf{y}, H_1^1 \; true) &= E^{\pi_{10}(\sigma_\alpha^2|\mathbf{y})}[\frac{N_K\sigma_\alpha^2}{\sigma^2 + N_K\sigma_\alpha^2}(\bar{y}_{i..} - \bar{y}_{...})], \qquad (3.30)\\
var(\alpha_i|\mathbf{y}, H_1^1 \; true) &= E^{\pi_{10}(\sigma_\alpha^2|\mathbf{y})}\{\frac{I\sigma_\alpha^2\sigma^2 + N_K\sigma_\alpha^4}{I(\sigma^2 + N_K\sigma_\alpha^2)} + [\frac{N_K\sigma_\alpha^2}{\sigma^2 + N_K\sigma_\alpha^2}(\bar{y}_{i..} - \bar{y}_{...})\\
&\quad -E(\alpha_i|\mathbf{y}, H_1^1 \; true)]^2\}, \qquad (3.31)\\
cov(\alpha_i, \alpha_r|\mathbf{y}, H_1^1 \; true) &= E^{\pi_{10}(\sigma_\alpha^2|\mathbf{y})}\{\frac{N_K\sigma_\alpha^4}{I(\sigma^2 + N_K\sigma_\alpha^2)}\\
&\quad +[\frac{N_K\sigma_\alpha^2(\bar{y}_{i..} - \bar{y}_{...})}{\sigma^2 + N_K\sigma_\alpha^2} - E(\alpha_i|\mathbf{y}, H_1^1 \; true)]\\
&\quad \times[\frac{N_K\sigma_\alpha^2(\bar{y}_{r..} - \bar{y}_{...})}{\sigma^2 + N_K\sigma_\alpha^2} - E(\alpha_r|\mathbf{y}, H_1^1 \; true)]\}, \qquad (3.32)\\
E(\beta_j|\mathbf{y}, H_1^2 \; true) &= E^{\pi_{01}(\sigma_\beta^2|\mathbf{y})}[\frac{IK_j\sigma_\beta^2}{\sigma^2 + IK_j\sigma_\beta^2}(y_{.j.} - \tilde{y}_{...})], \qquad (3.33)\\
var(\beta_j|\mathbf{y}, H_1^2 \; true) &= E^{\pi_{01}(\sigma_\beta^2|\mathbf{y})}\{\frac{\sigma^2\sigma_\beta^2}{\sigma^2 + IK_j\sigma_\beta^2} + \frac{IK_j^2\sigma_b^2\sigma_\beta^4}{(\sigma^2 + IK_j\sigma_\beta^2)^2}\\
&\quad +[\frac{IK_j\sigma_\beta^2(y_{.j.} - \tilde{y}_{...})}{\sigma^2 + IK_j\sigma_\beta^2} - E(\beta_j|\mathbf{y}, H_1^2 \; true)]^2\}, \qquad (3.34)\\
cov(\beta_j, \beta_s|\mathbf{y}, H_1^2 \; true) &= E^{\pi_{01}(\sigma_\beta^2|\mathbf{y})}\{\frac{IK_jK_s\sigma_b^2\sigma_\beta^4}{(\sigma^2 + IK_j\sigma_\beta^2)(\sigma^2 + IK_s\sigma_\beta^2)}\\
&\quad +[\frac{IK_j\sigma_\beta^2(y_{.j.} - \tilde{y}_{...})}{\sigma^2 + IK_j\sigma_\beta^2} - E(\beta_j|\mathbf{y}, H_1^2 \; true)]\\
&\quad \times[\frac{IK_s\sigma_\beta^2(y_{.s.} - \tilde{y}_{...})}{\sigma^2 + IK_s\sigma_\beta^2} - E(\beta_s|\mathbf{y}, H_1^2 \; true)]\}. \qquad (3.35)
\end{aligned}
$$

These quantities do not depend on π_1 nor π_2. The above expressions also give the posterior estimates of α_i and β_j if the priors $\pi(\beta_j) \equiv 1$ and $\pi(\alpha_i) \equiv 1$ are used, *respectively.*

Proof. Apply Lemma 3.3, Lemma 3.7 and Theorem 3.8. □

Note that Theorem 3.8 and Corollary 3.9 allow estimation and construction of variances for contrasts $\alpha_i - \alpha_r$ and $\beta_j - \beta_s$, even in conjunction with hypothesis testing.

