P. R. Krishnaiah, ed., *Handbook of Statistics, Vol. 1*
©North-Holland Publishing Company (1980) 117–132

4

Bayesian Inference in MANOVA

S. James Press

1. Introduction

This article focuses attention on the various models of the analysis of variance from the Bayesian inference point of view. In particular, we consider the multivariate analysis of variance (MANOVA).

Let y_j denote an $N\times 1$ vector of observations on the j^{th} dependent variable (yield of the j^{th} crop for N farms, score on the j^{th} type of examination for N subjects, etc.). Let $X: N\times q$ denote the (design) matrix of observations of q independent variables (including a constant term) that are related to the j^{th} independent variable (q characteristics of each of the N farms, q characteristics of each of the subjects who took the j^{th} type of examination, etc.). Then, assuming linearity of the relationships,

$$\begin{aligned}
\underset{(N\times 1)}{y_1|X} &= \underset{(N\times q)}{X}\ \underset{(q\times 1)}{\beta_1} + \underset{(N\times 1)}{u_1}\\
y_2|X &= X\ \beta_2 + u_2\\
&\ \ \vdots\\
y_p|X &= X\ \beta_p + y_p,
\end{aligned}$$

represents a system of N observations on p dependent variables and q independent variables; $\beta_j: q\times 1$ denotes the j^{th} unknown coefficient vector, and $u_j: N\times 1$ denotes the vector of errors corresponding to the j^{th} dependent variable, for each of the N cases. The error vectors may of course include omitted variables, misspecification structure of the model (nonlinearities, etc.), and errors in observing the dependent variable. We write the above system of equations in the more convenient compact form

$$\underset{(N\times p)}{Y|X} = \underset{(N\times q)}{X}\ \underset{(q\times p)}{B} + \underset{(N\times p)}{U}, \tag{1}$$

where

$$Y=(y_1,\ldots,y_p), \qquad U=(u_1,\ldots,u_p), \qquad B=(\beta_1,\ldots,\beta_p).$$

Also let $U'\equiv(v_1,\ldots,v_N)$, so that v_i' denotes the i^{th} row of U, and the prime denotes transpose. The Bayesian model of MANOVA will be developed in Section 4 as a special case of the Bayesian multivariate regression model. We assume the fixed effects model. We first develop the Bayesian multivariate regression model.

2. Assumptions of the model

To effect inferences about the unknown parameters of the model it is necessary to make some assumptions. Violations of these assumptions will lead to new models and modifications that must be made to the basic model.

1. $p+q\leqslant N$; this assumption permits all parameters to be estimated.

2. rank of $X_{(N\times q)}=q$; if rank of $X<q$ estimators of B will not be unique.

3. $E(v_i)=0$; $\operatorname{var}(v_i)\equiv E(v_iv_i')=\Sigma$, for all i. Σ is assumed to be an arbitrary, positive definite, symmetric matrix of order p. Thus, we are guaranteed that v_i has a density in p-dimensional space. The fact that Σ does not depend upon i means that the error vectors all have the same covariance matrix. This assumption is the multivariate version of homoscedasticity; it helps to provide estimators of the elements of B that have minimum variance.

4. The v_i vectors are mutually independent. Then, $E(v_iv_j')=0$ for $i\neq j$; that is, the v_i vectors are mutually uncorrelated. Assumptions 3 and 4 together provide the basis for minimum variance estimators. The implication of the assumption is that the N farm yields (or test scores of different subjects) are independent.

5. $\mathcal{L}(v_i)=N(0,\Sigma)$, for all $i=1,\ldots,N$; that is, the probability law of the v_i vectors is normal with mean zero and covariance matrix Σ. Thus, yields of the p crops or scores on the p types of examinations are correlated.

6. X is fixed and known; this implies that we need not make any assumptions about the distribution of X.

3. Estimation

The unknown parameters in the model are (B,Σ). They will be estimated by combining subjective (prior) information with the observed data, via Bayes theorem.

