SOME MODELS FOR DESIGNED EXPERIMENTS

The analysis of designed experiments usually consists, among other things, of an analysis of variance, which is a way to determine the effect of various factors on a response variable y. For example, in a completely randomized design t treatments are assigned at random to n experimental units. Suppose that n_i is the number of units receiving the i-th treatment. If y_{ij} denotes the response measured from the j-th unit receiving treatment i, a one-way analysis of variance is performed in order to determine if each treatment produces the same average response.

The analysis of variance partitions the total sum of squares into "between" treatment and "within" treatment components, and if the ratio of their corresponding mean squares is "large," significant differences between the treatments are said to exist, that is, the average responses produced by the treatments are not the same. Fisher (1953), Kempthorne (1952), Cochran and Cox (1957), and Cox (1958) are standard references for the design and analysis of experiments and Graybill (1961) and Searle (1971) give the theory for those models which are appropriate for the analysis of designed experiments.

This section will give the Bayesian analysis of designed experiments and is the analog of the classical approach given by Graybill (1961) and Searle (1971). The so-called one-way model

$$y_{ij} = \theta_i + e_{ij}, \quad i = 1,2,\ldots,t, \quad j = 1,2,\ldots,n_i \tag{3.92}$$

models the completely randomized design, where y_{ij} is the response measured from the j-th unit receiving treatment i, θ_i is the mean of the i-th treatment, and the e_{ij}'s are n.i.d. $(0, \tau^{-1})$, where $\tau > 0$, thus the average response from treatment i is θ_i, which is considered a real unknown parameter.

The analysis of variance is a way to test H_0: $\theta_1 = \theta_2 = \ldots = \theta_t$ versus the alternative that is not true, and it is equivalent to the likelihood-ratio test. This is shown by Graybill (1961) and Searle (1971). The AOV test is to reject H_0 at level α whenever

$$F = BMS/WMS > F_{\alpha/2,t-1,n-t,} \tag{3.93}$$

where

$$BMS = \sum_{i=1}^{t} (y_{i\cdot} - y_{\cdot\cdot})^2 / (t-1)$$

and

$$WMS = \sum_{1}^{t} \sum_{1}^{n_i} (y_{ij\cdot} - y_{i\cdot})^2 / (n-t).$$

The Bayesian approach to the analysis of the completely randomized design is based on an HPD region for the vector of treatment effects $\theta = (\theta_1, \theta_2, \ldots \theta_t)'$

The Completely Randomized Design

The one-way model (3.92) is a special case of the general linear model of Chapter 1, where y is $n \times 1 (n = \sum_1^t n_i)$ and

$$Y = \begin{matrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(t)} \end{matrix}$$

where $y^{(i)}$ is $n_i \times 1$ with components $y_{i1}, y_{i2}, \ldots, y_{in_i}$, and the design matrix X is $n \times t$, where

$$x = \begin{matrix} x_1 \\ x_2 \\ : \\ x_t \end{matrix}$$

where x_i is a $n_i \times t$ matrix with i-th column consisting of ones, and the remaining columns are zero. The error vector ε is $n \times 1$ and contains n.i.d. $(0, \tau)$ random variables.

The likelihood function for θ and τ is

$$L(\theta,\tau|y,x) \propto \tau^{n/2} \exp -\tfrac{\tau}{2} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2, \quad \theta \in R^t, \quad \tau > 0 \tag{3.94}$$

and suppose the prior information of the parameters is a normal-gamma distribution with parameters $y \in R^p$, p, $\alpha > 0$, and $\beta > 0$, then by Bayes theorem

$$\xi(\theta,\tau|y,x) \propto \tau^{(n+2\alpha+1)/2-1}\exp -\tfrac{\tau}{2}\{2\beta+\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2 + (\theta-\mu)'p(\theta-\mu)\}, \tag{3.95}$$

where $\theta \in R^p$ and $\tau > 0$, is the joint posterior density of θ and τ, which may be expressed as

$$\xi(\theta,\tau|y,x) \propto \tau^{(n+2\alpha+1)/2-1}\exp -\tfrac{\tau}{2}\left\{[\theta-A^{-1}B]'A[\theta-A^{-1}B] + C - B'A^{-1}B\right\}, \tag{3.96}$$

where

$$A=\mathrm{Diag}\,(n_1,n_2,\ldots,n_t)+p=D+p, \tag{3.97}$$

$$B=\begin{pmatrix}y_{1\cdot}\\ y_{2\cdot}\\ \vdots\\ y_{t\cdot}\end{pmatrix}+p\mu=T+p\mu \tag{3.98}$$

and

$$C=y'y+2\beta+\mu'p\mu, \tag{3.99}$$

where$y'y=\sum_{i=1}^{t}\sum_{j=1}^{n_i}y_{ij}^2$ and $y_i=\sum_{j=1}^{n_i}y_{ij}$ for i=1,2,…,t.

