Chapter 7

Two-way layout

The general normal linear model is, as usual

$$y \sim \mathrm{N}(\mu, \sigma^2 I),\ \mu \in \mathcal{E} \subset \Re^n$$

The *two-way layout* or two-factor design, provides a particular case for the estimation space $\mathcal{E}$. It illustrates many of the important issues in fixed effects linear models.

Table 7.1: Sample data for a two-way design.

	Age 1	Age 3	Age 6	Age 9	Age 12
4 ml	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
8 ml	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25

As an example, we will use data from an experiment given by Oehlert (2000), Table 8.1. This experiment had ten treatments, consisting of all combinations of $r = 2$ levels of added water, either 4 ml or 8 ml, and the $c = 5$ ages of barley seeds, either 1, 3, 6, 9 or 12 weeks. Each combination was repeated $m = 3$ times for a total of $n = rcm = 30$ observations. The response is the number of barley seeds out of 100 that germinated (these data could also be treated using binomial

regression models, since the responses may be binomially distributed). The data are given in Table 7.1.

7.1 Equal replications

We begin with the equal replication case as in Table 7.1. The table consists of $t = rc$ *cells*, each corresponding to a treatment combination. It is convenient to write the vector y of responses as having elements with three subscripts $y = (y_{ijk})$. We will write the vector y with the last subscript for replications always changes the fastest, and the first subscript for rows always changes the slowest. Any convention for ordering the elements of y could be adopted, but consistency is important. For example, in the case $r = 2, c = 3, m = 2$, the vector y is written in the order

$$y' = (y_{111}, y_{112}, y_{121}, y_{122}, y_{131}, y_{132}, y_{211}, y_{212}, y_{221}, y_{222}, y_{231}, y_{232}) \tag{7.1}$$

Definition 7.1 (vec notation) *Let A be an $m \times n$ matrix with columns $A = (a_1, \ldots, a_n)$. We will define* $\mathrm{vec}(A)$ *to be the $mn \times 1$ vector obtained by stacking the columns of A,*

$$\mathrm{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Similarly,

$$\mathrm{vec}(A') = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix}$$

where A_j is the jth row of A. We can extend the definition vec to multidimensional arrays. Suppose we use the notation (y_{ijk}) for a three dimensional array with $i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K$. Then

$$\mathrm{vec}((y_{ijk})) = \begin{pmatrix} \mathrm{vec}((y_{ij1})) \\ \mathrm{vec}((y_{ij2})) \\ \vdots \\ \mathrm{vec}((y_{ijm})) \end{pmatrix}$$

By this definition, we see that in (7.1), $y = \text{vec}((y_{ijk})')$.

Theorem 7.1 *Suppose a and b are $n \times 1$ vectors, α and β are scalars, and A, B and C are matrices of appropriate dimension. Then:*

1. $\text{vec}(a) = \text{vec}(a')$.
2. $\text{vec}(ab') = b \otimes a$.
3. $\text{vec}(\alpha A + \beta B) = \alpha\text{vec}(A) + \beta\text{vec}(B)$.
4. $tr(A'B) = (\text{vec}(A))'\text{vec}(B)$.
5. $\text{vec}(ABC) + (C' \otimes A)\text{vec}(B)$.

Proof: Schott (1997, Section 7.5).

Returning to the two-way model, we will write $\theta = (\theta_{ij})$ to be the matrix with r rows and c columns whose (i, j) element is $\theta_{ij} = \text{E}(y_{ijk})$. The jth column of θ is $(\theta_{1j}, \ldots, \theta_{rj})'$, and so unlike (7.1), $\text{vec}(\theta)$ will put the elements of θ in a order with level of row changing the fastest. To get the same order as (7.1), use $\text{vec}(\theta')$. Finally, we will have $\mu = (\text{vec}(\theta') \otimes J_m)$. Similarly, if we define $\bar{y}$ to be an $r \times c$ array with elements $\bar{y}_{ij+}$, then $\text{vec}(\bar{y}')$ is an $rc \times 1$ vector of cell means in the correct order.

Without further structure on the cell means, $\mu \in \mathcal{E}$, where $\mathcal{E} = \mathcal{R}(I_{rc} \otimes J_m)$. The matrix $I_{rc} \otimes J_m$ has $n = rcm$ rows and rc columns, one column for each combination of i and j. The ℓ-th column of this matrix has m ones in the positions for the ℓth treatment combination, and zeroes everywhere else.

As usual, there are many parametric versions of this model, each basis for $\mathcal{E}$ implying a different parameterization. One choice is:

$$\text{E}(y) = (I_{rc} \otimes J_m)\text{vec}(\theta') \tag{7.2}$$

The matrix $(I_{rc} \otimes J_m)$ is an $n \times rc$ matrix, and the parameters are the *cell means*. The scalar version of this model is more familiar,

$$y_{ijk} = \theta_{ij} + \varepsilon_{ijk};\ i = 1, \ldots, r, j = 1, \ldots, c; k = 1, \ldots, m \tag{7.3}$$

Model (7.2) or (7.3) does not make explicit reference to the additional subscript. We can do this by extending the definition of the Kronecker product to more than two matrices,

$$A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C)$$

and then since $I_{rc} = (I_r \otimes I_c)$, we can write (7.2) as

$$\mathrm{E}(y) = (I_r \otimes I_c \otimes J_m)\mathrm{vec}(\theta')$$

We can get other parametric versions of this model by choosing a different basis for $\mathcal{E}$. For example,

$$X_1 = ((J_t, e_2, \ldots, e_{t-1}) \otimes J_m)$$

provides another basis for $\mathcal{E}$, equivalent to the parametric model

$$\begin{aligned} y_{11k} &= \mu_0 + \varepsilon_{ijk} \\ y_{ijk} &= \mu_0 + \alpha_{ij} + \varepsilon_{ijk}, \text{ for } (i,j) \neq (1,1) \end{aligned} \tag{7.4}$$

This is the parameterization that sets the first level of the factor to zero, so the mean for the first cell is μ_0, and the mean for any other cell is $\mu_0 + \alpha_{ij}$.

As practice with Kronecker products, we will compute the projection. We find:

$$\begin{aligned} P_{\mathcal{E}} &= (I_{rc} \otimes J_m)\left\{(I_{rc} \otimes J_m)'(I_{rc} \otimes J_m)\right\}^{-1}(I_{rc} \otimes J_m)' \\ &= (I_{rc} \otimes J_m)\left\{(I_{rc} \otimes J_m')(I_{rc} \otimes J_m)\right\}^{-1}(I_{rc} \otimes J_m') \\ &= (I_{rc} \otimes J_m)\left\{(I_{rc} \otimes J_m' J_m)\right\}^{-1}(I_{rc} \otimes J_m') \\ &= (I_{rc} \otimes J_m)(I_{rc} \otimes 1/m)(I_{rc} \otimes J_m') \\ &= (I_{rc} \otimes J_m J_m'/m) \\ &= (I_{rc} \otimes P_{J_m}) \end{aligned}$$

where P_{J_m} is the projection on J_m. Thus

$$\begin{aligned} \hat{\mu} &= P_{\mathcal{E}} y = (I_{rc} \otimes P_{J_m})y \\ &= \mathrm{vec}(\bar{y}') \otimes J_m \end{aligned}$$

which is a vector that repeats each cell mean m times. Since we can write $\mathrm{vec}(\theta') = (1/m)(1_{rc} \otimes I_m')\mu$, the cell means are estimable, and

$$\widehat{\mathrm{vec}(\theta')} = (1/m)(1_{rc}' \otimes I_m)\hat{\mu} = \mathrm{vec}(\bar{y}')$$

After all this fancy stuff, we get the answer that the estimate of a population cell mean is an observed cell mean.