It is well known (and straightforward to show) that the maximum likelihood estimators of B, and Σ are given by

$$\underset{(q\times p)}{\hat{B}} = (\hat{\beta}_1,\ldots,\hat{\beta}_p) = (X'X)^{-1}X'Y,$$

$$\underset{(p\times p)}{\hat{\Sigma}} = \frac{1}{N}(Y-X\hat{B})'(Y-X\hat{B}).$$

Thus, if the matrix of residuals (estimated errors) is given by

$$\hat{U} = Y - X\hat{B}, \quad \text{and} \quad V \equiv \hat{U}'\hat{U},$$

an unbiased estimator of Σ is given by

$$\tilde{\Sigma} = \frac{1}{N-q}\hat{U}'\hat{U}.$$

It will be useful to rely upon an *orthogonality property of MLE's*, namely

$$(Y-XB)'(Y-XB) = V + (B-\hat{B})'(X'X)(B-\hat{B}). \tag{2}$$

This property is readily demonstrated by writing $(Y-XB)$ as $[(Y-X\hat{B}) - X(B-\hat{B})]$, expanding the sum of squares and noting that the cross product terms vanish.

3.1. Likelihood function

Since the v_i vectors are independent, the probability density[1] of the error matrix is

$$p(U|\Sigma) = p(v_1,\ldots,v_N|\Sigma) = \prod_{j=1}^{N} p(v_j|\Sigma).$$

Since

$$p(v_j|\Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-1/2(v_j'\Sigma^{-1}v_j)},$$

$$p(U|\Sigma) \propto |\Sigma|^{-N/2} \exp\left\{\left(-\frac{1}{2}\right)\sum_{j=1}^{N}(v_j'\Sigma^{-1}v_j)\right\},$$

where $\propto$ denotes proportionality. Since $U'U = \Sigma_1^N v_j v_j'$, the likelihood becomes $p(U|\Sigma) \propto |\Sigma|^{-N/2}\exp\{(-\frac{1}{2})\operatorname{tr} U'U\Sigma^{-1}\}$, where $\operatorname{tr}(\cdot)$ denotes the

[1]We will be using $p(\cdot)$ to denote densities in a generic way and will not change letters to distinguish one density from another; that will be accomplished by changing arguments.

trace operation. Now change variables to find the density of Y. Since $U=Y-XB$, the Jacobian of the transformation is unity and the new density becomes

$$p(Y|X,B,\Sigma)\propto|\Sigma|^{-N/2}\exp\left\{\left(-\tfrac{1}{2}\right)\mathrm{tr}(Y-XB)'(Y-XB)\Sigma^{-1}\right\}.$$

Use of eq. (2) gives the likelihood function

$$p(Y|X,B,\Sigma)\propto|\Sigma|^{-N/2}\exp\left\{\left(-\tfrac{1}{2}\right)\mathrm{tr}\left[V+(B-\hat{B})'(X'X)(B-\hat{B})\right]^{-1}\right\}. \tag{3}$$

3.2. *Prior to posterior analysis*

To make posterior inferences (those made after observing the sample data) about (B,Σ), we must first assess a prior distribution (one based only upon information available prior to observing the sample data) for (B,Σ). This prior distribution is very personal, in that it is the distribution representing the subjective beliefs of, and prior information available to, the decision maker.[2] The posterior distribution will of course be relative to the same individual. Thus, the entire MANOVA analysis yields changes in subjective beliefs, predictions, or decisions all based upon the assessed *prior of a given individual*. We will consider two classes of priors, non-informative, and informative.

Case 1. Non-informative prior

For this case we adopt the prior density (see Geisser and Cornfield, 1963)

$$p(B,\Sigma)=p(B)p(\Sigma),$$

where

$$p(B)\propto\text{constant},\qquad p(\Sigma)\propto|\Sigma|^{-(p+1)/2}.$$

That is,

$$p(B,\Sigma)\propto|\Sigma|^{-(p+1)/2}. \tag{4}$$

This assessment assumes B and Σ are independent, a priori; the elements of B are all independent, a priori, and they are all uniformally distributed over the entire real line. Thus, the prior distribution of B is improper. Eq. (4) also implies an improper prior distribution for Σ. In one dimension

[2]The term "decision maker" will be used throughout this article to refer to the individual for whom the MANOVA analysis is being carried out, whether or not he will actually make a decision at this time.

($p=1$), this assumption would imply that $\log\Sigma$ is uniformly distributed over the entire real line; so we are adopting a p-dimensional generalization of that idea. The distribution we have adopted, though improper, will lead to proper posteriors, and meaningful statistical inferences. We propose this prior distribution for many reasons.