Since the treatment effects are the parameters of interest, the marginal posterior distribution of θ is a t distribution with mean

$$E[\theta|x,y]=A^{-1}B=(D+\mu)^{-1}(T+p), \tag{3.100}$$

precision matrix

$$P[\theta|x,y]=\frac{(n+2\alpha)A}{C-B'A^{-1}B}, \tag{3.101}$$

and $n + 2\alpha$ degrees of freedom. The posterior dispersion matrix of θ is

$$D[\theta|x,y]=(C-B'A^{-1}B)A^{-1}(n+2\alpha-2)(n+2\alpha)^{-1} \tag{3.102}$$

Now how does one test H_0: $\theta_1 = \theta_2 = \ldots = \theta_t$? Following Box and Tiao (1973), one finds an HPD region for γ, where $\gamma = G\,\theta$ such that H_0 is true if and only if $\gamma = 0$. For example, γ might be a $(t-1)\times 1$ vector of contrasts γ_i of the t treatment effects, say

$$\gamma_i=\theta_i-\theta_{i+1},\quad i=1,2,\ldots,\quad t-1,$$

hence G is a $(t-1)\times t$ matrix

$$G=\begin{pmatrix}1 & -1 & 0 & \cdots & 0 & 0\\ 0 & 1 & -1 & \cdots & 0 & 0\\ & & \vdots & & & \\ 0 & 0 & 0 & \cdots & 1 & -1\end{pmatrix}.$$

Transforming from θ to γ gives a t-distribution for γ with mean $GA^{-1}B$ and precision matrix $[GP^{-1}[\theta|x,y]G']^{-1}$ with $n + 2\alpha$ degrees of freedom and a $1-\Delta$ HPD region for γ is

$$\mathrm{HPD}_\Delta(\gamma)=\{\gamma : F(\gamma)\leqslant F_{\Delta;t-1,n+2\alpha}\} \tag{3.103}$$

where

$$F(\gamma)=[\gamma-GA^{-1}B]'[GP^{-1}[\theta|x,y]G']^{-1}[\gamma-GA^{-1}B](t-1)^{-1} \tag{3.104}$$

has an F distribution with $t-1$ and $n+2\alpha$ degrees of freedom.

Thus, if $F(0) > F_{\Delta;t-1,\,n+2\alpha}$, then H_0 is rejected at level Δ, $0<\Delta<1$. One important problem remains to be solved, namely how does one assign values to the hyperparameters μ, p, α and ^{2}? First, let us see what happens when one assigns a Jeffreys' prior

$$\xi(\theta,\tau)\propto\tau^{-1},\quad \theta\in R^t,\quad \tau>0, \tag{3.105}$$

$$\xi(\theta,\tau) \propto \tau \quad , \quad \theta \in R^{t}, \quad \tau > 0, \qquad (3.105)$$

which is equivalent to letting $\beta \to 0$, $p \to 0(t \times t)$, and $\alpha \to -t/2$ in the posterior density of θ. This yields

$$A^{-1}B = \begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \vdots \\ \bar{y}_{t\cdot} \end{pmatrix} \qquad (3.106)$$

for the posterior mean of θ, which is the vector of treatment means, while the precision matrix is

$$P[\theta|x,y] = \frac{D}{s^2}, \qquad (3.107)$$

where $(n-t)s^2 = \sum_{i=1}^{t} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{t} y_{i\cdot}^2/n_i$, and s^2 is called the within mean square of the AOV. Substituting these into F(γ), then $F(0) > F_{\Delta;t-1,\,n-t}$, if and only if BMS/WMS > $F_{\Delta,\,t-1,\,n-t}$, which is the AOV test of H_0 the equality of treatment effects. The HPD method of testing H_0, using an improper prior density, is equivalent to the AOV; however, the HPD region for γ, given by (3.103), is more general than the AOV because the former allows prior information in the form of a normal-gamma density with parameters μ, p, α and β. But this poses another problem, since the hyperparameters must be assigned legitimate values that actually express one's prior information. How should values for μ, p, α and β be assigned?

By using the improper prior density one is saying the treatment effects and error precision are independent, that the treatments all have the same effect on the average response, and finally that error precisions closer to zero are more likely than those far away from zero, thus a Jeffreys' improper prior density weakly supports the null hypothesis.

