Ann. Inst. Statist. Math.
35 (1983), Part A, 77-90

AN APPROACH TO DEFINING THE PATTERN OF INTERACTION EFFECTS IN A TWO-WAY LAYOUT

C. Hirotsu

(Received Nov. 5, 1981; revised June 29, 1982)

Summary

A method is given to classify rows and columns into subgroups so that additivity holds within each of the subtables made of the grouped rows or the grouped columns. The least squares estimators of the cell means are easily obtained for the resulting linear model together with their variances. An estimator of the error variance σ^2 is given when there is only one observation per cell. A treatment of an ordered table is also given.

1. Introduction

Suppose that we are given two-way observations with replications y_{ijk} and assume the model

$$y_{ijk}=\mu_{ij}+\varepsilon_{ijk} \qquad (i=1,\cdots,a;\ j=1,\cdots,b;\ k=1,\cdots,r),$$

where the ε_{ijk} are independently and normally distributed with mean zero and variance σ^2.

In testing the hypothesis of no interaction the usual F-test can be applied. If the hypothesis is accepted, the μ_{ij} can be modeled by $\mu_{ij}=\mu+\alpha_i+\beta_j$. When the hypothesis of no interaction is rejected, however, we are faced with a more complicated model since the degrees of freedom for interactions are usually large, and it is desirable to simplify the structure of the interaction.

Practically it would be helpful if we can partition the sets of levels of both the factors so that interaction exists only between pairs of factor levels belonging to different sets, that is, $\mu_{ij}-\mu_{i'j}-\mu_{ij'}+\mu_{i'j'}=0$ unless both i and i', j and j' belong to different sets. In this paper we give a method for finding such a partition with the proposition that the probability of judging any partition to be significant when actually the additivity model holds is at most equal to a preassigned significance level.

When $r=1$, special structures in interaction effects are often assumed such as in Tukey [15], Mandel [11], [12] and Johnson and Graybill [9]. Although these methods are often successful, it would be helpful to have other approaches for those cases where there is not any intrinsic interest in the special models. Our method can also be

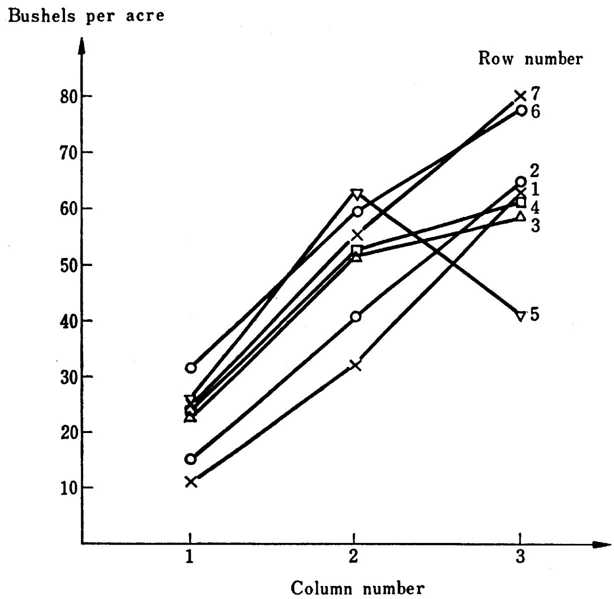

Figure 1. Yields of corn in bushels per acre.

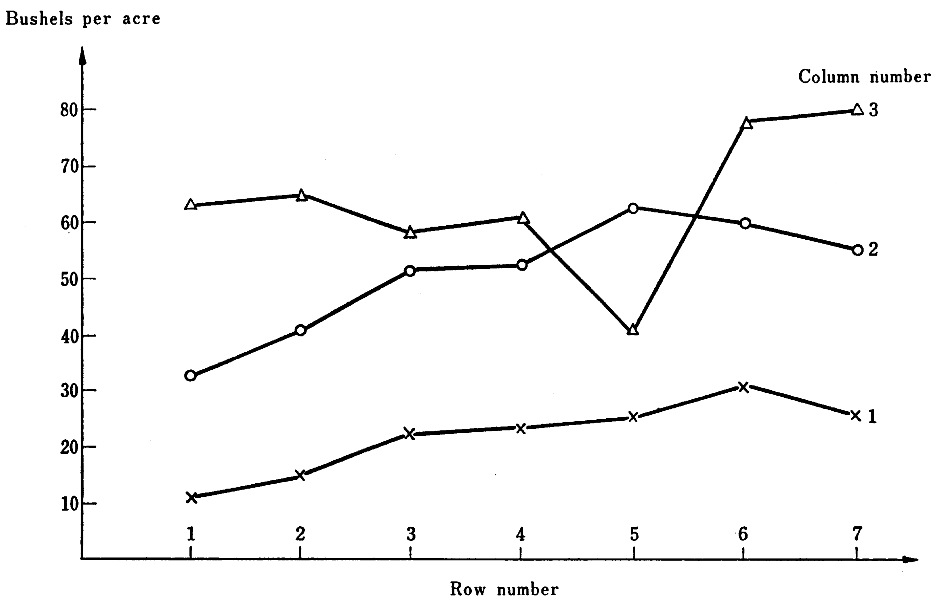

Figure 2. Yields of corn in bushels per acre.

applied to the case when $r=1$. For this case Johnson and Graybill [8] gave an approach in which they also assumed no special structures in the interaction effects. Our method is expected to complement the graphical technique introduced by Bradu and Gabriel [2] to find subtables with simple structures.