As usual, $\| (I-P)y \|^2 \sim \sigma^2\chi^2(n-rc)$, and

$$\begin{aligned} \| (I-P)y \|^2 &= \| y - \mathrm{vec}(\bar{y}) \otimes J_m \|^2 \\ &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{ij+})^2 \end{aligned}$$

The estimate of σ^2 is $\hat{\sigma}^2 = \| (I-P)y \|^2/(rcm - rc)$. If $m = 1$, then there is no estimate of σ^2 for this model without making further assumptions. The scalar versions of the estimates are

$$\hat{\mu}_{ij} = \bar{y}_{ij+};\ \mathrm{var}(\hat{\mu}_{ij}) = \sigma^2/m$$

In the barley sprouting example, here is the fit of this model in R. The variable `trts` consists of a label giving the level of Age and the level of Water for that particular treatment, so it defines a factor with ten levels.

```
> # Barley sprouting data from Oehlert, p. 166
> Age <- rep(c(1,3,6,9,12),rep(6,5))
> Water <- rep(rep(c(4,8),c(3,3)),5)
> y <- c(11,9,6,8,3,3,7,16,17,1,7,3,9,19,35,5,9,9,13,35,28,
+        1,10,9,20,37,45,11,15,25)
> # create a data frame
> barley <- data.frame(Age=Age, Water=Water, y=y)
> # display the data in a nice format
> aperm(array(y,c(3,2,5),dimnames=
+   list(paste("R",1:3,sep=""),c("4ml","8ml"),
+        paste("Age",c(1,3,6,9,12),sep=""))),
+   c(1,3,2))
, , 4ml

   Age1 Age3 Age6 Age9 Age12
R1   11    7    9   13    20
R2    9   16   19   35    37
R3    6   17   35   28    45

, , 8ml

   Age1 Age3 Age6 Age9 Age12
R1    8    1    5    1    11
R2    3    7    9   10    15
```

```
R3    3    3    9    9    25

>
> # Create a factor with one level for each of the ten treatments
> levels <- paste("Age", rep(c(1,3,6,9,12), rep(2,5)),
+                 rep(paste("Water",c(4,8),sep=""),5),sep="")
> barley$trts <- factor(rep(1:10, rep(3,10)), ordered=FALSE,
+                 labels=levels)
> # Fit the one-way model
> m0 <- lm(y ~ trts, data = barley)
> summary(m0)

Call:
lm(formula = y ~ trts, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)       8.667      4.462   1.942  0.06632 .
trtsAge1Water8   -4.000      6.311  -0.634  0.53335
trtsAge3Water4    4.667      6.311   0.740  0.46819
trtsAge3Water8   -5.000      6.311  -0.792  0.43747
trtsAge6Water4   12.333      6.311   1.954  0.06478 .
trtsAge6Water8   -1.000      6.311  -0.158  0.87568
trtsAge9Water4   16.667      6.311   2.641  0.01567 *
trtsAge9Water8   -2.000      6.311  -0.317  0.75458
trtsAge12Water4  25.333      6.311   4.014  0.00068 ***
trtsAge12Water8   8.333      6.311   1.321  0.20156
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939,     Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269
>
> anova(m0)
Analysis of Variance Table
```

```
Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
trts       9 2708.13  300.90  5.0375 0.001269 **
Residuals 20 1194.67   59.73
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In model `m0` above, the analysis of variance uses $\mathcal{E}_0 = J_t \otimes J_m$ without consulting the user. The model uses parameterization (7.4). To get parameterization (7.3), we need to delete the intercept from the model (the output is edited for space):

```
> m1 <- update(m0,formula=~.-1)   Reparameterize by deleting intercept
> summary(m1)
Call:
lm(formula = y ~ trts - 1, data = barley)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
trtsAge1Water4      8.667      4.462   1.942 0.066323 .
trtsAge1Water8      4.667      4.462   1.046 0.308118
trtsAge3Water4     13.333      4.462   2.988 0.007268 **
trtsAge3Water8      3.667      4.462   0.822 0.420925
trtsAge6Water4     21.000      4.462   4.706 0.000135 ***
trtsAge6Water8      7.667      4.462   1.718 0.101215
trtsAge9Water4     25.333      4.462   5.677 1.48e-05 ***
trtsAge9Water8      6.667      4.462   1.494 0.150777
trtsAge12Water4    34.000      4.462   7.620 2.45e-07 ***
trtsAge12Water8    17.000      4.462   3.810 0.001097 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared:  0.88,      Adjusted R-squared: 0.8199
F-statistic: 14.66 on 10 and 20 DF,  p-value: 3.921e-07
> anova(m1)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
trts      10 8757.3   875.7  14.661 3.921e-07 ***
```