3.1.5 Collapsibility

It is clear from Lemma 3.3, Corollaries 3.6 and 3.9 that, when $\pi(\mu) = 1$, the ranking probabilities, posterior means and variances can be calculated by integrating over either σ^2_α or σ^2_β alone. These results are shown here to be the same as those obtained from the one-way models constructed by collapsing the 2-way model over rows or columns. The same is true if either $\pi(\alpha_i) \equiv 1$ or $\pi(\beta_j) \equiv 1$. Furthermore it is interesting to note that, contrary to the result in Corollary 3.9 for $\pi(\mu) = 1$, numerical integration is not needed to obtain estimates and variances of $\alpha_i - \alpha_r$ (given $\pi(\alpha_i) \equiv 1$) and $\beta_j - \beta_s$ (given $\pi(\beta_j) \equiv 1$). This result does not apply for estimation of individual α_i and β_j.

Theorem 3.10 *If $\pi(\mu) \equiv 1$, $\pi(\alpha_i) \equiv 1$, or $\pi(\beta_j) \equiv 1$, the ranking probabilities, posterior means and variances associated with an additive model can be found from the* *reduced one-way models obtained by collapsing over rows or columns, respectively. In particular, given* $\mathbf{y}$, *the posterior distributions of* $\alpha_i - \alpha_r$ *(with* $\pi(\alpha_i) \equiv 1$*) and* $\beta_j - \beta_s$ *(with* $\pi(\beta_j) \equiv 1$*) are* $N(\bar{y}_{i..} - \bar{y}_{r..}, 2\sigma^2/N_K)$ *and* $N(y_{.j.} - y_{.s.}, (\sigma^2/K_j + \sigma^2/K_s)/I)$, *respectively. Furthermore,* $cov(\alpha_i - \alpha_r, \alpha_i - \alpha_l) = \sigma^2/N_K, cov(\beta_j - \beta_s, \beta_j - \beta_l) = \sigma^2/(IK_j)$ *and* $cov(\alpha_i - \alpha_j, \alpha_k - \alpha_l) = cov(\beta_i - \beta_j, \beta_k - \beta_l) = 0$.

Proof. Given in the appendix. □

3.2 Unknown σ^2

Conditional on H_1^1 and H_1^2 being true, the posterior distribution of $\sigma_\alpha^2, \sigma_\beta^2$ and σ^2 is given by

Theorem 3.11

$$\begin{aligned}\pi_{11}(\sigma_\alpha^2, \sigma_\beta^2, \sigma^2|\mathbf{y}) &= K_{111}^{-1}\sigma_\mu\sigma^{-IN_K+I+J-1}(\sigma^2/N_K + \sigma_\alpha^2)^{-(I-1)/2} \\ &\times[\prod_{j=1}^{J}(\sigma^2/K_j + I\sigma_\beta^2)]^{-1/2}[1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2]^{-1/2} \\ &\times \exp\{\frac{-1}{2}[\frac{S_3}{\sigma^2} + \frac{S_1}{\sigma^2/N_K + \sigma_\alpha^2} + I\sum_{j=1}^{J}\frac{(y_{.j.} - \tilde{y}_{...})^2}{\sigma^2/K_j + I\sigma_\beta^2} \\ &+ \frac{I(\tilde{y}_{...} - w)^2}{\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2}]\}\pi_1^*(\sigma_\alpha^2)\pi_2^*(\sigma_\beta^2)\pi^*(\sigma^2), \qquad (3.36)\end{aligned}$$

where $S_3 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K_j}[(y_{ijk} - y_{ij.})^2 + (y_{ij.} - \bar{y}_{i..} - y_{.j.} + \bar{y}_{...})^2]$,

$y_{ij.} = \sum_{k=1}^{K_j} y_{ijk}/K_j$ *and* K_{111} *is the appropriate normalizing constant.*

Proof. This is essentially identical to the proof of Theorem 3.1. □

From the theorem, one may also obtain the posterior densities $\pi_{10}(\sigma_\alpha^2, \sigma^2|\mathbf{y})$ and $\pi_{01}(\sigma_\beta^2, \sigma^2|\mathbf{y})$ by substituting zero for σ_β^2 and σ_α^2 in (3.36), respectively. The corresponding normalizing constants, K_{101} and K_{011}, can be found from the resulting densities. For hypothesis testing, results shown above in the known σ^2 case will still hold if the following substitutions are made:

$$K_{11}' = K_{111},\ K_{10}' = K_{101},\ K_{01}' = K_{011},\ K_{00}' = K_{001},$$

where K_{001} is the normalizing constant for the expression in (3.36) when $\sigma_\alpha^2 \equiv 0$ and $\sigma_\beta^2 \equiv 0$.

