Lesson 1.2. The Method of Fitting Constants

Method of fitting constants

In situations where approximate methods are inappropriate, such as when empty cells occur (some $n_{ij=0}$) or when the n_{ij} are dramatically different, the experimenter must use an exact analysis. The approach used to develop sums of squares for testing main effects and interactions is to represent the analysis of variance model as a regression model, fit that model to the data, and use the general regression significance test approach. The approach used to develop sums of squares for testing main effects and interactions is to represent the analysis of variance model as a regression model, fit that model to the data, and use the general regression significance test approach. However, this may be done in several ways, and these methods may result in different values for the sums of squares. Furthermore, the hypotheses that are being tested are not always direct analogs of those for the balanced case, nor are they always easily interpretable.

This approach is pplicable to data from well designed experiments as well as to data where good design is an impossibility. However, the method is computationally complex. Since we express the ANOVA problem as a regression problem, covariates may be included.

Note that (1) if interaction is present and the model assumed is that without interaction, then the estimates of the main effects are only approximates, becoming increasingly poor as the degree of disproportion increases, and (2) if there is no interaction and a model with interaction is assumed, main effects are inefficiently estimated though the estimates remain unbiased.

Example:

Consider the same data.

Material	Temperature (°F)		
Type	15	70	125
1	130 155 74	34 40 80	 70 58
2	159 126	136 115	45
3	138 160	150 139	96

The table of totals together with the class frequencies are shown below.

Material	Temperature (°F)			Material
Type	15	70	125	Total
1	359 (3)	154 (3)	128 (2)	641 (8)
2	285 (2)	251 (2)	45 (1)	581 (5)
3	298 (2)	289 (2)	96 (1)	683 (5)
Temp Total	942 (7)	694 (7)	269 (4)	1905 (18)

For this data, verify that

$$
\begin{aligned}
SST &= 32792.50; \quad df = 17 \\
SSA &= 10557.63; \quad df = 2 \\
SSB &= 12049.18; \quad df = 2 \\
SSTr &= 26961.67; \quad df = 8 \\
SSE &= SST - SSTr = 5830.83; \quad df = 9
\end{aligned}
$$

Suppose a model without interaction is applicable.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

with restrictions $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$.

In matrix notation, we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where: - $\mathbf{Y}$ is an $n \times 1$ vector of observed random variables,

- $\mathbf{X}$ is an $a + b + 1$ matrix of 0's and 1's, which we call as the **design matrix**
- $\boldsymbol{\beta}$ is an $a + b + 1$ vector of unknown parameters, the regression coefficients
- $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of unobserved random variables

For this data we have

$$
\mathbf{Y}_{18\times 1} = \left[\begin{array}{c} y_{111} \\ y_{112} \\ y_{113} \\ \hdashline y_{121} \\ y_{122} \\ y_{123} \\ \hdashline y_{131} \\ y_{132} \\ \hdashline y_{211} \\ y_{212} \\ \hdashline y_{221} \\ y_{222} \\ \hdashline y_{231} \\ \hdashline y_{311} \\ y_{312} \\ \hdashline y_{321} \\ y_{322} \\ \hdashline y_{331} \end{array}\right] = \left[\begin{array}{c} 130 \\ 155 \\ 74 \\ \hdashline 34 \\ 40 \\ 80 \\ \hdashline 70 \\ 58 \\ \hdashline 159 \\ 126 \\ \hdashline 136 \\ 115 \\ \hdashline 45 \\ \hdashline 138 \\ 160 \\ \hdashline 150 \\ 139 \\ \hdashline 96 \end{array}\right]
$$

while the design matrix is

$$
\mathbf{X}_{18\times 7} = \left[\begin{array}{ccccccc}
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 \\ \hdashline
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\ \hdashline
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hdashline
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\ \hdashline
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 \\ \hdashline
1 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hdashline
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\ \hdashline
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\ \hdashline
1 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}\right]
$$