```
Residuals 20 1194.7    59.7
```

In model `m1`, the subspace used in the analysis of variance is $\mathcal{E}_0 = \{0\}$.

7.2 Main effects and interactions

The additional richness of the design comes from exploiting the structure inherent in using two subscripts to describe each cell mean. The basic idea is to divide $\mathcal{E}$ into four spaces: a *grand mean*, an A-space which depends only on row means; a B-space that depends only on column means, and an AB-space that depends on the remainder after removing other effects. The estimation space is $\mathcal{E} = \mathcal{R}(I_r \otimes I_c \otimes J_m)$ of dimension $t = rc$.

Grand mean

The grand mean space corresponds to the average response in all treatment combinations, $\mathcal{E}_1 = \mathcal{R}(J_r \otimes J_c \otimes J_m)$, a space of dimension one. The projection onto this space is just $P_{\mathcal{E}_1} = (P_r \otimes P_c \otimes P_m)$, and $P_{\mathcal{E}_1} y = \bar{y}_{+++} J_{rcm}$ and $\| P_{\mathcal{E}_1} y \|^2 = rcm(\bar{y}^2_{+++})$:

$$\mathcal{E}_1 = \mathcal{R}(J_r \otimes J_c \otimes J_m) \tag{7.5}$$

A-space for row effects

The A space is for effects that depend on the first subscript of the cell but not on the second subscript, or effectively on the $\bar{y}_{i++}$. The full A space is given by $\mathcal{E}_A = \mathcal{R}(I_r \otimes J_c \otimes J_m)$, since

$$(I_r \otimes J'_c/c \otimes J'_m/m)y = \mathrm{vec}(\bar{y}_{i++})$$

This vector space is of dimension r, *but it includes* $\mathcal{E}_1$. To get the part of the A-space orthogonal to $\mathcal{E}_1$, we can multiply by on the left by any matrix G_r that is a linear transformation from $\Re^r$ to the $r-1$ dimensional subspace of $\Re^r$ that is orthogonal to J_r. One such matrix is $Q_r = I - P_r = I - J_r J'_r/r$, and

$$Q_r(I_r \otimes J'_c/c \otimes J'_m/m)y = Q_r \mathrm{vec}(\bar{y}_{i++}) = \mathrm{vec}(\bar{y}_{i++} - \bar{y}_{+++}) \tag{7.6}$$

which are the "usual" A-effects with the sum of the effects equal to zero. We can write the matrix on the left side of (7.6) as $(Q_r \otimes J'_c/c \otimes J'_m/m)$ and this provides a basis for the part of the estimation space for A that is orthogonal to $\mathcal{E}_1$. We

call this space $\mathcal{E}_{A^*}$. Replacing Q_r by the general G_r amounts to changing the parameterization for the row effects, using, for example, polynomial contrasts, the R parameterization of dropping the first level, the SAS parameterization, and so on. Substituting a general G_r, we find

$$\mathcal{E}_{A^*} = \mathcal{R}(G_r \otimes J_c \otimes J_m) \tag{7.7}$$

B-space for column effects

The development for the B or column space is the same as for rows, and we find $\mathcal{E}_B = \mathcal{R}(J_r \otimes I_c \otimes J_m)$ and the part of this space orthogonal to $\mathcal{E}_1$ is $\mathcal{E}_{B^*} = \mathcal{R}(J_r \otimes Q_c \otimes J_m)$. We can replace Q_c by any matrix G_c that spans the same space of Q_c, so for example instead of using the effects parameterization implied by Q_c, we can use polynomials, leave-out-a-level, or any other parameterization we choose. Projections will be the same, and so tests will be the same. The general form is

$$\mathcal{E}_{B^*} = \mathcal{R}(J_r \otimes G_c \otimes J_m) \tag{7.8}$$

AB-space for joint or interaction effects

The space $\mathcal{E}_{AB} = \mathcal{R}(I_r \otimes I_c \otimes J_m)$ is identical to $\mathcal{E}$, so we need further refinement. All of $\mathcal{E}_1, \mathcal{E}_A$ and $\mathcal{E}_B$ are subspaces, and so the unique part of $\mathcal{E}_{AB}$ is orthogonal to all three of these. It is given by $\mathcal{E}_{AB^*} = \mathcal{R}(Q_r \otimes Q_c \otimes J_m)$, and has dimension $(r-1)(c-1)$.

The interactions correspond to the effects due to the row i and the column j of the table, beyond the sum of the effects of the ith row by itself and the jth column by itself. For example, if the row variable is variety of a crop and the column variable is hours of daylight, suppose that one if the crops completely fails of the hours of daylight is too short, but the other varieties do not fail under these conditions. This would imply an interaction, since the effect of the column variable depends on the value of the row variable.

There are many useful ways to parameterize interactions. These are obtained by replacing Q_r and Q_c in the definition of the basis for $\mathcal{E}_{AB^*}$ by other matrices that span the same space that these matrices span. If we use G_r and G_c in place of Q_r and Q_c, we write the space as

$$\mathcal{E}_{AB^*} = \mathcal{R}(G_r \otimes G_c \otimes J_m) \tag{7.9}$$

7.2.1 Estimating parameters

As usual, parameters make interpretation and understanding of linear models more difficult. There is a special case: If the matrices G_r and G_c are orthogonal to the J_r and J_c respectively, then the parameters can be defined from the definitions of the subspaces (7.5)–(7.9). The parameter estimates are given by:

- Grand mean, averaging all rc treatment combinations,

$$\begin{aligned} \bar{\theta}_{++} &= (J_r'/r \otimes J_c'/c \otimes J_m'/m)y \\ &= A_1'y \end{aligned} \tag{7.10}$$

- A or row means, averaging over columns,

$$\begin{aligned} A \text{ or row parameters} &= (G_r \otimes J_c'/c \otimes J_m'/m)y \\ &= A_A'y \end{aligned} \tag{7.11}$$

  The choice of G_r determines how the parameters are defined. If $G_r = Q_r$ we get the effects parameterization, but otherwise we get different parameters.

- B or column means, averaging over rows,

$$\begin{aligned} B \text{ or column parameters} &= (J_r'/r \otimes G_c \otimes J_m'/m)y \\ &= A_B'y \end{aligned} \tag{7.12}$$

- AB or interactions,

$$\begin{aligned} \text{Interaction parameters} &= (G_r \otimes G_c \otimes J_m'/m)y \\ &= A_{AB}'y \end{aligned} \tag{7.13}$$

When either or both of G_r and G_c are not orthogonal to the column of ones, then these formulas won't work because the subspaces are not orthogonal and the estimates of parameters corresponding to one subspace must be adjusted for the other subspaces. The parameterizations used by R and by SAS are *not* orthogonal parameterizations, and so problems will arise here if you try to compute a formula for parameter estimates. Computer programs have no problem with this, since they: (1) have a matrix X whose columns span $\mathcal{E}$; (2) compute $X = QR$ using column pivoting, thereby excluding columns, from left to right, that are linear

combinations of the previous columns; and then estimate parameters by backsolving $R\hat{\beta} = Q'y$.

For the R parameterization in the balanced two way model, we can start with the cell-means approach with parameters $\mathrm{E}(y_{ijk}) = \theta_{ij}$, so that $\hat{\theta_{ij}} = \bar{y}_{ij+}$. We then have

$$\theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

where by the parameterization $\alpha_1 = \beta_1 = \gamma_{i1} = \gamma_{1j} = 0$, or in particular

$$\begin{aligned}
\theta_{11} &= \mu \\
\theta_{i1} &= \mu + \alpha_i, i > 1 \\
\theta_{1j} &= \mu + \beta_j, j > 1 \\