An example is given to illustrate some of the ideas. The two-way data in Table 4 of Johnson and Graybill [9] are plotted in Figures 1 and 2 by the method of Mandel [10]. They suggest the existence of some interaction. The first plot suggests that rows 1, 2, 6 and 7 do not interact with columns and that rows 3 and 4 do not interact with columns. The interaction is between the sets of rows {1, 2, 6, 7}, {3, 4} and {5}. Similarly the second plot suggests that columns 1 and 2 do not interact with rows and that the column 3 is very different from them. These plots suggest the pertinency of considering the row-wise interaction models introduced in Hirotsu [4]. A multiple comparison method is required since the grouping of rows and/or columns is not given in advance to obtaining the data.

2. Row-wise and column-wise interactions and a simplified model

Reparametrize the μ_{ij} as

$$\boldsymbol{\mu}=\underset{+}{\boldsymbol{\mu}}+(P_a\otimes P_b)\boldsymbol{\gamma}$$

where $\boldsymbol{\mu}$ is a column vector of μ_{ij}'s arranged in dictionary order, $\underset{+}{\boldsymbol{\mu}}$ is the additive part of $\boldsymbol{\mu}$ with the $b(i-1)+j$th element being $\mu_{i.}+\mu_{.j}-\mu_{..}$ and $\boldsymbol{\gamma}=(P_a'\otimes P_b')\boldsymbol{\mu}$ is the interaction part. We use the usual dot notation so that $\mu_{i.}$ means the average of μ_{ij} with respect to j, $\otimes$ denotes the Kronecker product and P_n' is such an $(n-1)$ by n matrix that $\begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ P_n' \end{bmatrix}$ is an orthogonal matrix of order n, so that $P_nP_n'=I_n-n^{-1}\boldsymbol{j}_n\boldsymbol{j}_n'$, where I_r is an identity matrix of order r and $\boldsymbol{j}_r$ an r dimensional vector of 1's.

The contribution of two particular rows, the mth and the nth rows, say, to $\boldsymbol{\gamma}$ is given by

$$L(m;n)=(1/\sqrt{2})P_b'(\boldsymbol{\mu}_m-\boldsymbol{\mu}_n)\ , \tag{2.1}$$

where $\boldsymbol{\mu}_i=(\mu_{i1},\cdots,\mu_{ib})'$. This is called an interaction element between the two rows (Hirotsu [4]). If it is known to be zero, one can take into consideration only those contrasts which are orthogonal to it. Thus if by any means one can classify rows into homogeneous subgroups so that in each of them all interaction elements are zero, one can have a much simplified model. The same can be done with the columns.

The resulting model may be expressed in terms of nonzero elements of γ. However, a more convenient expression of the model is,

$$\mu_{ij}=\mu_{i.}+\mu_{.j}-\mu_{..}+(\alpha\beta)_{ij} \tag{2.2}$$

with $(\alpha\beta)_{i.}=0$, $(\alpha\beta)_{.j}=0$ and $(\alpha\beta)_{ij}=(\alpha\beta)_{i'j'}$ if $i, i' \in H_u$ and $j, j' \in J_v$, where H_u, $u=1,\cdots,A$, and J_v, $v=1,\cdots,B$, denote homogeneous subgroups of rows and columns, respectively.

The least-squares estimators $\hat{\mu}$, $\hat{\gamma}$ and $\hat{L}(m;n)$ are obtained simply by replacing μ_{ij} by $y_{ij.}$ in their defining equations. The variance-covariance matrix of $\hat{\mu}$ is also easily obtained by virtue of orthogonality relations among coefficients. They lead to formulas (2.3) and (2.4) for the μ_{ij}, $i \in H_u$, $j \in J_v$.

$$\begin{aligned}\hat{\mu}_{ij}=y_{i..}+y_{.j.}-y_{...}+[\sum_{i\in H_u}\sum_{j\in J_v} y_{ij.}/\{n(H_u)n(J_v)\}\\ -\sum_{i\in H_u} y_{i..}/n(H_u)-\sum_{j\in J_v} y_{.j.}/n(J_v)+y_{...}]\end{aligned} \tag{2.3}$$

$$\mathrm{V}(\hat{\mu}_{ij})=[(a+b-1)/ab+\{a-n(H_u)\}\{b-n(J_v)\}/\{abn(H_u)n(J_v)\}](\sigma^2/r)\ . \tag{2.4}$$

In (2.3) and (2.4) $n(H_u)$ and $n(J_v)$ denote the number of rows and columns contained in the sets H_u and J_v, respectively.

