8 Factorial Models

8.1 Introduction

In the experimental designs that have concerned us so far, we have been chiefly interested in how the levels of a single factor affect the variation in the response variable. However, in some situations, it may be suspected that the responses may be affected by two or more factors. The introduction of a second factor will allow us to model and test partial relationships (*main effects* of factors) as well as *interactions*.

In a *crossed design* the levels of one factor are reproduced at all levels of a second.

8.2 The Completely Randomized Two-Factor Factorial Design

A general representation of the two-factor design may be given as follows:

		Factor B					
		1	**2**	...	**j**	...	**b**
Factor	**1**	$y_{111}\cdots y_{11n}$	$y_{121}\cdots y_{12n}$	...	$y_{1j1}\cdots y_{1jn}$	...	$y_{1b1}\cdots y_{1bn}$
	2	$y_{211}\cdots y_{21n}$	$y_{221}\cdots y_{22n}$	...	$y_{2j1}\cdots y_{2jn}$	...	$y_{2b1}\cdots y_{2bn}$
	...	...	...	...	...	...	...
A	**i**	$y_{i11}\cdots y_{i1n}$	$y_{i21}\cdots y_{i2n}$	...	$y_{ij1}\cdots y_{ijn}$	...	$y_{ib1}\cdots y_{ibn}$
	...	...	...	...	...	...	...
	a	$y_{a11}\cdots y_{a1n}$	$y_{a21}\cdots y_{a2n}$	...	$y_{aj1}\cdots y_{ajn}$	...	$y_{ab1}\cdots y_{abn}$

The *completely randomized two-factor factorial design* is characterized by:

• a collection of responses $\{y_{ijk}\}$, where y_{ijk} represents the response corresponding to the k-th replicate at the i-th level of A, and the j-th level of B, with $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n$
• $n \geq 2$ so that there are at least 2 replications in each cell of the table;
• the abn responses are taken in random order (thus *completely randomized*).

It is usual to associate the following linear statistical model to this design:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \qquad \begin{array}{l} i=1,\ldots,a \\ j=1,\ldots,b \\ k=1,2,\ldots,n \end{array} \tag{8.1}$$

where
μ is the *overall mean*[1],
τ_i is the effect at the i-th level of Factor A,
β_j is the effect at the j-th level of Factor B,
$(\tau\beta)_{ij}$ is the interaction effect at the i-th level of factor A and j-th level of factor B, and
$\epsilon_{ijk} \sim \text{NID}(0, \sigma^2)$ is a random error term.

[1] when using the *sum-to-zero* constraints.

We further suppose that all of the effects defined above are fixed. Thus, we impose the convenient (and intuitively reasonable) '*sum to zero*' constraints:

$$\textstyle\sum_{i=1}^{a} \tau_i = 0 \tag{8.2}$$
$$\textstyle\sum_{j=1}^{b} \beta_j = 0 \tag{8.3}$$
$$\textstyle\sum_{i=1}^{a} (\tau\beta)_{ij} = 0 \quad j = 1, \ldots, b \tag{8.4}$$
$$\textstyle\sum_{j=1}^{b} (\tau\beta)_{ij} = 0 \quad i = 1, \ldots, a. \tag{8.5}$$

Differences between the various levels of the two factors are of equal interest this time, and so we test between the following pairs of hypotheses:

$$\begin{aligned} H_{0A} &: \ \tau_1 = \tau_2 = \ldots = \tau_a = 0 \\ H_{1A} &: \ \tau_i \neq 0 \quad \text{for some } i \end{aligned}$$

$$\begin{aligned} H_{0B} &: \ \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_{1B} &: \ \beta_j \neq 0 \quad \text{for some } j \end{aligned}$$

$$\begin{aligned} H_{0AB} &: \ (\tau\beta)_{ij} = 0 \quad \text{for all } i, j \\ H_{1AB} &: \ (\tau\beta)_{ij} \neq 0 \quad \text{for some } i, j \end{aligned}$$