\theta_{ij} &= \mu + \alpha_i + \beta_j + \gamma_{ij}, i > 1, j > 1
\end{aligned}$$

Substituting estimates for parameters, and solving for the parameters, we find

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{11+} \\
\hat{\alpha}_i &= \bar{y}_{i1+} - \bar{y}_{11+}, i > 1 \\
\hat{\beta}_j &= \bar{y}_{1j+} - \bar{y}_{11+}, j > 1 \\
\hat{\gamma}_{ij} &= \bar{y}_{ij+} - \bar{y}_{i1+} - \bar{y}_{1j+} + \bar{y}_{11+}
\end{aligned}$$

The estimates in the SAS parameterization are similar, except that the subscript "1" must be replace by either r for rows or c for columns.

7.3 Variances

We will derive the variances for the parameters, assuming that $G_r = Q_r$, $G_c = Q_c$, and the model is written as

$$y_{ijk} = \theta + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

For the αs, the A main effects, from (7.11),

$$\begin{aligned}
\mathrm{Var}(\mathrm{vec}(\hat{\alpha})) &= \mathrm{Var}({A_A}'y) \\
&= {A_A}'\mathrm{Var}(y)A_A \\
&= \sigma^2 {A_\alpha}' A_\alpha
\end{aligned}$$

$$\begin{aligned} &= \sigma^2 \frac{1}{c^2 m^2} (Q_r \otimes J_c' \otimes J_m')(Q_r \otimes J_c \otimes J_m) \\ &= \frac{\sigma^2}{mc^2} (Q_r \otimes J_c' J_c) \\ &= \frac{\sigma^2}{cm} Q_r \end{aligned}$$

and the covariance between typical elements of $\hat{\alpha}$ are

$$\text{cov}(\hat{\alpha}_i, \hat{\alpha}_j) = \frac{\sigma^2}{rcm} \begin{cases} (r-1) & i = j \\ -1 & i \neq j \end{cases}$$

A similar computation for B or column effects gives

$$\text{Var}(\text{vec}(\hat{\beta})) = \frac{\sigma^2}{rm} Q_c$$

with typical elements

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \frac{\sigma^2}{rcm} \begin{cases} (c-1) & i = j \\ -1 & i \neq j \end{cases}$$

For interactions, typical elements of the covariance matrix are given by

$$\text{cov}(\hat{\gamma}_{ij}, \hat{\gamma}_{i'j'}) = \frac{\sigma^2}{rcm} \begin{cases} (r-1)(c-1) & i = i', j = j' \\ -(r-1) & i = i', j \neq j' \\ -(c-1) & i \neq i', j = j' \\ 1 & i \neq i', j \neq j' \end{cases}$$

Recall that if $Y \sim \text{N}(\mu, \Sigma)$, then $A'Y$ and $B'Y$ are independent if and only if $A'\Sigma B = 0$. Using this result, it is easy to verify that row effects, column effects and interactions are independent, since, by direct multiplication, ${A_A}' A_B = {A_A}' A_{AB} = {A_B}' A_{AB} = 0$.

7.4 Reduction to cell means

The data $y = (y_{ijk})$ can be reduced to *cell means*, and replaced by $\text{vec}(\bar{y}')$, which is a $t = rc \times 1$ vector. Assuming m_{ij} observations in cell (i, j), we can write

$$\text{vec}(\bar{y}') \sim \text{N}(\text{vec}(\theta'), \sigma^2 \text{diag}(1/m_{ij})) \tag{7.14}$$

and we can base the linear model on the cell means rather than the original observations. What is lost in this reduction is the error space, since $\mathrm{vec}(\bar{y}')$ is effectively the projection on the estimation space. We can always compute the residual sum of squares as

$$RSS = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m_{ij}}(y_{ijk} - \bar{y}_{ij+})^2$$

and $RSS \sim \chi^2(\nu)$ with $\nu = \sum\sum m_{ij} - rc$. Thus a model-free estimate of σ^2 is $\hat{\sigma}^2 = RSS/\nu$. We this of this as a *pure error* estimate because it depends on the largest possible model that could be fit, and will be independent of any function of $P_{\mathcal{E}}y$, or of $P_{\mathcal{E}_0}y$ if $\mathcal{E}_0 \subset \mathcal{E}$.

For the balanced two-way design, (7.14) holds, but with the covariance matrix given by $(\sigma^2/m)I$. For the cell means model, we leave it as an exercise to show that the projections on the four subspaces are:

$$\begin{aligned} P_1 &= (P_r \otimes P_c) \\ P_{A^*} &= (Q_r \otimes P_c) \\ P_{B^*} &= (P_r \otimes Q_c) \\ P_{AB^*} &= (Q_r \otimes Q_c) \end{aligned} \tag{7.15}$$

Similarly, the matrices $A_A, A_B, and A_{AB}$ defined by (7.11)–(7.13) can be defined for the cell means model.

If the m_{ij} are not all equal, the problem becomes more complex because the covariance matrix in (7.14) is no longer diagonal, and so the inner product now depends on the m_{ij}, and the four spaces defined by (7.15) need not be orthogonal. We return to this in Section 7.6.

7.5 Hypothesis testing

In the cell means model with m observations per cell mean, $Y \sim \mathrm{N}(\mu, (\sigma^2/m)I)$, consider a test of NH: $A'\mu = 0$ versus AH: $A'\mu \neq 0$. Given that we have an independent estimator $\nu\hat{\sigma}^2$ which has a $\sigma^2\chi^2(\nu)$ distribution, and assuming that $\mathcal{R}(A) \subset \mathcal{E}$ (where $\mu \in \mathcal{E}$), the F test of the null hypothesis is given by

$$F = \frac{m\| P_A Y \|^2}{\rho(A)\hat{\sigma}^2} \sim F(\rho(A), \nu, m\| P_A\mu \|^2/\sigma^2) \tag{7.16}$$

This differs from previous formulas for F tests only in the inclusion of the m in the numerator to account for using cell means, and the assumption that $\mathcal{R}(A) \subset \mathcal{E}$. If

Table 7.2: Analysis of Variance for the balanced two-way model

Source	*Space*	*df*	*SS*	*E(MS)*
Mean	A_μ	1	$rcm\bar{y}_{+++}^2$	$\sigma^2 + rcm\bar{\mu}_{+++}^2$
A	A_α	$r-1$	$mc\| \hat{\alpha} \|^2$	$\sigma^2 + mc\| \alpha \|^2/(r-1)$
B	A_B	$c-1$	$mr\| \hat{\beta} \|^2$	$\sigma^2 + mr\| \beta \|^2/(c-1)$
AB	A_{AB}	$(r-1)(c-1)$	$m\| C(\hat{\gamma}) \|^2$	$\sigma^2 + m\| C(\gamma) \|^2/(c-1)(r-1)$
Error	$\mathcal{E}^\perp$	$rc(m-1)$	$rc(m-1)\hat{\sigma}^2$	σ^2

this latter were not so, just replace P_A by $P_{P_{\mathcal{E}}A}$. We use this simplification because all the A matrices for the effects have column spaces that are subsets of $\Re^{rc}$, which is all we need.

For the A or row effects, we want to test $\alpha_1 = \cdots = \alpha_r = 0$, which is the same as testing $A_A{}'\text{vec}(\theta') = 0$. For this null hypothesis, we can use any parameterization for G_r we choose, so we use $G_r = Q_r$. Redefining A_A for the cell-means model, the numerator of the F statistic (7.16) is just

$$\begin{aligned} m\| P_{A_A}\text{vec}(\bar{y}) \|^2 &= m\text{vec}(\bar{y})' P_A \text{vec}(\bar{y}) \\ &= m\text{vec}(\bar{y}) A_A (A_A{}' A_A)^{-1} A_A{}' \text{vec}(\bar{y}) \\ &= cm\hat{\alpha}' Q_r \hat{\alpha} \\ &= cm\| \hat{\alpha} \|^2 \\ &= cm \sum_{i=1}^{r} (\bar{y}_{i++} - \bar{y}_{+++})^2 \end{aligned} \tag{7.17}$$

For other effects, can similarly get the following results:

- B or column effects:

$$m\| P_{A_B}\text{vec}(\bar{y}) \|^2 = rm\| \hat{\beta} \|^2 = rn \sum_{j=1}^{c} (\bar{y}_{+j+} - \bar{y}_{+++})^2$$

- Interaction effects:

$$m\| P_{A_{AB}}\text{vec}(\bar{y}) \|^2 = cm\| C(\hat{\gamma}) \|^2 = n \sum_{i=1}^{r} \sum_{j=1}^{c} (\bar{y}_{ij+} - \bar{y}_{i++} - \bar{y}_{+j+} + \bar{y}_{+++})^2$$

- Mean effect:

$$m\| \ P_{A_\mu}\text{vec}(\bar{y}) \ \|^2 = rcm\hat{\mu}^2 = rcm(\bar{y}_{+++})^2$$

All these results can be summarized in an Analysis of Variance as in Table 7.2.

Here is the result of fitting this model using R. Although the results above are given in terms of the effects parameterization, R does *not* use this parameterization, but rather it uses the parameterization obtained by deleting the first level of each factor and interaction.

