COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 384

Analysis of Variance With A Certain Linear Restriction

F. A. Bobkoski and L. G. Telser

April 21, 1954

Suppose that we have a two way analysis of variance problem where the random variable X_{ij} can be written as

$$X_{ij} = a_i + b_j + \mu + c_{ij} + u_{ij} \qquad \begin{array}{l} i = 1,\ldots, I \\ j = 1,\ldots, J \end{array}$$

The a_i's, b_j's, μ, and c_{ij}'s are to be regarded as fixed parameters. The variable u_{ij} is the random disturbance in the equation and $\mathrm{Var}\ u_{ij} = \sigma^2$ and $E\ u_{ij} = 0$. The problem as stated so far is the usual model for a two way analysis of variance.

However in addition to these requirements suppose that it is known that $\sum_j u_{ij} = K$ where K is some constant for all i. This means that the error terms are dependent. In order to proceed with the analysis it becomes necessary to make an assumption regarding the nature of the dependence of the u_{ij}'s for a given i. We shall assume that $E\ u_{ij}\ u_{ij'}$ are equal for each i, j, and j'. This implies that the covariance of the u_{ij}'s have a particularly simple value.

Consider Var $\sum_j u_{ij}$. It is clear that Var $\sum_j u_{ij} = 0$

$$\text{Var} \sum_j u_{ij} = E\,(\sum_j u_{ij})^2$$

$$E\,(\sum_j u_{ij})^2 = E\,(\sum_j \sum_k u_{ij}\, u_{ik}) = \sum_{j=1}^{J} \sum_{k=1}^{J} E\,(u_{ij}\, u_{ik})$$

But when $j = k$ we have $E\, u_{ij}\, u_{ik} = \sigma^2$ and when $j \neq k$ we have $E\, u_{ij}\, u_{ik} = \text{Cov}\; u_{ij}\, u_{ik} = \lambda$ say.

Therefore

$$E\,(\sum_j u_{ij})^2 = J\,\sigma^2 + 2\,\binom{J}{2}\,\lambda = 0$$

Hence

$$\lambda = -\frac{J\sigma^2}{2\binom{J}{2}} = -\frac{\sigma^2}{(J-1)}$$

Since there is no replication it is impossible to estimate the interaction term of first order c_{ij}. Furthermore we assume that $\sum_j X_{ij} = 0$ for all i. This is the motivation for the assumption that $\sum_j u_{ij} = K$. In particular this means that $\sum_i \sum_j X_{ij} = X_{..} = 0$. Hence the grand mean of the observation will also be zero. uppose we want Markoff estimates of a_i, b_j, and μ.

$$\sum_j X_{ij} = Ja_i + b. + J\mu + \sum_j u_{ij} = 0$$

To obtain the Markov estimates we wish to

$$\min \sum_{i,j} (X_{ij} - a_i - b_j - \mu)^2 \quad \text{subject to} \quad \sum_j u_{ij} = -(Ja_i + b. + J\mu)$$

so that we have I constraints. We introduce I Lagrangean parameters ρ_i and minimize the following expression:

$$\sum_{i,j} (X_{ij} - a_i - b_j - \mu)^2 + \sum_i \rho_i \,(Ja_i + b_i + J\mu)$$

$$\left.\begin{aligned}
\frac{\partial}{\partial a_i} &: \; 2\sum_j (X_{ij} - a_i - b_j - \mu) + \rho_i J = 0 \\
\frac{\partial}{\partial b_j} &: \; 2\sum_i (X_{ij} - a_i - b_j - \mu) + \sum_i \rho_i = 0 \\
\frac{\partial}{\partial \mu} &: \; 2\sum_{i,j} (X_{ij} - a_i - b_j - \mu) + J\sum_i \rho_i = 0 \\
& Ja_i + b. + J\mu = 0
\end{aligned}\right\}$$