If a noninformative prior option has not been exercised for any of the parameters μ, α_i and β_j, then the previous results on ranking and selection and on estimation will remain true with the obvious modification that the expectations are now taken with respect to the new densities (hence an additional integration over σ^2 is needed).

When $\pi(\mu) = 1$, it is no longer true that π_{11} equals $\pi_{10}\pi_{01}$; thus the collapsibility result no longer appears to hold and corollaries 3.6 and 3.9 are no longer valid. However, the assumption does lead to some simplifications in the computation of the required quantities of interest. From Theorem 3.11, the following corollary is immediate.

Corollary 3.12 *If $\pi(\mu)=1$, then*

$$\pi_{11}(\sigma_\alpha^2,\sigma_\beta^2,\sigma^2|\mathbf{y}) \propto f_1(\sigma_\alpha^2,\sigma^2)f_2(\sigma_\beta^2,\sigma^2)\sigma^{-IN_K+I+J-1}\exp(\frac{-S_3}{2\sigma^2})\pi^*(\sigma^2), \qquad (3.37)$$

where

$$f_1(\sigma_\alpha^2,\sigma^2) = (\sigma^2/N_K+\sigma_\alpha^2)^{-(I-1)/2}\exp\{\frac{-S_1}{2(\sigma^2/N_K+\sigma_\alpha^2)}\}\pi_1^*(\sigma_\alpha^2),$$

$$f_2(\sigma_\beta^2,\sigma^2) = \sigma_b[\prod_{j=1}^{J}(\sigma^2/K_j+I\sigma_\beta^2)]^{-1/2}\exp\{\frac{-I}{2}\sum_{j=1}^{J}\frac{(y_{.j.}-\tilde{y}_{...})^2}{\sigma^2/K_j+I\sigma_\beta^2}\}\pi_2^*(\sigma_\beta^2).$$

If the primary interest is to make inference about β_j and the noninformative prior (2.4) for σ_α^2 is used, it is possible to integrate out σ_α^2 analytically from the above density (after changing variables to σ_α^2 and $\eta=\sigma^2/\sigma_\alpha^2$ in f_1), thus eliminating one level of integration in subsequent calculations. This reduction of dimension of integration is also possible if the noninformative prior $\pi(\alpha_i)\equiv 1$ is used.

4 METHOD OF COMPUTATION

The method we recommend to evaluate the required integrals in the text is Monte Carlo simulation with importance sampling, based on the hierarchical representation for the posterior distribution. As observed earlier, calculation of the ranking probabilities usually involves three- or four-dimensional integration (for the known σ^2 case; add one dimension if σ^2 is unknown), which is a large enough dimension to make the

Monte Carlo approach appealing. Also, of course, the Monte Carlo approach leads immediately to estimates of the accuracy of the evaluation. The reason for generating random deviates in accord with the hierarchical representation is that it simplifies the task of choosing a good importance function. At each stage, only a one-dimensional importance function need be found.

The suggested hierarchical Monte Carlo approach is to first generate σ^2; then, conditional on σ^2, generate σ_α^2 and σ_β^2; and conditional on them, depending on whether we are finding p_{i*} or p_{*j}, generate Z or β followed by ϕ_j (cf. Lemma 3.4). Note that the conditional distributions for Z, β and ϕ_j are all normal, and so these random variables can be generated directly. If $\pi(\mu) = 1$, it may be possible to integrate out σ_α^2 analytically, as discussed at the end of Section 3. In any case, both σ_α^2 and σ_β^2 can be efficiently generated using two-point mixtures of an inverse gamma and an exponential distribution, each mixture chosen so that the ratio of the height at 0 to that at a non-zero point (the mode if it exists) will agree with the corresponding ratio for the posterior distribution (π_{10} or π_{01}), and so that the right tail of the importance function is similar to and no sharper than that of the posterior. Finally, an inverse gamma distribution ($\mathcal{IG}((IN_K - I - J - 1)/2, 2/S_3)$, cf. Theorem 3.11) is used to generate σ^2 provided that $\pi^*(\sigma^2)$ is relatively flat (like the noninformative priors 1 and σ^{-2}).