```
> # show the type contrasts used for the factor
> getOption("contrasts")
        unordered           ordered
"contr.treatment"      "contr.poly"
>
> # Effects contrasts:
> barley$Eage <- factor(barley$Age, ordered=FALSE)
> barley$Eage <- C(barley$Eage,contr.sum(c(1,3,6,9,12)),5)
> contrasts(barley$Eage)
   [,1] [,2] [,3] [,4]
1     1    0    0    0
3     0    1    0    0
6     0    0    1    0
9     0    0    0    1
12   -1   -1   -1   -1
> # Drop first level contrasts, the default
> barley$Fage <- factor(barley$Age, ordered=FALSE)
> contrasts(barley$Fage)
   3 6 9 12
1  0 0 0  0
3  1 0 0  0
6  0 1 0  0
9  0 0 1  0
12 0 0 0  1
> # Polynomial contrasts
> barley$Oage <- factor(barley$Age, ordered=TRUE)
> contrasts(barley$Oage)
               .L         .Q            .C         ^4
1  -6.324555e-01  0.5345225 -3.162278e-01  0.1195229
3  -3.162278e-01 -0.2672612  6.324555e-01 -0.4780914
6  -3.287978e-17 -0.5345225  6.329957e-17  0.7171372
```

```
9   3.162278e-01 -0.2672612 -6.324555e-01 -0.4780914
12  6.324555e-01  0.5345225  3.162278e-01  0.1195229
> # do the same for the water factor
> barley$Fwater <- as.factor(barley$Water) # the default
> barley$Ewater <- C(barley$Fwater,contr.sum(c(4,8)),1) # the default
>
> # the main effects + interaction model
> e1 <- lm(y ~ Eage*Ewater, data = barley)
> summary(e1)

Call:
lm(formula = y ~ Eage * Ewater, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.2000     1.4111  10.063 2.84e-09 ***
Eage1          -7.5333     2.8221  -2.669 0.014731 *
Eage2          -5.7000     2.8221  -2.020 0.057011 .
Eage3           0.1333     2.8221   0.047 0.962786
Eage4           1.8000     2.8221   0.638 0.530829
Ewater1         6.2667     1.4111   4.441 0.000251 ***
Eage1:Ewater1  -4.2667     2.8221  -1.512 0.146211
Eage2:Ewater1  -1.4333     2.8221  -0.508 0.617086
Eage3:Ewater1   0.4000     2.8221   0.142 0.888706
Eage4:Ewater1   3.0667     2.8221   1.087 0.290116
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939, Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269

> anova(e1)
Analysis of Variance Table

Response: y
```

```
            Df  Sum Sq Mean Sq F value   Pr(>F)
Eage         4 1321.13  330.28  5.5293 0.003645 **
Ewater       1 1178.13 1178.13 19.7232 0.000251 ***
Eage:Ewater  4  208.87   52.22  0.8742 0.496726
Residuals   20 1194.67   59.73
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> m1 <- lm(y ~ Fage*Fwater, data = barley)
> summary(m1)

Call:
lm(formula = y ~ Fage * Fwater, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)       8.667      4.462   1.942  0.06632 .
Fage3             4.667      6.310   0.740  0.46819
Fage6            12.333      6.310   1.954  0.06478 .
Fage9            16.667      6.310   2.641  0.01567 *
Fage12           25.333      6.310   4.014  0.00068 ***
Fwater8          -4.000      6.310  -0.634  0.53335
Fage3:Fwater8    -5.667      8.924  -0.635  0.53265
Fage6:Fwater8    -9.333      8.924  -1.046  0.30812
Fage9:Fwater8   -14.667      8.924  -1.643  0.11592
Fage12:Fwater8  -13.000      8.924  -1.457  0.16072
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939, Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269

> anova(m1)
Analysis of Variance Table

Response: y
```

```
            Df  Sum Sq Mean Sq F value   Pr(>F)
Fage         4 1321.13  330.28  5.5293 0.003645 **
Fwater       1 1178.13 1178.13 19.7232 0.000251 ***
Fage:Fwater  4  208.87   52.22  0.8742 0.496726
Residuals   20 1194.67   59.73
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> o1 <- lm(y ~ Oage*Fwater, data = barley)
> summary(o1)

Call:
lm(formula = y ~ Oage * Fwater, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     20.4667     1.9956  10.256 2.06e-09 ***
Oage.L          19.8169     4.4622   4.441 0.000251 ***
Oage.Q           1.2472     4.4622   0.280 0.782725
Oage.C           0.4216     4.4622   0.094 0.925659
Oage^4           1.6733     4.4622   0.375 0.711608
Fwater8        -12.5333     2.8221  -4.441 0.000251 ***
Oage.L:Fwater8 -11.0680     6.3105  -1.754 0.094760 .
Oage.Q:Fwater8   3.4744     6.3105   0.551 0.588023
Oage.C:Fwater8   1.5811     6.3105   0.251 0.804712
Oage^4:Fwater8   1.4741     6.3105   0.234 0.817672
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939, Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269

> anova(o1)
Analysis of Variance Table

Response: y
```

```
            Df  Sum Sq Mean Sq F value   Pr(>F)
Oage         4 1321.13  330.28  5.5293 0.003645 **
Fwater       1 1178.13 1178.13 19.7232 0.000251 ***
Oage:Fwater  4  208.87   52.22  0.8742 0.496726
Residuals   20 1194.67   59.73
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7.6 Unbalanced data

We first return to the one-way model with t groups and m_i observations in the i-th group, $n = \sum m_i$, with the model

$$y_{ij} = \theta_i + e_{ij}$$

Let $\mathcal{E}_0 = \mathcal{R}(J_n)$, and so $\| P_{\mathcal{E}_0} y \|^2 = y' J_n J_n' y = (\sum\sum y_{ij})^2/n = n\bar{y}_{++}^2$, the same answer as in the balanced case. Define $D(z)$ to be a diagonal matrix whose diagonal elements are the elements of z, and let $D(J_{m_i})$ be a matrix with "diagonals" J_{m_i}; if all the m_i are equal, then $D(J_{m_i}) = I_t \otimes J_m$. Then the most general estimation space is $\mathcal{E} = \mathcal{R}(D(J_{m_i}))$, and since the columns of $D(J_{m_i})$ form an orthogonal basis for $\mathcal{E}$, if $(u_1, \ldots, u_t)$ are the columns of $D(J_{m_i})$,

$$P_{\mathcal{E}} y = \sum \frac{(u_i, y)}{(u_i, u_i)} u_i = \sum \bar{y}_{i+} u_i = \begin{pmatrix} \bar{y}_{1+} J_{m_1} \\ \vdots \\ \bar{y}_{t+} J_{m_t} \end{pmatrix}$$

and $\| P_{\mathcal{E}} y \|^2 = \sum m_i \bar{y}_{i+}^2$. The usual F-test is then

$$f = \frac{\sum m_i \bar{y}_{i+}^2 - n\bar{y}_{++}^2}{(t-1)\hat{\sigma}^2} \sim F(t-1, n-t, \delta^2)$$

with $\hat{\sigma}^2 = \sum\sum (y_{ij} - \bar{y}_{i+})^2/(n-t)$. Although all this looks like what we have done before, we get something new for the noncentrality parameter,

$$\delta^2 = \sum m_i (\theta_i - \bar{\theta})^2/\sigma^2$$

Since the m_i appear inside the sum, the *noncentrality parameter, and hence the parametric hypothesis tested, depends on the sample sizes*. More weight is given

to means based on larger sample sizes. This is probably undesirable, but unavoidable with unbalanced data. This hardly invalidates the test, however, although power is surely effected.

In the two-way problem, we have

$$y_{ijk} = \theta_{ij} + \varepsilon_{ijk}$$

with $i = 1, \ldots, r; j = 1, \ldots, c$, and $k = 1, \ldots, m_{ij}$, where for now all the $m_{ij} > 0$ and $n = \sum\sum m_{ij}$. Unbalance is much more challenging because it compromises the standard interpretation of the four subspaces of interest. We temporarily reduce by sufficiency, to simplify the notation. The problem of interest is now

$$\mathrm{vec}(\bar{y}') \sim \mathrm{N}(\mathrm{vec}(\mu), \sigma^2 D(m_{ij}^{-1}))$$

and $\hat{\sigma}^2 = \sum\sum\sum(y_{ijk} - \bar{y}_{ij+})^2/(n - rc)$, which is independent of $\mathrm{vec}(\bar{y}')$.