These equations simplify to:

$$\left.\begin{aligned}
& X_{i.} - Ja_i - b. - J\mu + \frac{\rho_i}{2} J = 0 \\
& X._j - a. - Ib_j - I\mu + \frac{1}{2}\sum_i \rho_i = 0 \\
& X.. - Ja. - Ib. - IJ\mu + \frac{J}{2}\sum_i \rho_i = 0 \\
& Ja_i + b. + J\mu = 0
\end{aligned}\right\}$$

Hence we see that $\rho_i = 0$ for every i and that $X._j - a. - Ib_j - I\mu = 0$ is the only set of independent equations remaining. If we make the usual assumption that $a. = b. = 0$ we see that $\mu = 0$ so that

$$\hat{b}_j = \frac{X._j}{I} = \overline{X}._j \quad .$$

None of the a_i parameters are estimable because our restriction on the $X_{i.}$

causes all of these equations to vanish identically. Ordinarily we would estimate μ by $X..$ and we would take $\hat{b}_j = \bar{X}._j - \bar{X}..$, and $\hat{a}_i = \bar{X}_{i.} - \bar{X}..$ We see that the estimate for b_j is what it would be without the restrictions when we recall that $\bar{X}.. = 0$.

In order to make tests on the significance of the $\hat{b}_j$'s we have to specify a distribution for the error terms, u_{ij}'s. We shall assume that the u_{ij}'s are normal with means, variance and covariance as we have specified.

First let us learn the expected value of $\hat{b}_j$ and its variance. Obviously $E\hat{b}_j = b_j$

$$\text{Var } \hat{b}_j = E\,[\bar{X}._j - b_j]^2 = \frac{1}{I^2} E\,(X._j)^2 - b_j^2$$

After some calculation we find that

$$\text{Var } \hat{b}_j = \frac{\sigma^2}{I}$$

$$\sum_{i,j} (X_{ij} - \mu - a_i - b_j)^2 = \sum_{ij} (X_{ij} - \mu + \hat{\mu} - \hat{\mu} + \hat{a}_i - \hat{a}_i + \hat{b}_j. - \hat{b}_j - a_i - b_j)^2$$

We may as well take $a_i = \hat{a}_i \equiv 0$ and $\mu = \hat{\mu} = 0$ since it turns out that their variance is zero and our estimates of them are identically zero. Hence

$$\sum_{ij} (X_{ij} - \mu - a_i - b_j)^2 = \sum_{ij} (X_{ij} - \hat{b}_j)^2 + \sum_{ij} (\hat{b}_j - b_j)^2$$

Taking expected values on both sides we have

$$I\,J\,\sigma^2 - I\,J\,.\,\frac{\sigma^2}{I} = E \sum_{ij} (X_{ij} - \hat{b}_j)^2$$

$$\sigma^2\,J\,(I - 1) \qquad = E \sum_{ij} (X_{ij} - \hat{b}_j)^2 \quad .$$

We claim that the estimate of the residual sums of squares has $(I - 1)(J - 1)$ degrees of freedom. Since there are I restrictions on the observations the total number of degrees of freedom is

$$I\,J - I = I\,(J - 1)$$

The degrees of freedom due to the estimate of b_j is $J - 1$. Therefore the degrees of freedom of the residual is

$$I\,(J - 1) - J - 1 = (I - 1)\,(J - 1).$$

$$1/(I - 1)\,(J - 1) \cdot E \sum_{ij} (x_{ij} - \hat{b}_j)^2 = \frac{\sigma^2 J(I-1)}{(I-1)(J-1)} = \sigma^2 \frac{J}{J-1}$$

Similarly the $\dfrac{E \sum_{ij} (\hat{b}_j - b_j)^2}{J - 1} = \dfrac{\sigma^2 J}{(J-1)}$.

$$\therefore \frac{\frac{1}{(J-1)} \sum_{j} (\bar{x}_{\cdot j}{}^2_{\cdot})\ .}{\frac{1}{(I-1)(J-1)} \sum_{ij} (x_{ij} - \bar{x}_{\cdot j})^2}$$ has the F distribution with

$(J-1)$ and $(I-1)\,(J-1)$ degrees of freedom respectively.

A model similar to this was used to analyze the gains and losses of traders in future markets.