A principal reason for using the non-informative prior is that in many respects the distribution corresponds to our intuitive notion of prior ignorance about (B, Σ); that is, the idea of our subjective feelings about (B, Σ), a priori, being very vague, and not well defined; we are inclined to feel that all possible values are equally likely; or we wish to adopt a posture of trying our best to let the data "speak for themselves", objectively, without "contaminating" the inferences with individual, subjective beliefs. The posterior inferences made on the basis of this non-informative prior distribution will often be the same inferences that would be made by the frequentist statistician. For this reason, the non-informative prior provides sort of a benchmark, or point of departure, for Bayesian inferences. That is, inferences based upon the non-informative prior are those dictated by what we would believe in the "absence of prior information", and those based upon the informative prior are those based upon what we would believe when we try to combine our substantive subjective prior beliefs with observation.

This non-informative prior distribution can be shown to correspond to minimal information in a Shannon information sense (Shannon, 1948). It also has the property that it is invariant under parameter transformation groups (Jeffreys, 1961). Thus, probability statements made about observable random variables should remain invariant under changes in the parameterizations of the problem. In proposing fiducial inference R. A. Fisher implied a non-informative distribution on the parameters (for $p=1$), as did Fraser (1968) in his development of structural inference. A summary of these developments is given in Press (1972, Chapter 3).

Bayes theorem implies that the posterior density is proportional to the product of the prior density and the likelihood function. In the multivariate regression problem, we multiply eqs. (3) and (4) to yield the joint posterior density

$$p(B, \Sigma | X, Y) \propto |\Sigma|^{-(N+p+1)/2} \times \exp\left\{\left(-\tfrac{1}{2}\right) \operatorname{tr} \Sigma^{-1}\left[V + (B-\hat{B})'(X'X)(B-\hat{B})\right]\right\}. \tag{5}$$

Inferences about B, without regard to Σ, may be made from the marginal posterior density of B. The marginal posterior density of B is given by

$$p(B|X,Y)=\int_{\Sigma>0} p(B,\Sigma|X,Y)\,\mathrm{d}\Sigma. \tag{6}$$

We will actually want to make inferences about $\theta_{(p\times q)}\equiv B'$. Let $\hat{\theta}\equiv\hat{B}'$. The integration in (6) is readily carried out using the properties of the Wishart distribution (see, e.g. Press, 1972, p. 229). The result is

$$p(\theta|X,Y)\propto|V+(\theta-\hat{\theta})(X'X)(\theta-\hat{\theta})'|^{-N/2}.$$

The expression on the right is readily recognized to be the kernel of a matrix T-distribution (Kshirsagar, 1960). The complete density is given, for $-\infty<\theta<+\infty$, by

$$p(\theta|X,Y)=k|V|^{(N-q)/2}|X'X|^{p/2}|V+(\theta-\hat{\theta})'(X'X)(\theta-\hat{\theta})|^{-N/2}, \tag{7}$$

where $N>p+q-1$, $V>0, X'X>0$, and

$$k\equiv\frac{\Gamma_q(N/2)}{\pi^{pq/2}\Gamma_q\left(\frac{N-p}{2}\right)}.$$

This result is identical with that found by Geisser (1965, eq. (4.8)).

The notation $V>0$, for any matrix V, means that V is positive definite symmetric. The notation $\Gamma_q(t)$ denotes the q-dimensional gamma function, defined as

$$\Gamma_p(t)=\int_{X>0}|X|^{t-(p+1)/2}\mathrm{e}^{-\mathrm{tr}(X)}\,\mathrm{d}X$$
$$=\pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma\left(t-\frac{j-1}{2}\right).$$

The result in eq. (7) implies, in part, that the rows and columns of θ, a posteriori, follow multivariate Student t-distributions, and the individual elements of θ follow univariate Student t-distributions. It also follows that $E(\theta|X,Y)=\hat{\theta}$, and if $\theta_{(p\times q)}\equiv\left((\theta_1)_{(p\times 1)},\dots,(\theta_q)_{(p\times 1)}\right)$, and $\theta'_{(1\times pq)}\equiv(\theta'_1,\dots,\theta'_q)$, $\mathrm{var}(\theta|X,Y)=(1/(N-p-q-1))V\otimes(X'X)^{-1}$. Other properties of the distribution in (7) have been given by Dickey (1967) and Geisser (1965).