In the balanced case the two-way model led to examination of four subspaces, for the grand mean, $U_0 = P_r \otimes P_c$, for rows, $U_1 = Q_r \otimes P_c$, for columns, $U_2 = P_r \otimes Q_c$, and for interaction, $U_3 = Q_r \otimes Qc$. All these U matrices are orthogonal projections ($U_i^2 = U_i = U_i'$), orthogonal to each other ($U_i U_j = 0$ if $i \neq j$), and they span interesting and interpretable subspaces. When the m_{ij} are not equal, however, the inner product to use is $(a, b)_{D_{m_{ij}}}$, and relative to this inner product the U_j are neither orthogonal projections nor are they orthogonal to each other. Unequal sample sizes has nothing particular to do with the hypotheses that we find interesting, but they do change our ability to study those hypotheses. Any analysis of unbalanced data is therefore a compromise.

7.6.1 Wilkinson-Rogers notation

Return to the original two-way model, not the reduction based on sufficiency to cell means. A useful new notation is called *reduction notation*. As before, we start with the Wilkinson-Rogers notation for the two-way model with interaction,

$$y \sim 1 + A + B + A.B \ldots \textit{In R, use A:B}$$

and let $\mathcal{E}$ be the vector space implied by this model. We then define.

$$\textit{Red}(1, A, B, A.B) = \| y \|^2 - \| P_{\mathcal{R}(1+A+B+A.B)} y \|^2 \tag{7.18}$$

which is the reduction in sum of squares or the *residual sum of squares for the space $\mathcal{E}$ defined by the Wilkinson-Rogers model*. There is no ambiguity about this

sum of squares. It has a central χ^2 distribution if $\mu \in \mathcal{E}$, with df determined by the dimension of $\mathcal{E}$ and a non-central χ^2 distribution if $\mu \notin \mathcal{E}$.

For any other Wilkinson-Rogers model, say $y \sim 1+A+B$, we would similarly define the corresponding reduction sum of squares by

$$\textit{Red}(1, A, B) = \| y \|^2 - \| P_{\mathcal{R}(1+A+B)} y \|^2 \tag{7.19}$$

For comparison of any Wilkinson-Rogers model with terms (θ_1, θ_2) to a smaller model with terms θ_1, the reduction sum of squares is defined by

$$\textit{Red}(\theta_2|\theta_1) = \textit{Red}(\theta_1, \theta_2) - \textit{Red}(\theta_1)$$

so, for example, $\textit{Red}(A.B|1, A, B) = \textit{Red}(1, A, B, A.B) - \textit{Red}(1, A, B) > 0$ is the additional sum of squares explained by interactions after main effects are fit. An important characteristic of this notation is that *we implicitly define the space corresponding to A.B is the orthogonal complement of the range space of* $1+A+B$ *relative to the space* $1 + A + B + A.B$, so any hypotheses tested will depend on sample sizes.

7.6.2 Marginality principle

The *marginality principle* asserts that an Wilkinson-Rogers model that includes a term like $A.B$ must include all possible lower-order terms with these letters, so if $A.B$ is included, then in the two-way crossed design so is A and B. According to this paradigm, the test for row effects should be of NH:$y\ 1 + B$ versus AH : $y \sim 1 + A + B$. For column effects, NH:$y \sim 1 + A$ versus AH : $y \sim 1 + A + B$, and for interactions, NH:$y \sim 1 + A + B$ versus AH : $y \sim 1 + A + B + A.B$. For each, we use the same estimate of error. This leads to the analysis of variance table shown in Table 7.3. These tests are often called *SAS Type II*. The recommended procedure is to start at the bottom of the table, testing for an interaction first. Only if interactions are negligible do tests of main effects make any sense.

7.6.3 Two other approaches

The most common method of computing tests that is inconsistent with the approach to linear models that we have discussed so far is usually called *SAS Type III*. It corresponds to testing using the sums of squares shown in Table 7.4. Although the parametric hypotheses tested by SAS Type III seem sensible, *these violate the marginality principle* and therefore do not correspond to what we have

Table 7.3: Sum of squares and parameteric hypotheses tested for SAS Type II. *These tests obey the marginality principle.*

Source	Sum of Squares
Rows	$\mathit{Red}(A\|1, B)$
Columns	$\mathit{Red}(B\|1, A)$
Interaction	$\mathit{Red}(A.B\|1, A, B)$

Table 7.4: Numerator sum of squares and parameteric hypotheses tested for SAS Type III. *These tests violate the marginality principle.*

Source	SS	Hypothesis tested
Rows	$\mathit{Red}(A\|1, B, A.B)$	$\bar{\theta}_{i+}$ are all equal
Columns	$\mathit{Red}(B\|1, A, A.B)$	$\bar{\theta}_{+j}$ are all equal
Interaction	$\mathit{Red}(A.B\|1, A, B)$	$\mu_{ij} - \bar{\theta}_{i+} - \bar{\theta}_{+j} + \bar{\theta}_{++}$ are all zero

done so far in this course. For example, testing for A effects adjusting for AB interactions doesn't seem reasonable by what a scientist might want to mean by A and AB effects. In this analysis, each of the spaces is adjusted for all the others, and so the sums of squares do not add up. Even main effects are adjusted for interactions, which might seem counter-intuitive.

The final method is called the *sequential* or *regression* approach, in which the subspaces are ordered and sequentially orthogonalized, as in Table 7.5. For this method, the sums of squares do add to the total SS, but the hypotheses tested are quite unusual. Rows and columns are treated asymmetrically. The sequential approach is the default in R and in many other programs. In SAS, this is called *Type I*. The third approach, called SAS type II, is the same as the sequential approach, except that all main effects are adjusted for all main effects, all two-factor effects for other two-factor effects, and so on. In Table 7.5, this would amount to replacing $\mathit{Red}(\alpha|\mu)$ by $\mathit{Red}(\alpha|\mu, \beta)$.

Here is a small unbalanced example, with the following data:

Table 7.5: Sum of squares and non-centrality parameters for sequential fitting, SAS Type I.

Source	N	Hypothesis tested
Rows	$\mathit{Red}(\alpha\|\mu)$	$\sum_j (m_{ij}\mu_{ij}/m_{i+})$ all equal
Columns	$\mathit{Red}(\beta\|\mu,\alpha)$	$\sum_i m_{ij}\mu_{ij} = \sum_i \sum_k m_{ij}m_{jk}\mu_{ik}/m_{i+}$
Interaction	$\mathit{Red}(\gamma\|\mu,\alpha,\beta)$	$\mu_{ij} - \bar{\mu}_{i+} - \bar{\mu}_{+j} + \bar{\mu}_{++}$ are all zero