In case $r=1$, an estimator of σ^2 is given by

$$\tilde{\sigma}^2=\sum_i\sum_j (y_{ij}-\hat{\mu}_{ij})^2/f=\{T-\|\hat{\gamma}\|^2\}/f\ , \tag{2.5}$$

where $T=\sum\sum(y_{ij}-y_{i.}-y_{.j}+y_{..})^2$, $\|v\|^2$ is the squared norm of a vector v and $f=(a-1)(b-1)-(A-1)(B-1)$ with $(A-1)(B-1)$ being the number of orthogonal interaction contrasts remaining in the model. It should be noted that the $\hat{\mu}_{ij}$ and the $\tilde{\sigma}^2$ are better estimators of the μ_{ij} and σ^2 than those which might be obtained from separate subtables.

In Sections 3 and 4 simultaneous tests for no interaction are given for the cases $r\geqq 2$ and $r=1$, respectively, so that the probability of judging any interaction element, which is actually equal to zero, to be significantly different from zero is at most equal to the preassigned significance level.

3. Simultaneous test procedures for interaction elements when $r\geqq 2$

3.1. *Simultaneous tests for interaction elements between rows*

The hypotheses

$$H(m;n):L(m;n)=0 \qquad (m, n=1,\cdots,a)$$

are tested simultaneously by comparing the sums of squares

$$S(m;n)=r\|\hat{L}(m;n)\|^2=(r/2)\sum_j\{y_{mj\cdot}-y_{m\cdot\cdot}-(y_{nj\cdot}-y_{n\cdot\cdot})\}^2 \quad (m,n=1,\cdots,a)$$

with $(a-1)(b-1)\hat{\sigma}^2F\{(a-1)(b-1),ab(r-1);\alpha\}$, where $\hat{\sigma}^2$ denotes the usual unbiased estimate of the variance σ^2 and $F(\nu_1,\nu_2;\alpha)$ denotes the upper $100\alpha\%$ point of the F distribution with the degrees of freedom (ν_1,ν_2). It is easy to verify that the null hypothesis of additivity H_0 is simply the union of all the $H(m;n)$. The function $S(m;n)$ is first introduced in Hirotsu [4] and is called the squared distance between the two rows. Note that the usual sum of squares for interaction can be written in matrix notation as

$$T=r\|(P_a'\otimes P_b')\boldsymbol{y}\|^2, \tag{3.1}$$

where $\boldsymbol{y}$ is a vector of $y_{ij\cdot}$'s arranged in dictionary order. Then it is easy to see that the $S(m;n)$ are the components of T and we have

$$\Pr[\max_{m,n}S(m;n)\geqq(a-1)(b-1)\hat{\sigma}^2 \times F\{(a-1)(b-1),ab(r-1);\alpha\}\,|\,H_0]\leqq\alpha\,. \tag{3.2}$$

As will be mentioned in Section 4, $S(m;n)$ is more sharply bounded by the largest latent root of a Wishart matrix and the use of it is discussed in Johnson [7] in a slightly different context. However, we use the criterion (3.2) here because of a procedure to be given in Subsection 3.3. Rows are classified into subgroups so that in each of them no significant element is included. This is most easily done by making an a by a matrix of squared distances and then by rearranging the elements so that the (s,t)th element is $S(m_s;m_t)$ where $S(m_1;m_a)$ is the largest $S(m;n)$ $(m,n=1,\cdots,a)$ and $S(m_1;m_i)$ is the $a-i+1$th largest $S(m_1;n)$ $(n\neq m_1)$, see Tables 1a and 1b. Some desirable properties of the $S(m;n)$ for this procedure are shown in Hirotsu [4]. In our experience there will not be much difference if the elements are rearranged according to the size of $S(m;m_a)$ $(m\neq m_a)$.

3.2. *Interaction element between subgroups*

In this section a procedure is given which measures the contribution of two subgroups of rows to the sum of squares for interaction. Without loss of generality let the first subgroup be composed of the first p_1 rows and the second subgroup be composed of the subsequent p_2 rows $(p_1+p_2\leqq a)$. Then we define the interaction element between the two subgroups by

$$\begin{aligned}L(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)\\ =\{p_1p_2(p_1+p_2)\}^{-1/2}\{(p_2,\cdots,p_2,\ -p_1,\cdots,-p_1,\ 0,\cdots,0)\otimes P_b'\}\boldsymbol{\mu}\,.\end{aligned} \tag{3.3}$$

The sum of squares for (3.3) is

$$(3.4)\quad \begin{aligned} &S(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)\\ &=r\|\hat{L}(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)\|^2\\ &=\frac{2p_1p_2}{p_1+p_2}\times\frac{r}{2}\sum_{j=1}^{b}\left\{\frac{1}{p_1}\sum_{i=1}^{p_1}(y_{ij\cdot}-y_{i\cdot\cdot})-\frac{1}{p_2}\sum_{i=p_1+1}^{p_1+p_2}(y_{ij\cdot}-y_{i\cdot\cdot})\right\}^2 . \end{aligned}$$

Since it is also a component of T, it can also be compared to $(a-1)(b-1)\cdot\hat{\sigma}^2F\{(a-1)(b-1),\ ab(r-1);\ \alpha\}$. Sometimes it will reveal an interaction effect between groups even when no interaction element between two rows from different groups is significant. This is because pooling homogeneous rows amounts to increasing the replication number. It should be noted that the sum of squares of (3.4) is $2p_1p_2/(p_1+p_2)$ times the usual sum of squares for interaction in the $2\times b$ table made of averaged observations over p_1 and p_2 rows.