5 AN ILLUSTRATIVE EXAMPLE

Let us consider a randomized complete block design for comparing several treatments. Suppose the treatments take differing amounts of time to implement, and that time-consuming treatments are done less often, yielding an unbalanced design. In particular, suppose there are four treatments and three blocks with the data given in Table 1. (This data was chosen artificially, so as to provide a clear illustration of the possible "order reversal" effect.) It is desired, based on the given data, to choose the treatment with the largest treatment effect. If we use α_i, $i = 1,2,3$, and β_j, $j = 1,2,3,4$, to represent the block and the treatment effects, respectively, then the problem is to select the largest β_j. The subsequent analysis is performed conditional on $H_0^1 : \alpha_1 = \alpha_2 = \alpha_3$ and $H_0^2 : \beta_1 = \beta_2 = \beta_3 = \beta_4$ being false.

We assume that the exchangeable priors discussed in the paper are deemed to be appropriate, and use the noninformative priors (2.4) and (2.5) for σ_α^2 and σ_β^2, respectively, and the constant improper prior 1 for both μ and σ^2. The posterior probabilities of each β_j being the largest, as well as estimates of β_j and their corresponding standard errors (the square roots of the posterior variances of the β_j) are given in Table 2. Note that treatment A has the highest probability of being the best, in spite of the fact that treatment D has the largest sample mean. This is the "order reversal" effect that can arise when variances are unequal. Contrasts with A, the largest estimated treatment, are presented in Table 3.

For comparison, use of a noninformative prior, $\pi(\alpha_i) \equiv 1$, for the block effects was also considered. The results were very similar to the results from the two-way analysis given above (and hence are omitted), indicating that the assumption of exchangeability of the α_i has little effect here (in contrast to the assumption of exchangeability of the β_j, which resulted in an order reversal).

Table 1. Observations on 4 Treatments in 3 Blocks

Block	Treatment									
	A				B		C		D
1	550	500	550	450	550	500	450	400	400
2	400	450	500	450	300	400	350	450	500
3	600	500	650	550	350	350	550	450	650
Average	512.50				408.33		441.67		516.67

Finally, if it is desired to test the various hypotheses about the block and the treatment effects, proper priors must be used for both σ_α^2 and σ_β^2. As an example, assume that priors of the form (2.7) are chosen. Suppose a user has a prior belief

Table 2. Ranking Probability, Mean and Standard Error for the Treatment Effect

j	A (j=1)	B (j=2)	C (j=3)	D (j=4)
Ranking Probability	.551±.007	.010±.001	.035±.003	.411±.012
Mean of β_j	34.29±0.20	-42.32±0.38	-19.18±0.17	27.21± 0.35
Standard Error	50.10±2.01	52.59±1.93	51.13±1.99	54.36±1.91

Table 3. Mean and Standard Error for the Treatment Contrast $\beta_1 - \beta_j$

	j=2	j=3	j=4
Mean	76.61±0.55	53.47±0.35	7.08±0.16
Standard Error	40.09±0.27	36.47±0.27	40.40±0.41

that, with probability 0.8, σ_α^2 is less than 10.0 and, independently, with probability 0.8, σ_β^2 is less than 20.0; the user would then use values of r and d in (2.7) for $\pi_1^*(\sigma_\alpha^2)$ and $\pi_2^*(\sigma_\beta^2)$ of (0.8, 10.0) and (0.8, 20.0), respectively. Also, suppose the user chooses $\pi_1 = \pi_2 = 1/2$ and the constant improper prior 1 for both μ and σ^2 as a noninformative choice. The posterior probabilities of the various combinations of hypotheses being true are given in Table 4. Note that the joint probability of both effects being present is highest. If one considers the marginal probabilities (simply sum the rows or columns), these is also moderately strong evidence that each effect exists individually.

Table 4. Joint Probabilities for Various Hypotheses in the Additive Model

	H_0^1	H_1^1
H_0^2	0.12	0.18
H_1^2	0.24	0.46

To perform the integrations in this example, 3000 random vectors of deviates were generated to calculate the probabilities, means and standard errors. This yielded sufficient accuracy for all practical purposes (numbers after the plus-minus sign represent the simulation errors). The program was run on an IBM 3090-400 machine using 3.4

seconds of total processor time for the evaluation of Tables 2 and 3. To compute Table 4, we found that it is convenient to use existing programs in IMSL (the numerical integration subroutines like DMLIN, DBLIN and DCADRE) to evaluate the normalizing constants, $K_{111}, K_{101}, K_{011}$ and K_{001}. The total time required was 0.7 seconds.