From an operational viewpoint, it is useful for computing confidence regions on θ to note that (see Geisser, 1965)

$$U \equiv \frac{|V|}{|V+(\theta-\hat{\theta})'(X'X)(\theta-\hat{\theta})|}$$

has an $U_{p,q,N-q}$ distribution, as defined by Anderson (1958); i.e., it is distributed as the product of independent beta variates. Thus, the posterior distribution of U (where θ is the random variable) is the same as the sampling distribution of U (for fixed θ). So a posterior region for θ is found from the relation

$$P\{U(\theta) \leqslant U_{\alpha,p,q,N-q}\} = 1-\alpha,$$

where $U_{\alpha,p,q,N-q}$ is the α^{th} percentage point. Therefore the Bayesian region on θ is equivalent to the confidence region.

Inferences about Σ, without regard to B, may be made from the marginal posterior distribution of Σ. The posterior density of Σ is found by integrating eq. (5) with respect to B, and is given by

$$p(\Sigma|X,Y) \propto |\Sigma|^{-(N+p-q+1)/2} \exp\{(-\tfrac{1}{2}) \operatorname{tr} \Sigma^{-1} V\}.$$

The expression on the right is readily recognized as the kernel of an inverted Wishart distribution (see e.g. Press, 1972, p. 109). The complete density is given, for $V>0$, by

$$p(\Sigma|X,Y) = \frac{|V|^{(N-q)/2}}{c|\Sigma|^{(N+p-q+1)/2}} \exp\{(-\tfrac{1}{2}) \operatorname{tr} \Sigma^{-1} V\}, \tag{8}$$

where

$$c \equiv 2^{(N-q)(p)/2} \Gamma_p\left(\frac{N-q}{2}\right).$$

It follows from (8) that

$$E(\Sigma|X,Y) = \frac{V}{N-p-q-1},$$

where $N-p-q-1>0$. Variances and covariances of $(\Sigma|X,Y)$ may be found, e.g., in Press (1972, p. 112). Posterior inferences regarding the diagonal elements of Σ (or blocks of diagonal elements) may be made from the marginal densities of the distribution in (8). The marginals of the diagonal elements of Σ follow inverted gamma distributions while the marginals of the block diagonal elements of Σ also follow inverted Wishart distributions (see, e.g., Press, 1972, p. 111).

Case 2. Informative prior

We now treat the case in which the analyst has some specific subjective prior information he would like to interpose in this problem. The mechanism we propose for introducing this information involves the so-called (generalized) natural conjugate (this class was introduced by Raiffa and Schlaifer (1961)) family of distributions. The approach we suggest is to represent the prior distribution of (B, Σ) by a parametric family of distributions whose members are indexed by certain fixed, but as yet undetermined, parameters (often called hyperparameters to distinguish them from the parameters that index the sampling distribution). The hyperparameters are then assessed for the decision maker on the basis of his specific prior information.

For example, the decision maker might not know the value of a regression coefficient θ_{ij}, but he might feel θ_{ij} is most likely equal to about θ_{ij}^* although it could be greater or less than θ_{ij}^* with probabilities that get steadily smaller as we depart from θ_{ij}^* in either direction, symmetrically. That is, θ_{ij} is assumed to follow some unimodal, symmetric distribution (such as normal) centered at θ_{ij}^*. The decision maker could be "pressed" further and he might conjecture that in his view it is unlikely that the value of θ_{ij} would lie outside some stated range. These assertions could be used to assess some of the hyperparameters by taking the roughly stated coefficient value to be the mean of the corresponding prior distribution; the stated range could be used as the value of three standard deviations of the corresponding prior distribution. Extending these ideas to many parameters will yield a complete assessment of the hyperparameters.

The assessment problem is not simple and must be carried out carefully. It involves forcing the decision maker to introspect about the problem and to draw upon both his past experiences, and any theory he believes about the phenomenon at issue. There are now a number of computer programs available for assisting the analyst to assess prior information from the decision maker (see Press, 1979). Such problems greatly facilitate the problem of assessment in a multiparameter problem such as MANOVA.

The regression coefficients will be assumed to be jointly normal, a priori, while the covariances will be assumed, a priori, to follow an inverted Wishart distribution. The regression coefficients are first expressed as a long concatenated vector. Recall that $B_{(q\times p)} \equiv \left((\beta_1)_{(q\times 1)}, \dots, (\beta_p)_{(q\times 1)}\right)$. Now define $\beta'_{(pq\times 1)} \equiv (\beta'_1, \dots, \beta'_p)$; similarly for $\hat{\beta}$. Next note the convenient identity relationship:

$$(\beta-\hat{\beta})'[\Sigma^{-1}\otimes(X'X)](\beta-\hat{\beta})=\operatorname{tr}\Sigma^{-1}[(B-\hat{B})'(X'X)(B-\hat{B})]. \tag{9}$$