The same support for the null hypothesis can be achieved with the conjugate prior density if the components of μ are all equal and if p is diagonal. Still another way to express prior knowledge of the treatment effects θ_i is to assume them to be exchangeable, i.e., one's prior opinion about θ_1 is the same as that of θ_3, or any other θ_i, and similarly for pairs and triplets. In particular, one could assume the treatment effects are i.i.d. n(0, η), and the θ_i (i = 1, 2, …, t) would be exchangeable, and the null hypothesis would be supported, a priori. Exchangeability is formulated by de Finetti (1972) and applied by Lindley and Smith (1972) to the Bayesian analysis of linear models, but will not be emphasized here.

It is relatively easy to support the null hypothesis with a normal-gamma prior, however it remains to assign values to the hyperparameters. If one has a past experiment

$$z = x^{*}\theta + \varepsilon, \qquad (3.108)$$

with the same treatments, one may use the prior predictive distribution of z to assign values to the hyperparameters, and this method was explained in the second section of the chapter. If (3.108) represents a future experiment, the experimenter is asked to predict z; the future observations y, when the design is completely randomized; that is, the experimenter must assign at least one response for each of the t treatments, then the values of the hyperparameters are assigned by fitting them to the prior predictive density (3.3). Of course, past experiments of the same form as the one which will be performed give less subjective values for the hyperparameters than hypothetical future experiments.

Returning to the analysis of a completely randomized design, the AOV or the HPD method of testing the equality of treatment effects is only one step in the analysis of such a design. It should be remembered, the marginal posterior distribution of θ tells the complete story about the treatment effects in the sense that once the data y are observed, all inferences are based on this distribution of θ.

If H_0 is rejected, a careful inspection of the marginal posterior distribution of θ is necessary in order to see in what way the treatments are affecting the average response. For example, are the treatment effects partitioned into various groups or clusters?

Consider three treatments and their effects θ_1, θ_2, and θ_3. Suppose for every ε > 0, $P[|\theta_1 - \theta_2| < \varepsilon \,|y] > P[|\theta_2 - \theta_3| < \varepsilon|y]$, then the data suggest θ_2 is "closer" to θ_1 than it is to θ_3, and thus $\{\theta_2, \theta_1\}$ forms one cluster, while $\{\theta_3\}$ is by itself, so to speak. These probabilities are easily evaluated, because $\theta_i - \theta_j$ has a known univariate t distribution and one may compute $P[|\theta_i - \theta_j| < \varepsilon|y]$ by Student's t-tables.

THE ANALYSIS OF TWO-FACTOR EXPERIMENTS

In the previous section, the completely randomized experiment was analyzed. Such a one-way layout is an example of a one-factor experiment, where the various levels of that factor may have distinct influences on the level of the average response. The factor was the treatment and the treatment effects were the levels of the treatment factor.

Such an experiment was analyzed as if the t treatments were fixed, that is, one is only interested in making inferences about the set of treatments which were included in the experiment and not to a population of treatments, from which the treatments were selected.

In the latter case, the treatment factor would have been considered random, that is the different treatments would have been selected at random from a population of treatments, and the experiment would have been examined by the techniques of Chapter 4. This chapter deals only with fixed factors.

With a two-factor experiment, the average response sij may be influenced by the levels of the two factors, namely the i-th level of the first and the j-th level of the second factor, where i = 1, 2, …, t and j = 1, 2, …, b. Suppose for each of the bt combinations of factor levels, one has n_{ij} measurements on the response and suppose the $n = \sum_{1}^{t}\sum_{1}^{b} n_{ij}$. observations are independent, then if y_{ijk} is the k-th measurement, when level i of the first factor and level j of the second are present,

$$y_{ijk} = s_{ij} + e_{ijk}$$

where the e_{ijk}'s are independent random variables with mean zero and precision τ, i.e., the average response is $E(y_{ijk}) = s_{ij}$ for all i, j, and k.

If the two factors are additive, i.e., $s_{ij} = m + a_i + b_j$ then a_i is referred to as the i-th effect of the first factor and b_j the effect of the j-th level of the second factor, where the a_i the b_j and m are unknown real parameters. If the two factors are not additive $s_{ij} = m + a_i + b_j + (ab)_{ij}$, $(ab)_{ij} \neq 0$, for some i and

second factor, where the a_i the b_j and m are unknown real parameters. If the two factors are not additive $s_{ij} = m + a_i + b_j + (ab)_{ij}$, $(ab)_{ij} \neq 0$, for some i and j, and $(ab)_{ij}$ is called the interaction between the i-th and j-th levels of the two factors, where the tb interaction effects are unknown real parameters, and in this case there are tb + t + b + 1 parameters, excluding the common precision parameter τ of the errors.

Additive Models