If $S(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)$ is significant and $S(1,\cdots,p_1+p_2;\ p_1+p_2+1,\cdots,p_1+p_2+p_3\,(=a))\}$ is not significant one is still recommended to have both of $\gamma_1=L(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)$ and $\gamma_2=L(1,\cdots,p_1+p_2;\ p_1+p_2+1,\cdots,a)$ remaining in the model, since they together suggest the existence of interaction among the three subgroups, see Example 1. The simple formulas (2.3) and (2.4) can be used only for this model. It should also be noted that, as it is seen from Figure 3, an equivalent model is obtained if we replace γ_1 and γ_2 by $\gamma_3=L(1,\cdots,p_1;\ p_1+p_2+1,\cdots,a)$ and $\gamma_4=L(1,\cdots,p_1,\ p_1+p_2+1,\cdots,a;\ p_1+1,\cdots,p_1+p_2)$ or by $\gamma_5=L(p_1+1,\cdots,p_1+p_2;\ p_1+p_2+1,\cdots,a)$ and $\gamma_6=L(1,\cdots,p_1;\ p_1+1,\cdots,a)$. When no interaction element is revealed to be significant by the procedures of Subsections 3.1 and 3.2 we can proceed as in Subsection 3.3.

3.3. *Adding procedure*

Classify rows into three subgroups noting carefully the elements of the matrix of the squared distances. Suppose they are composed of the first p_1 rows, the subsequent p_2 rows and the last p_3 rows, $\sum p_i=a$. Then compute the statistic

$$\begin{aligned} &S(1,\cdots,p_1;\ p_1+1,\cdots,p_1+p_2)+S(1,\cdots,p_1+p_2;\ p_1+p_2+1,\cdots,a)\\ &\quad=S(1,\cdots,p_1;\ p_1+p_2+1,\cdots,a)\\ &\qquad+S(1,\cdots,p_1,\ p_1+p_2+1,\cdots,a;\ p_1+1,\cdots,p_1+p_2)\\ &\quad=S(p_1+1,\cdots,p_1+p_2;\ p_1+p_2+1,\cdots,a)\\ &\qquad+S(1,\cdots,p_1;\ p_1+1,\cdots,a) \end{aligned}$$

and compare it with $(a-1)(b-1)\hat{\sigma}^2F\{(a-1)(b-1),\ ab(r-1);\ \alpha\}$. If the statistic is larger, the grouping into three subgroups is significant and we obtain the model with γ_1 and γ_2 of Subsection 3.2. Again the prob-

$$
\begin{array}{l|ccccc|ccccc|ccccc}
 & \multicolumn{5}{c|}{\overbrace{\hspace{12em}}^{p_1}} & \multicolumn{5}{c|}{\overbrace{\hspace{12em}}^{p_2}} & \multicolumn{5}{c}{\overbrace{\hspace{12em}}^{p_3}} \\
 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
p_1-1 & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & 1 & 1 & 1 & \cdots & -(p_1-1) & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
 & 0 & 0 & 0 & \cdots & 0 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
p_2-1 & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & 1 & \cdots & -(p_2-1) & 0 & 0 & 0 & \cdots & 0 \\
 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -1 & 0 & \cdots & 0 \\
p_3-1 & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & 1 & \cdots & -(p_3-1) \\
\text{for } \gamma_1 & p_2 & p_2 & p_2 & \cdots & p_2 & -p_1 & -p_1 & -p_1 & \cdots & -p_1 & 0 & 0 & 0 & \cdots & 0 \\
\text{for } \gamma_2 & p_3 & p_3 & p_3 & \cdots & p_3 & p_3 & p_3 & p_3 & \cdots & p_3 & -(p_1+p_2) & -(p_1+p_2) & -(p_1+p_2) & \cdots & -(p_1+p_2) \\
\text{for } \gamma_3 & p_3 & p_3 & p_3 & \cdots & p_3 & 0 & 0 & 0 & \cdots & 0 & -p_1 & -p_1 & -p_1 & \cdots & -p_1 \\
\text{for } \gamma_4 & p_2 & p_2 & p_2 & \cdots & p_2 & -(p_1+p_3) & -(p_1+p_3) & -(p_1+p_3) & \cdots & -(p_1+p_3) & p_2 & p_2 & p_2 & \cdots & p_2 \\
\text{for } \gamma_5 & 0 & 0 & 0 & \cdots & 0 & p_3 & p_3 & p_3 & \cdots & p_3 & -p_2 & -p_2 & -p_2 & \cdots & -p_2 \\
\text{for } \gamma_6 & (p_2+p_3) & (p_2+p_3) & (p_2+p_3) & \cdots & (p_2+p_3) & -p_1 & -p_1 & -p_1 & \cdots & -p_1 & -p_1 & -p_1 & -p_1 & \cdots & -p_1
\end{array}
$$

Figure 3. The matrix P'_a for partitioning rows into three subgroups (normalizing constants are omitted).

ability of judging this to be significant when H_0 is true is less than or equal to α since the statistic is still a component of T. If it is smaller we can try four subgroups, five subgroups and so on until each subgroup is composed of only one row, this obviously giving T. If no significant element is obtained in this process, we can apply the additive model for the whole table. Otherwise we proceed to classify rows and columns simultaneously.