References

Berger, J. O. (1985), Statistical Decision Theory and Bayesian Analysis, New York: Springer-Verlag, 2nd edition.

Berger, J. O., and Deely, J. J. (1988), "A Bayesian Approach to Ranking and Selection of Related Means with Alternatives to AOV Methodology," Journal of the American Statistical Association, 83, 364-373.

Berger, J. O., and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of P-Values and Evidence (with Discussion)," Journal of the American Statistical Association, 82, 112-139.

Box, G., and Tiao, G. (1968), "Bayesian Estimation of Means for the Random Effects Model," Journal of the American Statistical Association, 63, 174-181.

Dawid, A. P. (1977), "Invariant Distribution and Analysis of Variance Models," Biometrika, 64, 291-297.

Deely, J. J., and Zimmer, W. J. (1987), "Choosing a Quality Supplier - a Bayesian Approach," submitted to Bayesian Statistics 3.

Ghosh, M., and Meeden, G. (1984), "A New Bayesian Analysis of a Random

Effects Model," Journal of the Royal Statistical Society, Series B, 46, 474-482.

Goel, P., and Rubin, H. (1977), "On Selecting a Subset Containing the Best Population - a Bayesian Approach," The Annals of Statistics, 5, 969-983.

Gupta, S. S., and Hsu, J. C. (1980), "Subset Selection Procedures with Application to Motor Vehicle Fatality Data in a Two-Way Layout," Technometrics, 22, 543-546.

Gupta, S. S., and Yang, H. M. (1985), "Bayes-p^* Subset Selection Procedures for the Best Population," Journal of Statistical Planning and Inference, 12, 213-233.

Hill, B. (1965), "Inference About Variance Components in the One-Way Model," Journal of the American Statistical Association, 60, 806-825.

Huang, D. Y., and Panchapakesan, S. (1976), "A Modified Subset Selection Formulation with Special Reference to One-Way and Two-Way Layout Experiments," Communications in Statistics: Theory and Methods, A5, 621-633.

Kloek, T., and Van Dijk, H. K. (1978), "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," Econometrica, 46, 1-20.

Lindley, D. V. (1974), "A Bayesian Solution for Two-Way Analysis of Variance," in Progress in Statistics (European Meeting Statisticians, Budapest, 1972), North-Holland, Amsterdam, 475-496.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," Journal of the Royal Statistical Society, Series B, 34, 1-41.

Smith, A. F. M. (1973), "Bayes Estimates in One-Way and Two-Way Models," Biometrika, 60, 319-329.

Appendix

Proof of Theorem 3.1

Let $y_{ij.} = \sum_{k=1}^{K_j} y_{ijk}/K_j$ and $\bar{\alpha} = \sum_{i=1}^{I} \alpha_i/I$. Since

$$\sum_{i,j,k}(y_{ijk} - \mu - \alpha_i - \beta_j)^2 = \sum_{i,j,k}(y_{.j.} - \mu - \bar{\alpha} - \beta_j)^2 + \sum_{i,j,k}(\bar{y}_{i..} - \bar{y}_{...} - \alpha_i + \bar{\alpha})^2$$
$$+ \sum_{i,j,k}(y_{ij.} - \bar{y}_{i..} - y_{.j.} + \bar{y}_{...})^2 + \sum_{i,j,k}(y_{ijk} - y_{ij.})^2,$$

it is convenient to work with the following sets of quantities to obtain likelihood functions based on $\mathbf{y}$:

$$\{y_{.j.}\}, \quad \{\bar{y}_{i..} - \bar{y}_{...}\}, \quad \{y_{ij.} - \bar{y}_{i..} - y_{.j.} + \bar{y}_{...}\}, \quad \{y_{ijk} - y_{ij.}\}.$$