This identity is established readily by recalling the definition of direct product and examining the general element of both sides of the identity. Now assume that a priori, β and Σ are independent and follow densities

$$p(\beta|\phi,F)\propto\exp\{(-\tfrac{1}{2})(\beta-\phi)'F^{-1}(\beta-\phi)\},$$
$$p(\Sigma|G,m)\propto|\Sigma|^{-m/2}\exp\{(-\tfrac{1}{2})\operatorname{tr}[G\Sigma^{-1}]\},$$

for $G>0, m>2p$, so that the joint prior density is given by

$$p(B,\Sigma|\phi,F,G,m)\propto|\Sigma|^{-m/2}\exp\{(-\tfrac{1}{2})\operatorname{tr}[G\Sigma^{-1}+(\beta-\phi)'F^{-1}(\beta-\phi)]\}. \tag{10}$$

Note that (ϕ,F,G,m) are hyperparameters of the prior distribution that must be assessed. We now go forward in the analysis assuming (ϕ,F,G,m) are known for a given decision maker.

Bayes theorem yields the joint posterior distribution of (B,Σ), for the case of the non-informative prior, by multiplying the likelihood function in (3) by the prior in (10). The result is

$$\begin{aligned}(B,\Sigma|X,Y,\phi,F,G,m)\propto|\Sigma|^{-(m+N)/2}\exp\{(-(\tfrac{1}{2})[(\beta-\phi)'F^{-1}(\beta-\phi)\\ +\operatorname{tr}\Sigma^{-1}[(V+G)+(B-\hat{B})'\\ \times(X'X)(B-\hat{B})]]\}.\end{aligned} \tag{11}$$

Integrating eq. (11) with respect to Σ, in order to obtain the marginal posterior density of B, gives, for all $-\infty<B<+\infty$,

$$p(B|X,Y,\phi,F,G,m)\frac{\exp\{(-\tfrac{1}{2})(\beta-\phi)'F^{-1}(\beta-\phi)\}}{|(V+G)+(B-\hat{B})'(X'X)(B-\hat{B})|^{(N+m-p-1)/2}}. \tag{12}$$

This density being, the product of multivariate normal and matrix T-densities, is very complicated. It is therefore very difficult to use it to make posterior inferences, except numerically. It is straightforward to develop a large sample normal approximation, however.[3] The result is that for large N, it is asymptotically true that

$$\mathcal{L}(\beta|X,Y,\phi,F,G,m) \cong N(\beta_0, J^{-1}), \tag{13}$$

where

$$\beta_0 \equiv \left[F^{-1}+(V+G)^{-1}\otimes(X'X)\right]^{-1}\left\{F^{-1}\phi+\left[(V+G)^{-1}\otimes(X'X)\right]\hat{\beta}\right\},$$

and

$$J \equiv F^{-1}+(V+G)^{-1}\otimes(X'X).$$

Thus, in large samples, posterior inferences about the elements of β may be made from (13), without regard to Σ.

The marginal posterior density of Σ is readily found by using the identity in (9) in eq. (11), completing the square in β, and integrating the resulting normal density with respect to β. The resulting density is

$$\begin{aligned} p(\Sigma|X,Y,\phi,F,G,m) &\propto |\Sigma|^{-(m+N)/2}|F^{-1}+\Sigma^{-1}\otimes(X'X)|^{-1/2} \\ &\exp\left(-\tfrac{1}{2}\right)\mathrm{tr}\left\{\Sigma^{-1}(V+G)+\hat{\beta}'(\Sigma^{-1}\otimes X'X)\hat{\beta}\right. \\ &\quad -\left[F^{-1}\phi+(\Sigma^{-1}\otimes X'X)\hat{\beta}\right]'\left[F^{-1}+\Sigma^{-1}\otimes(X'X)\right]^{-1} \\ &\quad \left.\times\left[F^{-1}\phi+(\Sigma^{-1}\otimes X'X)\hat{\beta}\right]\right\}. \end{aligned} \tag{14}$$

4. MANOVA models

4.1. One way classification

Adopt the p-dimensional, one way layout (classification), fixed effects model. Specifically, assume

$$z_\alpha(t) = \theta_\alpha + v_\alpha(t), \tag{15}$$

where $\alpha = 1,\ldots,q$; $t = 1,\ldots,T_\alpha$; and $z_\alpha(t)$ is a $p\times 1$ response vector. Equivalently, assume

$$\mathcal{L}\left[z_\alpha(t)\right] = N(\theta_\alpha, \Sigma), \quad \Sigma > 0.$$

[3]The approximation is found by expressing the T-density portion of eq. (11) as an exponential, and then letting T become large.