and the parameter vector is

$$\boldsymbol{\beta}_{7\times 1} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Next we set up simultaneous equations to estimate the parameters: $\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$. To come up with unique estimates of the parameters we assume that

$$\sum_{i=1}^{a} \alpha_i = 0 \implies \alpha_1 + \alpha_2 + \alpha_3 = 0 \implies \alpha_3 = -\alpha_1 - \alpha_2$$

and

$$\sum_{j=1}^{b} \beta_j = 0 \implies \beta_1 + \beta_2 + \beta_3 = 0 \implies \beta_3 = -\beta_1 - \beta_2$$

The reparameterized design matrix and parameter vector are, respectively, given by

$$\mathbf{X}_{18\times 5} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

and

$$\boldsymbol{\beta}_{5\times 1} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

```r
#Data
Y = as.matrix(c(130,  155, 74, 34, 40 ,80 ,70,58, 159,126,136,115,45,138,160,150 ,139,96

#reparameterized design matrix
X1 <- matrix(c(1,   1,  0,  1,  0,
                      1,    1,  0,  1,  0,
                      1,    1,  0,  1,  0,
                      1,    1,  0,  0,  1,
                      1,    1,  0,  0,  1,
                      1,    1,  0,  0,  1,
                      1,    1,  0,  -1, -1,
                      1,    1,  0,  -1, -1,
                      1,  0,    1,  1,  0,
                      1,    0,  1,  1,  0,
                      1,    0,  1,  0,  1,
                      1,    0,  1,  0,  1,
                      1,    0,  1,  -1, -1,
                      1,    -1, -1, 1,  0,
                      1,    -1, -1, 1,  0,
                      1,    -1, -1, 0,  1,
                      1,    -1, -1, 0,  1,
                      1,    -1, -1, -1, -1),nrow=18, byrow = TRUE)

#Estimates of the parameters
solve(t(X1)%*%X1)%*%(t(X1)%*%Y)
```

```
##             [,1]
## [1,] 105.553763
## [2,] -29.301075
## [3,]   4.450538
## [4,]  33.203533
## [5,]  -2.225038
```

So we have

$$
\begin{aligned}
\hat{\mu} &= 105.553763 \\
\hat{\alpha}_1 &= -29.301075 \\
\hat{\alpha}_2 &= 4.450538 \\
\hat{\alpha}_3 &= 24.85054 \\
\hat{\beta}_1 &= 33.20353 \\
\hat{\beta}_2 &= -2.225038 \\
\hat{\beta}_3 &= -30.97849
\end{aligned}
$$

The reduction in the sum of squares due to fitting constant is given by

$$
\hat{\mu} \times Y_{...} + \sum_{i=1}^{a} \hat{\alpha}_i \times Y_{i..} + \sum_{j=1}^{b} \hat{\beta}_j \times Y_{.j.} =
$$

```r
estimates <- c(105.553763, -29.301075, 4.450538, 24.85054, 33.203533, -2.225038 , -30.97
totals <- c(1905,641,581, 683,942, 694,269)
sum(estimates*totals)
```

```
## [1] 223256.9
```

So,

$$
\begin{aligned}
\text{Adjusted model SS} &= 223256.9 - \frac{1905^2}{18} = 21644.4 \\
\text{SSA (adjusted for B)} &= 21644.4 - SSB = 21644.4 - 12049.18 = 9595.22 \\
\text{SSB (adjusted for A)} &= 21644.4 - SSA = 21644.4 - 10557.63 = 11086.77 \\
\text{Adjusted SSE} &= SST - 21644.4 = 32792.50 - 21644.4 = 11148.1
\end{aligned}
$$

SoV	df	SS	MS	F	p
Mat	2	9595.22	4797.61	5.594579	0.01766
Temp	2	11086.77	5543.385	6.46424	0.01125
Error	13	11148.1	857.5462		