8.3 Some Notation

For our particular model, let us introduce the following (sub-) totals:

$$\begin{aligned} y_{i..} &= \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad i = 1 \ldots, a \\ y_{.j.} &= \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad j = 1 \ldots, b \\ y_{ij.} &= \sum_{k=1}^{n} y_{ijk} \qquad i = 1 \ldots, a, \ j = 1 \ldots, b \\ y_{...} &= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \end{aligned}$$

with

$$\overline{y}_{i..} = y_{i..}/bn \quad \overline{y}_{.j.} = y_{.j.}/an \quad \overline{y}_{ij.} = y_{ij.}/n \quad \overline{y}_{...} = y_{...}/abn.$$

8.4 Parameter Estimation

We derive parameter estimates for our model based upon the principle of least squares. The form of these estimates will help to motivate our decomposition of the total sum-of-squares when we consider Analysis of Variance.

First we construct the quantity

$$\mathcal{L} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \mu - \tau_i - \beta_j - (\tau\beta)_{ij})^2$$

which is a measure of the distance between the data points $\{y_{ijk}\}$ and their expected values under the model (8.1).

Our least squares estimates will be those values of μ, $\{\tau_i\}$, $\{\beta_j\}$, $\{(\tau\beta)_{ij}\}$ that (jointly) minimize $\mathcal{L}$: let us denote these by $\hat{\mu}$, $\{\hat{\tau}_i\}$, $\{\hat{\beta}_j\}$, $\{\widehat{(\tau\beta)}_{ij}\}$, respectively.

Carrying out partial differentiations with respect to each of these parameters and equating to zero yields the following equations:

$$-2\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \widehat{(\tau\beta)}_{ij}\right) = 0 \tag{8.6}$$

$$-2\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \widehat{(\tau\beta)}_{ij}\right) = 0 \qquad i=1,\ldots,a \tag{8.7}$$

$$-2\sum_{i=1}^{a}\sum_{k=1}^{n}\left(y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \widehat{(\tau\beta)}_{ij}\right) = 0 \qquad j=1,\ldots,b \tag{8.8}$$

$$-2\sum_{k=1}^{n}\left(y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \widehat{(\tau\beta)}_{ij}\right) = 0 \qquad i=1,\ldots,a, \quad j=1,\ldots,b. \tag{8.9}$$

These can be re-arranged to yield

$$abn\hat{\mu} + bn\sum_{i=1}^{a}\hat{\tau}_i + an\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{(\tau\beta)}_{ij} = y_{\ldots} \tag{8.10}$$

$$bn\hat{\mu} + bn\hat{\tau}_i + n\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{j=1}^{b}\widehat{(\tau\beta)}_{ij} = y_{i..} \qquad i=1,\ldots,a \tag{8.11}$$

$$an\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}\widehat{(\tau\beta)}_{ij} = y_{.j.} \qquad j=1,\ldots,b \tag{8.12}$$

$$n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n\widehat{(\tau\beta)}_{ij} = y_{ij.} \qquad i=1,\ldots,a,\ j=1,\ldots,b. \tag{8.13}$$

Now (8.10)-(8.13) constitute $ab + a + b + 1$ equations in that many unknowns. However, they do not possess a unique solution. Indeed:

- the a equations in (8.11) can be generated by summing (8.13) over j for each i;
- the b equations in (8.12) can be generated by summing (8.13) over i for each j;
- (8.10) can be generated by summing (8.13) over all i and j.