That is, there are observations on q populations, each p-dimensional, with common covariance matrix, and we want to compare the mean vectors, and linear functions of their components. Accordingly, define

$$\underset{(p\times N)}{Y'} \equiv \left[z_1(1),\dots,z_1(T_1);\dots;z_q(1),\dots,z_q(T_q)\right];$$

$$\underset{(p\times N)}{U'} \equiv \left[v_1(1),\dots,v_1(T_1);\dots;v_q(1),\dots,v_q(T_q)\right];$$

$$\underset{(p\times q)}{B'} \equiv \left[(\theta_1)_{(p\times 1)},\dots,(\theta_q)_{(p\times 1)}\right] \equiv \underset{(p\times q)}{\theta};$$

$$\underset{(N\times q)}{X} \equiv \left[\begin{array}{c|c|c|c} 1 & & \cdots & \\ \vdots & & & \\ 1 & & & \\ \hline & 1 & & \\ & \vdots & & \\ & 1 & & \\ \hline & & \vdots & \\ \hline & & & 1 \\ & & & \vdots \\ & & & 1 \end{array}\right] \begin{array}{c} \\ T_1 \\ \\ \\ T_2 \\ \\ \vdots \\ \\ T_q \\ \\ \end{array}$$

Note that $N \equiv \Sigma_1^q T_\alpha$. With these definitions the MANOVA model in eq. (15) becomes the regression model $(Y = XB + U)$.

Suppose we are interested in a set of r comparisons of one dimensional means. Then define

$$\underset{(r\times 1)}{\psi} = \underset{(r\times p)}{C_1} \; \underset{(p\times q)}{B'} \; \underset{(q\times 1)}{C_2} \equiv C_1 \theta C_2, \tag{16}$$

where C_1 and C_2 are constant, preassigned matrices. The components of ψ are linear combinations of the elements of θ. We can make posterior inferences about ψ, or the elements of ψ, from the posterior distribution of ψ.

4.1.1. Non-informative prior

In the case of a non-informative prior distribution on (θ, Σ), the marginal posterior density of θ is the matrix T-density given in eq. (7). The linear transformation in (16) yields for the posterior density of ψ, the multivariate

Student t-density

$$p(\psi|X,Y) \propto \{ C_2'(X'X)^{-1}C_2 + (\psi-\hat{\psi})'(C_1VC_1')^{-1}(\psi-\hat{\psi}) \}^{-(\nu+r)/2}, \tag{17}$$

where

$$\hat{\psi} \equiv C_1\hat{\theta}C_2, \qquad \nu \equiv N-(p+q)+1.$$

Thus,

$$E[\psi|X,Y]=\hat{\psi}, \qquad \operatorname{var}[\psi|X,Y] = \left[\frac{C_2'(X'X)^{-1}C_2}{\nu-2} \right](C_1VC_1').$$

As an illustration of the use of eq. (17) suppose we would like to make posterior inferences about simple contrasts, i.e., simple differences in mean vectors, or in their components. Take $r=p$, so that C_1 is the identity matrix of order p, I. Take C_2 to be the q-vector given by $C_2' \equiv [1,-1,0,\ldots,0]$. Then

$$\underset{(p\times 1)}{\psi} = C_1\theta C_2 = \theta_1 - \theta_2,$$

$$\underset{(p\times 1)}{\hat{\psi}} = \hat{\theta}_1 - \hat{\theta}_2 = \bar{z}_1 - \bar{z}_2,$$

where

$$\bar{z}_\alpha \equiv \frac{1}{T_\alpha} \sum_{t=1}^{T_\alpha} z_\alpha(t).$$

In this case, eq. (17) gives the posterior density for the difference in the mean vectors of populations 1 and 2.