4. Procedure for a two-way layout with exactly one observation per cell

When $r=1$, we cannot apply the previous procedures since we do not have the within cell sum of squares and have no independent estimate of σ^2. It is shown, however, that the sum of squares for the interaction element between any two rows or between any two subgroups is bounded above by the largest latent root of a Wishart matrix.

LEMMA 1. *The maximum value of* $\|\{(a_1,\cdots,a_a)\otimes P_b'\}\boldsymbol{y}\|^2$ *with respect to* a_i *subject to the restrictions that* $\sum a_i=0$, $\sum a_i^2=1$ *is the largest latent root of a Wishart matrix which is distributed as* $W\{\sigma^2 I_{\min(a-1,b-1)}, \max(a-1,b-1)\}$ *under* H_0. *Here* $\boldsymbol{y}$ *is a column vector of* y_{ij}*'s arranged in dictionary order.*

PROOF. It is easily verified that

$$\begin{aligned}\max_{\sum a_i=0,\,\sum a_i^2=1} \|\{(a_1,\cdots,a_a)\otimes \boldsymbol{P}_b'\}\boldsymbol{y}\|^2 &= \max\,(\sum a_i\boldsymbol{z}_i)'(\sum a_i\boldsymbol{z}_i)\\ &= \text{maximum latent root of } (\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)'(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a),\\ &\quad \text{or equivalently maximum latent root of } \sum \boldsymbol{z}_i\boldsymbol{z}_i',\\ &\quad \text{or maximum latent root of } P_a'(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)'(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)P_a,\\ &\quad \text{since } (\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)\boldsymbol{j}_a=0\,,\end{aligned}$$

where $\boldsymbol{z}_i=\boldsymbol{P}_b'(y_{i1}-y_{i\cdot}-y_{\cdot 1}+y_{\cdot\cdot},\cdots,y_{ib}-y_{i\cdot}-y_{\cdot b}+y_{\cdot\cdot})'$. But by Theorem 4.1 of Johnson and Graybill [9] the null distribution of $\sum \boldsymbol{z}_i\boldsymbol{z}_i'$ is $W(\sigma^2\cdot I_{b-1}, a-1)$ when $a\geqq b$. When $b\geqq a$, similar arguments show that $P_a'(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)'(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_a)P_a$ is distributed as $W(\sigma^2 I_{a-1}, b-1)$.

The sum of squares for any interaction element between rows or between subgroups is obtained by specifying the a_i's in $\|\{(a_1,\cdots,a_a)\otimes P_b'\}\boldsymbol{y}\|^2$ subject to the restrictions on the a_i's.

Johnson and Graybill [9] studied the distribution of the ratio of the largest latent root l_1 of the Wishart matrix to its trace which is equal to T and tabulated the approximate $100\alpha\%$ points u_α for some values of α, a and b. They used it in another paper (Johnson and Graybill [8]) to make simultaneous comparisons of all two-by-two interaction contrasts. We can use a similar technique for all interaction elements between any two rows and also between any two subgroups. That is, we have

$$\Pr\left[S \geqq \{u_\alpha/(1-u_\alpha)\}(T-l_1)\,|\,H_0, \text{ for any } S\right] \leqq \alpha\ ,$$

where S stands for the sum of squares for any interaction element. Then the process to find significant elements is carried out similarly to Subsections 3.1 and 3.2. The exact tabulation of u_α is given by Schuurmann, Krishnaiah and Chattopadhay [14]. It should be noted that if we apply the process also to columns the total number of interaction elements compared amounts to

$$\left(\frac{1}{2}\right)\left\{\sum_{a_1=1}^{a-1}\binom{a}{a_1}\sum_{a_2=1}^{a-a_1}\binom{a-a_1}{a_2}+\sum_{b_1=1}^{b-1}\binom{b}{b_1}\sum_{b_2=1}^{b-b_1}\binom{b-b_1}{b_2}\right\}$$
$$=\left(\frac{1}{2}\right)(3^a+3^b)-(2^a+2^b)+1\ , \qquad a, b \geqq 3\ .$$

Even for a moderate value of a or b this will be too large for the Bonferroni inequality to be successful for obtaining the critical value.

When one factor, rows, say, is a classification factor, one may wish to deal with the separate subtables assuming additivity after classifying rows. This can be done in the contexts of Sections 3 and 4 just by leaving the columns unclassified, see Example 2.