Observing that

$$y_{.j.} = \mu + \bar{\alpha} + \beta_j + \epsilon_{.j.}; \qquad \bar{y}_{i..} - \bar{y}_{...} = \alpha_i - \bar{\alpha} + \bar{\epsilon}_{i..} - \bar{\epsilon}_{...};$$

$$y_{ij.} - \bar{y}_{i..} - y_{.j.} + \bar{y}_{...} = \epsilon_{ij.} - \bar{\epsilon}_{i..} - \epsilon_{.j.} + \bar{\epsilon}_{...}; \qquad y_{ijk} - y_{ij.} = \epsilon_{ijk} - \epsilon_{ij.},$$

where $\epsilon_{ij.} = \sum_{k=1}^{K_j} \epsilon_{ijk}/K_j$, $\epsilon_{.j.} = \sum_{i=1}^{I} \epsilon_{ij.}/I$, $\bar{\epsilon}_{i..} = \sum_{j=1}^{J}\sum_{k=1}^{K_j} \epsilon_{ijk}/N_K$, and $\bar{\epsilon}_{...} = \sum_{i=1}^{I} \epsilon_{i..}/I$, it is easy to check that, conditional on σ_α^2 and σ_β^2, the four sets of variables are independent of one another and

$$\sum_{i=1}^{I}(\bar{y}_{i..} - \bar{y}_{...})^2 \sim (\sigma^2/N_K + \sigma_\alpha^2)\chi_{I-1}^2,$$

$$\sum_{i,j,k}(y_{ij.} - \bar{y}_{i..} - y_{.j.} + \bar{y}_{...})^2 \sim \sigma^2\chi^2_{(I-1)(J-1)},$$

$$\sum_{i,j,k}(y_{ijk} - y_{ij.})^2 \sim \sigma^2\chi^2_{IN_K-IJ}.$$

Thus, to obtain the likelihood function $l(\sigma^2_\alpha, \sigma^2_\beta|\mathbf{y})$, it suffices to consider the likelihood functions corresponding to $\mathbf{y}_2 = (y_{.1.}, \ldots, y_{.J.})^T$ and $S_1 = \sum_{i=1}^I(\bar{y}_{i..} - \bar{y}_{...})^2$ alone, and multiply them together.

From Berger (1985, section 4.6), one has

$$\begin{aligned} l(\sigma^2_\alpha, \sigma^2_\beta|\mathbf{y}_2) \;\propto\; & \sigma_\mu[\prod_{j=1}^J(\sigma^2/K_j + I\sigma^2_\beta)]^{-1/2}[1 + (\sigma^2_\alpha + I\sigma^2_\mu)/\sigma^2_b]^{-1/2} \\ & \times \exp\{\frac{-I}{2}[\sum_{j=1}^J \frac{(y_{.j.} - \tilde{y}_{...})^2}{\sigma^2/K_j + I\sigma^2_\beta} + \frac{(\tilde{y}_{...} - w)^2}{\sigma^2_b + \sigma^2_\alpha + I\sigma^2_\mu}]\}. \end{aligned}$$

Since

$$l(\sigma^2_\alpha|S_1) \propto (\sigma^2/N_K + \sigma^2_\alpha)^{\frac{-(I-1)}{2}} \exp\{\frac{-S_1}{2(\sigma^2/N_K + \sigma^2_\alpha)}\},$$

the result follows immediately. (The first multiplicative factor, σ_μ, is an irrelevant constant included here for consistency with the expression obtained in the case of vague knowledge about μ.)

Proof of Lemma 3.3

In the Proof of Theorem 3.1, sending σ^2_μ to infinity (which can be shown to be equivalent to choosing $\pi(\mu) = 1$), the likelihood $l(\sigma^2_\alpha|S_1)$ remains unchanged but

$$l(\sigma^2_\alpha, \sigma^2_\beta|\mathbf{y}_2) \propto \sigma_b[\prod_{j=1}^J(\sigma^2/K_j + I\sigma^2_\beta)]^{-1/2} \exp\{\frac{-I}{2}\sum_{j=1}^J \frac{(y_{.j.} - \tilde{y}_{...})^2}{\sigma^2/K_j + I\sigma^2_\beta}\}. \qquad (*)$$

Using the decomposition result for the likelihood function in the Proof of Theorem 3.1, the first part of the lemma follows.

Treating σ_α^2 as known and letting $\sigma_\alpha^2 \to \infty$ results, after renormalization, in $l(\sigma_\beta^2|\mathbf{y}_2)$ being given by $(*)$. Hence π_{01} is the required posterior density when the option $\pi(\alpha_i) \equiv 1$ is used. Similarly, as $\sigma_\beta^2 \to \infty$, it can be shown that the likelihood function $l(\sigma_\alpha^2|\mathbf{y})$ is proportional to $l(\sigma_\alpha^2|S_1)$. Thus π_{10} is the posterior density of σ_α^2.