As a second example, take $r=1$, $(C_1)_{(r\times p)} \equiv (1,0,\ldots,0)$, and $(C_2')_{(1\times q)} \equiv (1,-1,0,\ldots,0)$. Then, $\psi \equiv (\theta_{11}-\theta_{12})$. That is, ψ denotes the difference in the first components of the mean vectors of populations 1 and 2. The posterior density of such a simple contrast is found from eq. (17) as

$$p(\psi|X,Y) \propto \left\{ (a_{11}-2a_{12}+a_{22}) + \left(\frac{1}{v_{11}}\right)(\psi-\hat{\psi})^2 \right\}^{-(N-p-q+2)/2} \tag{18}$$

where $(X'X)^{-1} \equiv A = (a_{ij})$, and $V \equiv (v_{ij})$. That is, for $\psi \equiv \theta_{11}-\theta_{12}$,

$$\frac{(\psi-\hat{\psi})}{\left[\frac{v_{11}(a_{11}-2a_{12}+a_{22}}{\nu} \right]^{1/2}}$$

follows a standard univariate Student t-distribution with $\nu \equiv N - p - q + 1$ degrees of freedom.

Note that while in the classical sampling theory case, only confidence interval statements can be made about ψ, and they must be made for what might happen if many samples were to be collected, in the case of Bayesian inference, the entire distribution (posterior) of ψ is available for inferences, and all assertions are made conditional on the single sample actually observed. Joint inferences about several simple contrasts can be made from the higher dimensional marginal densities of (17) (higher than univariate).

4.1.2. *Informative prior*

The marginal posterior distribution of B was given in eq. (12). It is very complicated, as would be inferences based upon it. For this reason we examine, instead, the large sample approximation case. The asymptotic posterior distribution of B is normal, and is given in eq. (13). Since all elements of B are jointly normally distributed in large samples, a posteriori, so are all linear comparisons of means. Thus, in large samples, the posterior distribution of $\psi \equiv C_1 B' C_2$ is also multivariate normal and all contrasts may be readily evaluated.

4.2. *Two-way (and higher) classifications*

4.2.1. *No interactions*

Adopt the p-dimensional, complete, two-way layout with fixed effects, no interaction between effects, and K observations per cell, $K > 1$. Then, the response vector is conventionally written as

$$(z_{ijk})_{(p\times 1)} = \mu + \alpha_i + \delta_j + v_{ij} \equiv (\theta_{ij})_{(p\times 1)} + v_{ijk}, \tag{19}$$

where $i = 1,\dots,I$; $j = 1,\dots,J$; $k = 1,\dots,K$. Here, of course, μ denotes the overall mean, α_i the main effect due to the first factor operating at level i, δ_j denotes the main effect due to the second factor operating at level j, and v_{ijk} denotes an error term. Note that all terms in eq. (19) are p-dimensional column vectors. Since the errors have been assumed to be normally distributed, it follows that

$$\mathcal{L}(z_{ijk} | \theta, \Sigma) = N(\theta_{ij}, \Sigma), \quad \Sigma > 0.$$

Now we place the MANOVA model in (19) in regression format. Accordingly, define

$$\underset{(p\times N)}{Y'} \equiv \left[(z_{111})_{(p\times 1)}, z_{112}, \ldots, z_{11k}; \ldots; z_{IJ1}, \ldots, (z_{IJK})_{(p\times 1)}\right],$$

$$\underset{(p\times N)}{U'} \equiv \left[(v_{111})_{(p\times 1)}, v_{112}, \ldots, v_{11k}; \ldots; v_{IJ1}, \ldots, (v_{IJK})_{(p\times 1)}\right],$$

$$\theta \equiv \underset{(p\times q)}{B'} \equiv \left[(\theta_{11})_{(p\times 1)}, \ldots, \theta_{I1}; \theta_{12}, \ldots, \theta_{I2}; \ldots; \theta_{1J}, \ldots, \theta_{IJ}\right],$$

$$\begin{array}{cc} & \begin{array}{cccc} \leftarrow K \rightarrow & \leftarrow K \rightarrow & \quad\quad & \leftarrow K \rightarrow \end{array} \\ \underset{(p\times N)}{X'} \equiv & \left[\begin{array}{c|c|c|c} 1,\ldots,1 & & & \\ \hline & 1,\ldots,1 & & \\ \hline & & & \\ \hline & & & 1,\ldots,1 \end{array}\right], \end{array}$$

where

$$q \equiv IJ, \qquad N = IJK.$$

Now

$$Y = XB + U.$$

For this situation, for $\psi_{(r\times 1)} = (C_1)_{(r\times p)}\theta_{(p\times q)}(C_2)_{(q\times 1)}$, if $r = p$, and $C_1 = I$, so that $\psi = \theta C_2$, and a non-informative prior, the posterior density of ψ is given by the multivariate Student t-density

$$p(\psi \mid X, Y) \propto \left\{ C_2'(X'X)^{-1}C_2 + (\psi - \hat{\psi})' V^{-1}(\psi - \hat{\psi}) \right\}^{-(\nu+p)/2},$$

where $\nu \equiv IJ(K-1) - (p-1)$.