5. A treatment of an ordered table

When there is a natural order in the levels of one factor, columns, say, not all interaction contrasts are interesting. For this case Hirotsu [6] showed the desirability of considering contrasts $\boldsymbol{\eta}=(P_a'\otimes P_b^{*\prime})\boldsymbol{\mu}$, where $P_b^{*\prime}$ is a $b-1$ by b matrix, the mth row being given by

$$\{mb(b-m)\}^{-1/2}(b-m,\cdots,b-m,\ -m,\cdots,\ -m)\ .$$

Following this we may reparameterize the μ_{ij} as

$$\boldsymbol{\mu}=\underset{+}{\boldsymbol{\mu}}+(P_a\otimes B^{*-1}P_b^*)\boldsymbol{\eta} \tag{5.1}$$

where $\boldsymbol{B}^*(\boldsymbol{j}_b\ P_b^*)(\boldsymbol{j}_b\ P_b^*)'$. Since $B^{*-1}P_b^*P_b^{*\prime}=I_b-b^{-1}\boldsymbol{j}_b\boldsymbol{j}_b'$ the equation (5.1) is well defined.

The least squares estimator for newly introduced parameters are obtained again by replacing $\boldsymbol{\mu}$ by $\boldsymbol{y}$ in their defining equations, where $\boldsymbol{y}$ is a column vector of $y_{ij\cdot}$. For example, the least squares estimator for $\boldsymbol{\eta}$ is found to be $\hat{\boldsymbol{\eta}}=(P_a'\otimes P_b^{*\prime})y$. The statistic $\|\hat{\boldsymbol{\eta}}\|^2$ is well approximated by a constant times a χ^2 variable, adjusted for the first two cumulants. That is, in case $r \geqq 2$, we have that the probability,

(5.2) $$\Pr[\|\hat{\eta}\|^2 \geqq (a-1)(b-1)\hat{\sigma}^2 F\{\nu, ab(r-1); \alpha\} \mid H_0],$$

is approximately less than or equal to α, where $\nu = (a-1)(b-1)^2/\mathrm{tr}(P_b^{*\prime} \cdot P_b^*)^2$. A better approximation is given in Hirotsu [6] together with the table of $(b-1)^2/\mathrm{tr}(P_b^{*\prime}P_b^*)^2$.

We can define modified interaction elements between two rows and also between two subgroups of rows in the similar way to previous sections and apply procedures in Subsections 3.1, 3.2 and 3.3 referring to the inequality (5.2). For example, the modified interaction element $\boldsymbol{L}^*$ between two subgroups of rows, $(1, \cdots, p_1)$ and $(p_1+1, \cdots, p_1+p_2)$, are defined by (3.3) with P_b replaced by P_b^*. The sum of squares $S^*(1, \cdots, p_1; p_1+1, \cdots, p_1+p_2)$ for $\boldsymbol{L}^*(1, \cdots, p_1; p_1+1, \cdots, p_1+p_2)$ is

$$\frac{rp_1p_2b}{p_1+p_2}\sum_{m=1}^{b-1}\left[\frac{m}{b-m}\left\{\sum_{i=1}^{p_1}\left(\sum_{j=1}^{m} y_{ij\cdot}/m - y_{i\cdot\cdot}\right)\Big/p_1 - \sum_{i=p_1+1}^{p_1+p_2}\left(\sum_{j=1}^{m} y_{ij\cdot}/m - y_{i\cdot\cdot}\right)\Big/p_2\right\}^2\right]$$

and is judged to be significant if it exceeds the value in the right-hand side of the inequality (5.2).

When $a \geqq b$, by similar arguments to those in Section 4, the maximum value of $\|\{(a_1, \cdots, a_a) \otimes P_b^{*\prime}\}\boldsymbol{y}\|^2$ with respect to a_i subject to the restrictions that $\sum a_i = 0$, $\sum a_i^2 = 1$ is bounded above by the largest latent root of a Wishart matrix which is distributed as $W((1/r)\sigma^2 P_b^{*\prime}P_b^*, a-1)$ under H_0. However, the author has not obtained the critical values like the u_α for $W(\sigma^2 I_{b-1}, a-1)$.

6. Examples

Example 1. We apply the procedure in Section 4 for the data in Table 4 of Johnson and Graybill [9]. The total sum of squares for non-additivity is $T = 1044.19$. The latent roots of the relevant Wishart matrix are $l_1 = 1029.05$ and $l_2 = 15.14$. The u_α is 0.9168 for $\alpha = 0.05$ by Table 1 of Johnson and Graybill [9]. Therefore, we compare the sum of squares for any interaction element with $\{u_\alpha/(1-u_\alpha)\}(T-l_1) = 166.79$.

The matrix of squared distances for rows is given in Table 1a. It is rearranged as Table 1b by noting the largest element $S(1;5)$. As mentioned at the end of Subsection 3.1 it is easy to interpret Table 1b. The fifth row is found to behave very differently from the other rows. Hence we extract a vector $\boldsymbol{q}_1 = 42^{-1/2}(1, 1, 1, 1, -6, 1, 1)'$. The sum of squares for it is found to be $S(1, 2, 7, 6, 4, 3; 5) = 777.32$. Although there is no significant element left in Table 1b, any two rows chosen one from each of the subgroups of rows (1, 2, 7, 6) and (4, 3) show somewhat large squared distance. Therefore, try the vector $\boldsymbol{q}_2 = 12^{-1/2}(1, 1, -2, -2, 0, 1, 1)'$ to find $S(1, 2, 7, 6; 4, 3) = 217.74$ which is significantly