Thus, there are $a+b+1$ linear dependencies, and so we need to impose an additional $a+b+1$ constraints in order to obtain unique estimates. In line with the constraints (8.2) to (8.5), we introduce the following $a+b+1$ constraints on our parameter estimates:

$$\sum_{i=1}^{a} \hat{\tau}_i = 0 \tag{8.14}$$
$$\sum_{j=1}^{b} \hat{\beta}_j = 0 \tag{8.15}$$
$$\sum_{i=1}^{a} (\widehat{\tau\beta})_{ij} = 0 \quad j=1,\ldots,b \tag{8.16}$$
$$\sum_{j=1}^{b} (\widehat{\tau\beta})_{ij} = 0 \quad i=1,\ldots,a. \tag{8.17}$$

Notice that (8.14) and (8.15) constitute 2 independent constraints, whereas (8.16) and (8.17) constitute only $a+b-1$ constraints.[2] After invoking these constraints, we obtain the following parameter estimates:

$$\hat{\mu} = \frac{y_{...}}{abn} = \overline{y}_{...} \tag{8.18}$$
$$\hat{\tau}_i = \frac{y_{i..}}{bn} - \hat{\mu} = \overline{y}_{i..} - \overline{y}_{...} \qquad i=1,\ldots,a \tag{8.19}$$
$$\hat{\beta}_j = \frac{y_{.j.}}{an} - \hat{\mu} = \overline{y}_{.j.} - \overline{y}_{...} \qquad j=1,\ldots,b \tag{8.20}$$

$$\begin{aligned}(\widehat{\tau\beta})_{ij} &= \frac{y_{ij.}}{n} - \hat{\tau}_i - \hat{\beta}_j - \hat{\mu} \\ &= \overline{y}_{ij.} - (\overline{y}_{i..} - \overline{y}_{...}) - (\overline{y}_{.j.} - \overline{y}_{...}) - \overline{y}_{...} \\ &= \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...} \qquad i=1,\ldots,a, \quad j=1,\ldots,b.\end{aligned} \tag{8.21}$$

And the estimated or *fitted* value of y_{ijk} is given by:

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + (\widehat{\tau\beta})_{ij} \\ &= \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) \\ &= \overline{y}_{ij.}\end{aligned}$$

for $i=1,\ldots,a, \quad j=1,\ldots,b, \quad k=1,\ldots,n.$

Therefore, each one of the abn observations is estimated by the average of the n observations in its corresponding cell of the table. The residuals for this model are given by the $\{e_{ijk}\}$ where

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} = y_{ijk} - \overline{y}_{ij.} \tag{8.22}$$

8.5 Analysis of Variance

Motivated by the form of the estimates for the effects, measures of the variation between levels of A, and the levels of B, may be defined as

$$SS_A = bn \sum_{i=1}^{a} (\overline{y}_{i..} - \overline{y}_{...})^2$$
$$SS_B = an \sum_{j=1}^{b} (\overline{y}_{.j.} - \overline{y}_{...})^2$$

respectively.

[2] Why so?

Also, considering the form of the $\{\widehat{(\tau\beta)}_{ij}\}$, then a measure of the *interaction* between factors A and B is

$$SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})^2$$

We can think of the *residual sum-of-squares* as literally being the sum of squares of the residuals, and so

$$SS_R = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \overline{y}_{ij.})^2$$

A measure of the total variation in the data, is given by the *total (corrected) sum-of-squares*, which is defined as

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \overline{y}_{...})^2$$

It turns out that SS_T can be decomposed to yield an identity which solely involves the previously defined quantities[3]:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_R \tag{8.23}$$

Degrees of Freedom

Since Factor A occurs at a levels, then we associate it with $a-1$ degrees of freedom. Similarly Factor B occurs at b levels, so this has $b-1$ degrees of freedom.

The interaction between A and B is associated with the degrees of freedom for the ab cells of the table, i.e. $ab-1$, less those for A, and B: thus SS_{AB} has $ab-1-(a-1)-(b-1)$, i.e. $(a-1)(b-1)$ degrees of freedom.

As for SS_R, within each of the ab cells of the design, there are n data points: thus there are $a \times b \times (n-1)$ degrees of freedom associated with the quantity.