	Variety		
Soil	1	2	3
1	6, 10, 11	13, 15	14, 22
2	12, 15, 19, 18	31	18, 9, 12

```
> #Table 4.1 from Searle, Unbalanced data
> soil <- rep(c(1,2),c(7,8))
> var  <- c(rep(c(1,2,3),c(3,2,2)),rep(c(1,2,3),c(4,1,3)))
> days <- c(6,10,11,13,15,14,22,12,15,19,18,31,18,9,12)
> d <- data.frame(days=days,soil=factor(soil,ordered=FALSE),
+                          var=factor(var,ordered=FALSE))
>
> anova(m1 <- aov(days~var*soil,d))
Analysis of Variance Table

Response: days
          Df Sum Sq Mean Sq F value Pr(>F)
var        2   93.3    46.7    3.50 0.0751
soil       1   83.9    83.9    6.29 0.0334
var:soil   2  222.8   111.4    8.35 0.0089
Residuals  9  120.0    13.3
> anova(m2 <- aov(days~soil*var,d))
Analysis of Variance Table

Response: days
          Df Sum Sq Mean Sq F value Pr(>F)
soil       1   52.5    52.5    3.94 0.0785
var        2  124.7    62.4    4.68 0.0405
soil:var   2  222.8   111.4    8.35 0.0089
Residuals  9  120.0    13.3
```

```
> drop1(m1,scope=~var*soil,test="F")
Single term deletions

Model:
days ~ var * soil
         Df Sum of Sq RSS AIC F value  Pr(F)
<none>                120  43
var       2       100 220  48    3.75 0.0654
soil      1        84 204  49    6.30 0.0333
var:soil  2       223 343  55    8.35 0.0089
```

```
title 'Searle p 79';
proc glm data = work.searle;
 class soil var;
 model days = soil var soil*var;
run;

Dependent Variable: days
                             Sum of
Source           DF         Squares  Mean Square  F Value  Pr > F
Model             5    400.0000000   80.0000000     6.00  0.0103
Error             9    120.0000000   13.3333333
Corrected Total  14    520.0000000

  R-Square     Coeff Var      Root MSE     days Mean
  0.769231      24.34322      3.651484      15.00000

Source     DF       Type I SS     Mean Square    F Value    Pr > F

soil        1      52.5000000      52.5000000       3.94    0.0785
var         2     124.7340426      62.3670213       4.68    0.0405
soil*var    2     222.7659574     111.3829787       8.35    0.0089


Source     DF     Type III SS     Mean Square    F Value    Pr > F

soil        1     123.7714286     123.7714286       9.28    0.0139
var         2     192.1276596      96.0638298       7.20    0.0135
soil*var    2     222.7659574     111.3829787       8.35    0.0089
```

```
proc glm data = work.searle;
 class soil var;
 model days = soil var soil*var/e e1 e2 e3;
run;
```

Source	DF	Type II SS	Mean Square	F Value	Pr > F
soil	1	83.9007092	83.9007092	6.29	0.0334
var	2	124.7340426	62.3670213	4.68	0.0405
soil*var	2	222.7659574	111.3829787	8.35	0.0089

7.6.4 Empty cells

7.7 Homework Problems

Specific assignments will be made in class.

1. Find the projection matrices that map onto the following spaces:
   (a) Space generated by the row means (as opposed to the row effects).
   (b) Space generated by the column means.
   (c) Space generated by the simple row effects.
   (d) Space generated by the simple column effects.
2. Let $\hat{\delta}$ denote the $2rc \times 1$ vector of estimated simple row effects followed by estimated simple column effects. Find Var$(\hat{\delta})$.
3. Verify the expected mean squares in Table 7.2.
4. Suppose that $r = 4$, $c = 5$, and primary interest is in the test of all B-or column effects equal versus the alternative that they are all different. Suppose it is believed that for the five levels of the B effects, $\theta_{+i} = \mu_0 + \eta^2 \times i, i = 1, \ldots, 5$. The hypothesis of no difference corresponds to $\eta = 0$, and as η gets larger, then the differences between the means gets larger. Assuming $\sigma^2 = 1$, for $\eta^2 = 0, 1, 2$, draw the power functions for the test that all B effects are equal against the general alternative, as a function of m.

5. The following table gives the results of an experiment to investigate the effects of a control (C) and two inorganic fertilizers (F_1 and F_2) on the yields of 3 varieties of rice. There were two replicates of each of the nine treatments.

Variety	C	F_1	F_2
A	106	168	112
	94	146	128
B	97	61	112
	86	88	91
C	62	125	134
	60	106	112

   Construct the Analysis of Variance table for these data and carry out the F-tests for main effects and interactions. Draw an interaction graph, and summarize your findings.

6. Consider an $r \times c$ table with n observations per cell. For $i = 1, \ldots, r$, define $\rho_i = \sum_j a_j \mu_{ij}$, where $\sum a_j = 0$, but not all the $a_j = 0$. As defined, ρ_i is a contrast, but it only concerns cells in the i-th row of the table. Derive the F-test for NH: $\rho_1 = \cdots = \rho_r$, against a general alternative. Use the data from the last problem to test the hypothesis that the difference $C - F_1$ is constant across varieties.

7. Using only the cell means from the data of Problem 5, carry out Tukey's test for additivity and test the hypothesis that $\delta_i = 0,\ i = 1, \ldots, r$ in the model

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \delta_i \beta_j$$

   assuming $\sum \delta_i = 0$.

8. The following model is often recommended as a useful generalization of the model corresponding to Tukey's additivity test:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \lambda \delta_i \gamma_j$$

   where $\sum \alpha_i = \sum \beta_j = \sum \delta_i = \sum \gamma_j = 0$. The analysis of this model is discussed by Johnson and Graybill (*Journal of the American Statistical Association*, 1972, page 862). Read their paper and carry out their test of NH: $\lambda = 0$ for the cell mean data of Problem 5.

9. An experiment is planned with two factors with $r = 5$ and $c = 6$, with n observations per treatment. Let $\beta_n(\text{rows})$ be the power of the F-test (at level $\alpha = 0.05$) concerning row effects. Draw a power curve (n versus $\beta_n(\text{rows})$) under the following assumptions.

   (a) Under the alternative hypothesis, the row means fall on the quadratic curve $\bar{y}_{i+} = di^2$, with $d = 0.01\sigma$.

   (b) Under the alternative hypothesis, the row effects are all equal, except for one row where the effect is different. The difference between the two means is σ.

Chapter 8

Multiple Testing

The goal of multiple testing methods is to control family-wise error rate when we do many tests, as compared to controlling the per-test error rate. We briefly discuss three methods.

8.1 Scheffé method

Let

$$Y \sim \mathrm{N}(X\beta, \sigma^2 I)$$

The goal is to find a region C_α^ℓ such that

$$\Pr(\beta'\ell \in C_\alpha^\ell, \ell \in L) = 1 - \alpha \tag{8.1}$$

where L is a subspace, and C_α^ℓ can depend on ℓ. Here are some examples of interesting choices of L.

1. $L = \{\ell | \ell' J = 0\}$, and L is the set of all linear contrasts among the elements of β.
2. $L = \{\ell \in \mathcal{R}(X')\}$, which will get the set of all fitted values/predictions.