Table 1. Matrix of squared distances among rows

a. *Original*

Row number	Row number 1	2	3	4	5	6	7
1	0	10.8	149.6	126.3	730.0	37.4	17.1
2		0	84.4	67.0	574.7	8.1	6.4
3			0	1.0	218.9	42.8	111.1
4				0	249.5	30.8	91.4
5					0	451.5	629.5
6						0	16.4
7							0

b. *Rearranged*

Row number	Row number 1	2	7	6	4	3	5
1	0	10.8	17.1	37.4	126.3	149.6	730.0
2		0	6.4	8.1	67.0	84.4	574.7
7			0	16.4	91.4	111.0	629.5
6				0	30.8	42.8	451.5
4					0	1.0	249.5
3						0	218.5
5							0

large. This together with $S(1, 2, 7, 6, 4, 3; 5)$ explains 95.3% of T.

The squared distances among columns are calculated as $S(1; 2)=70.42$, $S(1; 3)=548.58$ and $S(2; 3)=947.30$. The sum of squares $S(1, 2; 3)$ is found to be $T-70.42=973.77$. It explains 93.3% of T. These results agree very well with what suggested by Figures 1 and 2. Thus we reach the model

$$\boldsymbol{\mu}=\underset{+}{\boldsymbol{\mu}}+\{(\boldsymbol{q}_1, \boldsymbol{q}_2)\otimes 6^{-1/2}(1, 1, -2)'\}(\delta_1, \delta_2)' , \tag{6.1}$$

where

$$(\delta_1, \delta_2)'=\{(\boldsymbol{q}_1, \boldsymbol{q}_2)'\otimes 6^{-1/2}(1, 1, -2)\}\boldsymbol{\mu} . \tag{6.2}$$

The least squares estimates of δ_1 and δ_2 are obtained by replacing $\boldsymbol{\mu}$ by $\boldsymbol{y}$ in (6.2). They give $\hat{\delta}_1^2+\hat{\delta}_2^2=(-27.056)^2+(-14.555)^2=943.87$, which explains 90.4% of T. The departures of the $\hat{\mu}_{ij}$ from $y_{i.}+y_{.j}-y_{..}$ are given by the last term of (6.1) with δ_i's being replaced by their least squares estimates. They can also be obtained directly from the formula (2.3) and given in Table 2, where rows are rearranged.

The variance of $\hat{\mu}_{ij}$ is obtained by the formula (2.4) and an estimate of σ^2 is given by $\tilde{\sigma}^2=(T-943.87)/(12-2)=10.03$.

Table 2. Departures of the $\hat{\mu}_{ij}$ from $y_{i.}+y_{.j}-y_{..}$

Row number	Column number		
	1	2	3
1	−3.42	−3.42	6.84
2	−3.42	−3.42	6.84
7	−3.42	−3.42	6.84
6	−3.42	−3.42	6.84
4	1.73	1.73	−3.45
3	1.73	1.73	−3.45
5	10.23	10.23	−20.45

In classifying rows one may try the vector $\boldsymbol{q}_3=20^{-1/2}(1, 1, 0, 0, -4, 1, 1)'$ to find $S(1, 2, 7, 6; 5)=944.65$ being significantly large and wish to have this element only in the model because $S(1, 2, 7, 6, 5; 4, 3)=50.43$ is far from being significant. However, as stated at the end of Subsection 3.2, the author has the feeling that it should not be done without any intrinsic interest in the vector $\boldsymbol{q}_3$ because it specifies the model too strongly.

For comparisons the biplot of deviations from the over all mean was applied. It shows that two subgroups of row markers (g_1, g_2, g_6, g_7) and (g_3, g_4) are approximately collinear making angles close to 90° with an appoximately fitted line through column markers, which suggests the same result obtained above. Goodness of fit coefficients defined by Bradu and Gabriel [2] for the three columns in the biplot are 0.978, 0.583 and 0.938. They suggest that although in the plane the column markers appear nearly collinear, in three dimensional space they are not. This situation leads one to try a biplot of only two columns. The biplot of the first two columns is then verified to indicate that an additive model might be appropriate for the first two columns.

The next example is taken from Davies [3].

Example 2. Aluminium alloys. Corrosion resistance.

This relates to the testing of nine Aluminium alloys for their resistance to corrosion in a chemical plant atmosphere. Four sites in the factory were chosen, and at each of them a plate made from each alloy was exposed for a year. The plates were then submitted to four observers, who assessed their condition visually and awarded marks to each from 0 to 10 according to the degree of resistance to attack. Thus the data were originally of a $9\times4\times4$ experiment. However, we can treat them as a two-way table, Table 3 here, averaged over the observers since there is no evidence of interaction of observers with sites and alloys. According to Davies the unbiased estimate of the variance

Table 3. Averaged corrosion resistance of aluminium alloys ($r=4$)