Proof of Lemma 3.4

Let $\mathbf{y}_1^* = (\bar{y}_{1..}, \ldots, \bar{y}_{I..})^T - \bar{y}_{...}1_I$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$ and 1_n be a vector of n ones. From the Proof of Theorem 3.1, we have the following results:

<u>Result 1</u>. Conditional on $(\mu, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)$, $\mathbf{y}_1^*$ and $\mathbf{y}_2$ are sufficient for $\boldsymbol{\alpha}^* = \boldsymbol{\alpha} - \bar{\alpha}1_I$ and $\boldsymbol{\beta} + (\mu + \bar{\alpha})1_J$, respectively.

<u>Result 2</u>. Conditional on $(\mu, \boldsymbol{\alpha}^T, \sigma_\beta^2)$, $\tilde{y}_{...}$ is sufficient for $\mu + \bar{\alpha}$.

Proof. This follows from the decomposition result in the Proof of Theorem 3.1, the fact that

$$\sum_{j=1}^{J} \frac{(y_{.j.} - \mu - \bar{\alpha})^2}{\sigma_\beta^2 + \sigma^2/(IK_j)} = \sum_{j=1}^{J} \frac{(y_{.j.} - \tilde{y}_{...})^2}{\sigma_\beta^2 + \sigma^2/(IK_j)} + \sum_{j=1}^{J} \frac{(\tilde{y}_{...} - \mu - \bar{\alpha})^2}{\sigma_\beta^2 + \sigma^2/(IK_j)},$$

and an application of the factorization theorem.

<u>Result 3</u>. Conditional on $(\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2)$, $\bar{\alpha}$ is independent of $\alpha_i - \bar{\alpha}$ for all i.

Proof. From the Proof of Theorem 3.1 it is clear that, conditional on $(\boldsymbol{\alpha}^T, \sigma_\beta^2)$, $\mathbf{y}_1^*$ and

$\tilde{y}_{...}$ are independent and they are sufficient for $\boldsymbol{\alpha}^*$ and $\bar{\alpha}$, respectively. Also, given σ_α^2, $\bar{\alpha}$ is independent of $\alpha_i - \bar{\alpha}$ for all i. The result follows immediately.

From Berger and Deely (1988) and Result 1 above we have

$$\begin{aligned} Pr(\alpha_i \text{ largest}|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) &= Pr(\alpha_i - \bar{\alpha} \text{ largest}|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) \\ &= Pr(\alpha_i - \bar{\alpha} \text{ largest}|\mathbf{y}_1^*, \sigma_\alpha^2) \\ &= \eta_i(\sigma_\alpha^2) \end{aligned}$$

and

$$\begin{aligned} Pr(\beta_j \text{ largest}|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) &= Pr(\mu + \bar{\alpha} + \beta_j \text{ largest}|\mathbf{y}, \sigma_\alpha^2, \sigma_\beta^2) \\ &= Pr(\mu + \bar{\alpha} + \beta_j \text{ largest}|\mathbf{y}_2, \sigma_\alpha^2, \sigma_\beta^2) \\ &= \psi_j(\sigma_\alpha^2, \sigma_\beta^2). \end{aligned}$$

The required result follows easily.

Proof of Lemma 3.7

Since $\mathbf{y}_2|\mu, \bar{\alpha}, \boldsymbol{\beta} \sim N(\boldsymbol{\beta}+(\mu+\bar{\alpha})1_J, \Sigma)$, where Σ is a diagonal matrix with diagonal elements $\sigma^2/(IK_j)$, it follows from Berger (1985) that

$$\boldsymbol{\beta} + (\mu + \bar{\alpha})1_J|\mathbf{y}_2, \mu + \bar{\alpha}, \sigma_\beta^2 \sim N(\mathbf{y}_2 - \Sigma W_2(\mathbf{y}_2 - (\mu + \bar{\alpha})1_J), \Sigma - \Sigma W_2 \Sigma),$$

$$\mu + \bar{\alpha}|\mathbf{y}_2, \sigma_\alpha^2, \sigma_\beta^2 \sim N(\tilde{y}_{...} - [1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2]^{-1}(\tilde{y}_{...} - w), \frac{(\sigma_\alpha^2 + I\sigma_\mu^2)/I}{1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2}),$$