The *Scheffé intervals* are

$$\Pr\left(\ell'\beta \in \ell'\hat{\beta} \pm [q\hat{\sigma}^2 F(1-\alpha, p, n-p)\ell'(X'X)^{-1}\ell]^{1/2}, \ell \in L\right) = 1 - \alpha \tag{8.2}$$

where X is $n \times p$ rank p, $\dim(L) = q$, and $\hat{\sigma}^2$ is the usual estimate of σ^2. This interval should be compared to the t-interval,

$$\ell'\hat{\beta} \pm t(1-\alpha, n-p)\hat{\sigma}\sqrt{\ell'(X'X)^{-1}\ell}$$

Here is the outline of the derivation of the method. Suppose $\{b_1, \ldots, b_q\}$ is a basis for L, so we can write for any ℓ,

$$\ell = \sum \lambda_i b_i$$

Let B be a $q \times p$ matrix whose ith row is b_i'. A $1-\alpha$ confidence region for $B\beta$ is

$$\Pr\left([B(\beta - \hat{\beta})]'(B(X'X)^{-1}B')^{-1}[B(\beta - \hat{\beta})] \leq q\hat{\sigma}^2 F(1-\alpha, p, n-p)\right) = 1-\alpha \tag{8.3}$$

We can always reparameterize the model so that $(B(X'X)^{-1}B') = I$. In this case, (8.3) is

$$\Pr\left(\| B(\beta - \hat{\beta}) \|^2 \leq q\hat{\sigma}^2 F(1-\alpha, p, n-p)\right) = 1-\alpha \tag{8.4}$$

And this is a sphere centered at $\hat{\beta}$ with radius given by $(q\hat{\sigma}^2 F(1-\alpha, p, n-p))^{1/2}$.

The confidence interval for $\ell'\beta$ will consist of all values $\ell'\beta_0$ such that β_0 is in the region specified by (8.4). For all $\ell \in L$ simultaneously,

$$\Pr((\ell'\beta - \ell'\hat{\beta})^2 \leq q\hat{\sigma}^2 F(1-\alpha, p, n-p)) = 1-\alpha$$

which is the same as (8.1).

For example, in the one-way layout with t groups and m_i observations per group, $L = \{\ell | \ell' 1 = 0\}$ is the linear subspace of all possible contrasts. We easily compute $(X'X)^{-1} = \text{diag}(1/m_i)$, $\ell'(X'X)^{-1}\ell = \sum \ell_i^2/m_i$, and the Scheffé bounds are

$$\ell'\hat{\beta} \pm [(t-1)\hat{\sigma}^2 F(1-\alpha, t-1, n-t) \sum \ell_i^2/m_i]^{1/2}$$

Scheffé showed that if the rows of B are a basis for L, the F-test of $B\beta = 0$ is rejected at level α if and only if at least one simultaneous interval does not include zero.

8.1.1 Bonferroni method

Suppose we plan to do exactly K tests, and suppose in each case that the null hypothesis is correct. Let E_i be the probability that we fail to reject the i the hypothesis at levelα_i, so if $\bar{E}_i$ is the event that we reject the i-th hypothesis at level α_i, then $\Pr(\bar{E}_i) = \alpha$ and $\Pr(E_i) = 1-\alpha$.

The first Bonferroni bound states that

$$\Pr(\cap E_i) \geq 1 - \sum \Pr(\bar{E}_i)$$

or in the application to K tests,

$$\Pr(E_1 \cap E_2 \cdots \cap E_K) \leq 1 - \sum \alpha_i$$

which suggests that if we perform the i-th test at level $\alpha_i = \alpha/K$, the overall probability of all the E_i holding is at least $1 - \alpha$. The Bonferroni method can be used in almost any testing situation where we can specify the number of tests in advance. For testing contrasts, the intervals are

$$\ell'\hat{\beta} \pm t(1 - \alpha/K, n - p)\hat{\sigma}\sqrt{\ell'(X'X)^{-1}\ell} \tag{8.5}$$

For example, if we have $t = 5$ means with $m = 10$ observations per mean, at $\alpha = 0.05$ the critical value for testing a constrast with the Scheffé is $10.33\hat{\sigma}^2$. For the Bonferroni method, $t^2(.05/K, 45)$ is given in the following table:

1	2	5	10	15	20	25	30
4.06	5.38	7.23	8.70	9.61	10.30	10.82	11.22

so at least 20 comparisons are needed for the Scheffé method to be preferred to Bonferroni.

8.2 Tukey's method

Suppose $y_1, \ldots, y_n$ are independent N$(0 < \sigma^2)$, and we have an independent estimate $d\hat{\sigma}^2/\sigma^2 \sim \chi^2(d)$. Then the quantity

$$Q = (\max(y_i) - \min(y_i))/\hat{\sigma} = \max(y_i - y_j)/\hat{\sigma}$$

is called the *Studentized range*. Although its distribution is complicated, quantiles of this distribution are frequently tabled, and can even be computed in R using the function `qtukey`. The distribution of Q depends on n and of d.

Tukey's method, called the HSD or honestly significant difference, if designed to compare cell means in a factorial design. If we have a factor with a levels and corresponding means $\mu_1, \ldots, \mu_a$, then a 95% simultaneous confidence interval for the difference between two means is

$$|\hat{\mu}_i - \hat{\mu}_j| > \hat{\sigma}Q(\alpha, n, d)/\sqrt{h(m)}$$

If all the means are equally replicated, then $h(m) = m$. If the m_i are unequal, and approximate interval due to Tukey and Kramer is obtained if $h(m) = [(1/m_i + 1/m_j)/2]^{-1}$, which is the harmonic mean of the sample sizes.

For the table presented in the table in the last section, if we consider only comparisons of means rather than all possible contrasts, The comparable value for Tukey's method is $2 \times Q(1 - \alpha, 5, 45) = 8.12$, so for the problem for which it is intended, Tukey's method is better than its competitors.

In R, the function TukeyHSD does the computing for you.

```
> data(warpbreaks)
> summary(fm1 <- aov(breaks ~ wool * tension, data = warpbreaks))
             Df Sum Sq Mean Sq F value    Pr(>F)
wool          1  450.7   450.7  3.7653 0.0582130 .
tension       2 2034.3  1017.1  8.4980 0.0006926 ***
wool:tension  2 1002.8   501.4  4.1891 0.0210442 *
Residuals    48 5745.1   119.7
---
> TukeyHSD(fm1, "tension", ordered = TRUE)
  Tukey multiple comparisons of means
    95% family-wise confidence level
    factor levels have been ordered
Fit: aov(formula = breaks ~ wool + tension, data = warpbreaks)
$tension
         diff        lwr      upr
M-H  4.722222 -4.6311985 14.07564
L-H 14.722222  5.3688015 24.07564
L-M 10.000000  0.6465793 19.35342

> TukeyHSD(fm1, "wool:tension", ordered = FALSE)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = breaks ~ wool * tension, data = warpbreaks)
$"wool:tension"
             diff       lwr        upr
 [1,] -16.3333333 -31.63966  -1.027012
 [2,] -20.5555556 -35.86188  -5.249234
 [3,] -15.7777778 -31.08410  -0.471456
 [4,] -20.0000000 -35.30632  -4.693678
 [5,] -25.7777778 -41.08410 -10.471456
 [6,]  -4.2222222 -19.52854  11.084100
 [7,]   0.5555556 -14.75077  15.861877
 [8,]  -3.6666667 -18.97299  11.639655
```

```
 [9,]  -9.4444444 -24.75077   5.861877
[10,]   4.7777778 -10.52854  20.084100
[11,]   0.5555556 -14.75077  15.861877
[12,]  -5.2222222 -20.52854  10.084100
[13,]  -4.2222222 -19.52854  11.084100
[14,] -10.0000000 -25.30632   5.306322
[15,]  -5.7777778 -21.08410   9.528544
```

There are many other multiple testing methods that are particularly suited to an individual problem. One reasonable rule is to use the method that gives the shortest confidence intervals, and this can vary from problem to problem.

References: R. G. Miller, Jr., (1981). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York, NY; Hochberg, Yosef and Tamhane, Ajit. (1997) *Multiple Comparison Procedures*. New York: Wiley; Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Boca Raton: CRC Press.