Sites	Alloys								
	1	2	3	4	5	6	7	8	9
1	5.50	5.50	5.25	5.00	6.50	5.00	2.25	6.00	7.00
2	8.00	8.00	7.25	7.50	6.00	5.00	5.50	5.75	6.50
3	3.25	3.75	5.00	3.25	4.50	3.00	1.00	5.50	6.25
4	4.25	4.00	6.00	4.75	6.00	4.50	3.75	7.00	6.00
1,3,4	4.33	4.42	5.42	4.33	5.67	4.17	2.33	6.17	6.42
2	8.00	8.00	7.25	7.50	6.00	5.00	5.50	5.75	6.50

to assess the interaction involved in Table 3 is $\hat{\sigma}^2=0.90$ with the degrees of freedom 105. The purpose of the experiment is to choose an appropriate alloy for each of four sites which are considered to be an uncontrollable factor. Then it is preferrable if an alloy is suitable for as many sites as possible since it would be inconvenient to have to use different alloys in different sites in a factory. The squared distances between rows, which should be compared with $24\hat{\sigma}^2F(24, 105; \alpha)=35.26$ ($\alpha=0.05$), 43.04 ($\alpha=0.01$), are as follows: $S(1;2)=38.36^*$, $S(1;3)=8.53$, $S(1;4)=17.69$, $S(2;3)=52.75^{**}$, $S(2;4)=49.61^{**}$, $S(3;4)=11.69$. The site 2 seems to behave very differently from other sites and after eliminating it we can find no significant element. The squared distance between the subgroups is found to be $S(1, 3, 4; 2)=64.04$, which elucidates approximately 71.70% of $T=89.32$. The mean responses of alloys averaged over the sites in each subgroup are shown in the lower half of Table 3. From it we can derive a tentative conclusion that for sites 1, 3 and 4 the alloy 9 or 8 would be suitable and the alloy 1 or 2 for the site 2.

Another example given in Hirotsu [5] is a two-way layout without replication examining the adaptability of 18 varieties of rice to 44 combinations of regions and years. The procedure was applied there by using the percentage of the sum of the sums of squares for orthogonal interaction contrasts to the total sum of squares T since the distributional property of Section 4 was not obtained then. Nevertheless it succeeded in classifying varieties into four types, Formosan type, Indian type, Japanese and Korean type and the special variety called Hybrid. Regions were also classified properly into six groups, Korea and the northern part of Japan, southern part of Japan, tropical regions, Nepal, Egypt and Mexico.

7. Conclusion

Advantages of considering interaction elements here will be:

1. The simultaneous tests must be more powerful than the usual ones for interaction contrasts with one degree of freedom since the critical values used are the same in both tests.
2. The resulting model aids interpretation.

In fact by the adding procedure of Subsection 3.3. the test can have the same power as the usual F test when $r \geqq 2$. However, in case too many subgroups of rows and columns are necessary to define the model, the best thing to do will be to find some appropriate transformation for the data. To this point one should refer to Tukey [15], Box and Cox [1] and Schlesselman [13]. The method here is expected to suggest whether any kind of transformation is necessary or not.

Acknowledgements

The author is greatly indebted to Professors D. R. Cox and K. Takeuchi for their useful suggestions and discussions. His thanks are also due to the referees for their valuable comments.

TOKYO UNIVERSITY

REFERENCES

[1] Box, G. E. P. and Cox, D. R. (1964). Analysis of transformations, *J. R. Statist. Soc.*, B, **26**, 211-243.

[2] Bradu, D. and Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables, *Technometrics*, **20**, 47-68.

[3] Davies, O. L. (1954). *The Design and Analysis of Industrial Experiments*, Oliver and Boyd, London.

[4] Hirotsu, C. (1973). Multiple comparisons in a two-way layout, *Rep. of Statist. Appl. Union of Japanese Scientists and Engineers*, **20**, 1-10.

[5] Hirotsu, C. (1976). *Analysis of Variance*, Kyoiku Shuppan, Tokyo.

[6] Hirotsu, C. (1978). Ordered alternatives for interaction effects, *Biometrika*, **65**, 561-570.

[7] Johnson, D. E. (1976). Some new multiple comparison procedures for the two-way AOV with interaction, *Biometrics*, **32**, 929-934.

[8] Johnson, D. E. and Graybill, F. A. (1972a). Estimation of σ^2 in a two-way classification model with interaction, *J. Amer. Statist. Ass.*, **67**, 388-394.

[9] Johnson, D. E. and Graybill, F. A. (1972b). An analysis of a two-way model with interaction and no replication, *J. Amer. Statist. Ass.*, **67**, 862-868.

[10] Mandel, J. (1964). *The Statistical Analysis of Experimental Data*, Interscience-Wiley, New York.

[11] Mandel, J. (1969). The partitioning of interactions in analysis of variance, *J. Res. Nat. Bur. Stand.*, B, **73**, 309-328.

[12] Mandel, J. (1971). A new analysis of variance model for non additive data, *Technometrics*, **13**, 1-18.

[13] Schlesselman, James J. (1973). Data transformation in two-way analysis of variance, *J. Amer. Statist. Ass.*, **68**, 369-378.

[14] Schuurmann, F. J., Krishnaiah, P. R. and Chattopadhyay, A. K. (1973). On the distribution of the ratios of the extreme roots to the trace of the Wishart matrix, *J. Multivariate Anal.*, **3**, 445-453.

[15] Tukey, J. W. (1949). One degree of freedom for non-additivity, *Biometrics*, **5**, 232-242.