where $W_2 = (\Sigma + \sigma_\beta^2 I_{J\times J})^{-1}$ and $I_{J\times J}$ is the $J \times J$ identity matrix. Hence, conditional on $(\mathbf{y}_2, \sigma_\alpha^2, \sigma_\beta^2)$, the mean vector and the covariance matrix of $\boldsymbol{\beta}$ are respectively given by

$$
\begin{aligned}
E(\boldsymbol{\beta}|\mathbf{y}_2, \sigma_\alpha^2, \sigma_\beta^2) &= (I_{J\times J} - \Sigma W_2)(\mathbf{y}_2 - [\tilde{y}_{...} - \frac{\tilde{y}_{...} - w}{1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2}]1_J), \\
V(\boldsymbol{\beta}|\mathbf{y}_2, \sigma_\alpha^2, \sigma_\beta^2) &= \Sigma - \Sigma W_2 \Sigma + (I_{J\times J} - \Sigma W_2)\mathbf{1}_{J\times J}(I_{J\times J} - W_2\Sigma)\frac{(\sigma_\alpha^2 + I\sigma_\mu^2)/I}{1 + (\sigma_\alpha^2 + I\sigma_\mu^2)/\sigma_b^2},
\end{aligned}
$$

where $\mathbf{1}_{J\times J}$ is the matrix of all ones.

From the densities of $\mathbf{y}_1^*$ given $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}^*$ given σ_α^2, one has

$$
\boldsymbol{\alpha}^*|\mathbf{y}_1^*, \sigma_\alpha^2 \sim N(\frac{\sigma_\alpha^2 \mathbf{y}_1^*}{\sigma^2/N_K + \sigma_\alpha^2}, \frac{\sigma_\alpha^2 \sigma^2}{\sigma^2 + N_K \sigma_\alpha^2}[I_{I\times I} - (1/I)1_I 1_I^T]).
$$

Using Result 2 in the Proof of Lemma 3.4, the posterior density of $\bar{\alpha}$ can be obtained:

$$
\bar{\alpha}|\tilde{y}_{...}, \sigma_\alpha^2, \sigma_\beta^2 \sim N(\frac{\sigma_\alpha^2(\tilde{y}_{...} - w)}{\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2}, \frac{\sigma_\alpha^2(\sigma_b^2 + I\sigma_\mu^2)}{I(\sigma_b^2 + \sigma_\alpha^2 + I\sigma_\mu^2)}).
$$

Thus the posterior density of $\boldsymbol{\alpha}$ follows from Result 3 in the Proof of Lemma 3.4.

The lemma follows from Result 1 in the Proof of Lemma 3.4.

Proof of Theorem 3.10

The first part of the theorem follows from corollaries 3.6 and 3.9. To prove the second part, we consider the following two cases:

(i) $\pi(\alpha_i) \equiv 1$

It follows from the Proof of Lemma 3.4 that, conditional on $\mathbf{y}$, $\boldsymbol{\alpha}^*$ has the same distribution as

$$\boldsymbol{\alpha}^* | \mathbf{y}_1^* \sim N(\mathbf{y}_1^*, [I_{I\times I} - (1/I)1_I 1_I^T]\sigma^2/N_K).$$

Using the fact that $\alpha_i - \alpha_r = (\alpha_i - \bar{\alpha}) - (\alpha_r - \bar{\alpha})$ yields the required result.

(ii) $\pi(\beta_j) \equiv 1$

From the Proof of Theorem 3.1 it can be shown that, conditional on $(\boldsymbol{\beta}^T, \sigma_\alpha^2)$, $\mathbf{y}_2$ is sufficient for $\boldsymbol{\beta}$. Since

$$\mathbf{y}_2 | \boldsymbol{\beta}, \sigma_\alpha^2 \sim N(\boldsymbol{\beta} + w1_J, \Sigma + (\sigma_\mu^2 + \sigma_\alpha^2/I)1_J 1_J^T),$$

it follows that

$$\boldsymbol{\beta} | \mathbf{y}, \sigma_\alpha^2 \sim N(\mathbf{y}_2 - w1_J, \Sigma + (\sigma_\mu^2 + \sigma_\alpha^2/I)1_J 1_J^T).$$

Note that the derived distribution

$$\beta_j - \beta_s | \mathbf{y} \sim N(y_{.j.} - y_{.s.}, (\sigma^2/K_j + \sigma^2/K_s)/I)$$

does not depend on σ_α^2.