As an example, take the case in which $(C_2')_{(1\times q)} \equiv [1, -1, 0, \ldots, 0]$, so that

$$\underset{(p\times 1)}{\psi} = (\theta_{11})_{(p\times 1)} - (\theta_{21})_{(p\times 1)}.$$

But $\theta_{11} - \theta_{21} = \alpha_1 - \alpha_2$, and $\hat{\theta}_{11} - \hat{\theta}_{21} = \hat{\alpha}_1 - \hat{\alpha}_2$. By choosing C_2 appropriately, we obtain the posterior distribution of any linear combination of main effect vectors, and by choosing $C_1 \neq I$, we obtain linear combinations of their components.

For the case of an informative prior, normally distributed posteriors result for all contrasts, as in the case of a one way classification.

Inferences about contrasts for higher way layouts for both non-informative and informative priors are obtained in the same way as we have carried out the analyses for the one and two way layouts.

4.2.2. *MANOVA with interaction*

For the simplest case of a fixed effects model with interaction we examine the complete two way layout with K observations per cell, $K>1$. Then,

$$(z_{ijk})_{(p\times 1)}=\mu+\alpha_i+\delta_j+\gamma_{ij}+v_{ijk}\equiv(\theta_{ijk})_{(p\times 1)}+v_{ijk},$$

where all terms are p-dimensional; γ_{ij} denotes the effect of an interaction between the first factor operating at level i, and the second factor operating at level j; $i=1,\dots,I$; $j=1,\dots,J$; and $k=1,\dots,K$. It now follows that

$$\mathcal{L}(z_{ijk}|\theta,\Sigma)=N(\theta_{ij},\Sigma),\quad \Sigma>0.$$

Note that (X,Y,B,U) may all be defined exactly as they were above for the case of a two way layout without interaction (the difference between the two models becomes apparent only in the definition of θ). Thus, as before, if $\psi_{(p\times 1)}=\theta_{(p\times q)}(C_2)_{(q\times 1)}$, and $\hat{\psi}=\hat{\theta}C_2$, where $q\equiv IJ$, the posterior density of ψ is given by

$$p(\psi|X,Y)\propto\left\{C_2'(X'X)^{-1}C_2+(\psi-\hat{\psi})'V^{-1}(\psi-\hat{\psi})\right\}^{-(\nu+p)/2},$$

for $\nu\equiv IJ(K-1)-(p-1)$.

Now recall that to ensure estimability (identifiability) of all of the parameters, it is customary to impose the constraints

$$\alpha_+=0,\quad \delta_+=0,\quad \gamma_{i+}=0 \text{ for all } i,\quad \text{and } \gamma_{+j}=0 \text{ for all } j,$$

where a plus denotes an averaging over the subscript. Thus $\alpha_+\equiv I^{-1}\Sigma_1^I\alpha_i$. It now follows that

$$\gamma_{ij}=\theta_{ij}-\theta_{i+}-\theta_{+j}+\theta_{++}.$$

So every γ_{ij} is just a linear function of the θ_{ij} (as are the α_i and δ_j). Thus, posterior inferences about the main and interaction effects may be made by judicious selection of C_2.

Inferences in higher way layouts with interaction effects are made in a completely analogous way.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Dickey, J. M. (1967). Matric-variate generalizations of the multivariate t-distribution and the inverted multivariate t-distribution. *Ann. Math. Statist.* **38**, 511–518.

Fraser, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.

Geisser, S. (1965). Bayesian estimation in multivariate analysis. *Ann. Math. Statist.* **36** (1) 150–159.

Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. B* **25**, 368–376.

Jeffreys, H. (1961). *Theory of Probability* (third edition). Clarendon Press, Oxford.

Kshirsagar, A. M. (1960). Some extensions of the multivariate t-distribution and the multivariate generalization of the distribution of the regression coefficient. *Proc. Cambridge Phil. Soc.* **57**, 80–85.

Press, S. J. (1972). *Applied Multivariate Analysis*. Holt, Rinehart and Winston, New York.

Press, S. J. (1979). Bayesian Computer Programs, In: A. Zellner, ed., *Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys*. North-Holland, Amsterdam.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard University Press, Boston.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423, 623–656.