But notice that the aforementioned degrees of freedom sum to $abn-1$, i.e. those for SS_T:

$$abn - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1).$$

We summarize these results in the form of an ANOVA table:

[3] As usual, the cross terms that emerge yield zero, overall.

ANOVA TABLE

Source	DF	SS	MS	F
A	$a-1$	$bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2$	$SS_A/(a-1)$	$F_A=\frac{MS_A}{MS_R}$
B	$b-1$	$an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2$	$SS_B/(b-1)$	$F_B=\frac{MS_B}{MS_R}$
AB	$(a-1)(b-1)$	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$	$SS_{AB}/(a-1)(b-1)$	$F_{AB}=\frac{MS_{AB}}{MS_R}$
Residual	$ab(n-1)$	By subtraction	$SS_R/ab(n-1)$	
Total	$abn-1$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2$		

The F-statistics suggested in the final column are motivated by the fact that the expectations of the *mean sum-of-squares* take the following form:

$$\mathrm{E}[MS_A]=\mathrm{E}\left[\frac{SS_A}{a-1}\right]=\sigma^2+\frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$
$$\mathrm{E}[MS_B]=\mathrm{E}\left[\frac{SS_B}{b-1}\right]=\sigma^2+\frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$
$$\mathrm{E}[MS_{AB}]=\mathrm{E}\left[\frac{SS_{AB}}{(a-1)(b-1)}\right]=\sigma^2+\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$
$$\mathrm{E}[MS_R]=\mathrm{E}\left[\frac{SS_R}{ab(n-1)}\right]=\sigma^2$$

Large values of F_A, F_B, F_{AB}, lead us to doubt that H_{0A}, H_{0B}, H_{0AB} hold true, respectively. Furthermore, the relevant distribution theory allows us to conclude that we reject:

- H_{0A} if $F_A > F_{a-1,ab(n-1),\alpha}$ (each at the $100\alpha\%$ level of significance).
- H_{0B} if $F_B > F_{b-1,ab(n-1),\alpha}$
- H_{0AB} if $F_{AB} > F_{(a-1)(b-1),ab(n-1),\alpha}$

8.6 Checking the Fit of the Model

There are several avenues available for checking whether our model actually fits the data adequately: the ones we consider here are based around inspection of the *residuals*. The residuals e_{ijk}, are defined to be[4]

$$e_{ijk}=y_{ijk}-\hat{y}_{ijk}\quad i=1,\ldots,a;\ j=1,\ldots,b$$

We could consider plotting graphs of:
- e_{ijk} vs. $\hat{y}_{ijk}$, i.e. residuals against fitted values; any trends in this plot might indicate a poor fit;
- e_{ijk} vs. i,j, i.e. residuals against the factors (treatments), in order to check the assumption on the variance across factors.

[4] Alternatively, standardized residuals could be used. For the two-way ANOVA, these are $e'_{ijk}=\dfrac{e_{ijk}}{s\sqrt{1-1/n}}$

8.7 Example: Battery Operating Conditions (c.f. Montgomery p.234-235)

An engineer is investigating whether battery life is affected by:
• the material from which it is made;
• the prevailing temperature in its end-use environment

There are 3 different types of material (M_1,M_2,M_3), and three temperatures that are of particular interest (**1** for -9.4^oC, **2** for 21.1^oC, and **3** for 51.7^oC).

The design of the experiment is such that 4 different batteries are tested at each combination of levels of material and temperature, and all $3 \times 3 \times 4 = 36$ runs are tested in random order.

The battery lifetime data (in hours) arising from the experiment are given in the table below:

		Temperature					
		1		**2**		**3**	
Material	M_1	130	155	34	40	20	70
		74	180	80	75	82	58
	M_2	150	188	136	122	25	70
		159	126	106	115	58	45
Type	M_3	138	110	174	120	96	104
		168	160	150	139	82	60

This is an example of a *two-factor factorial design* (a 3×3 factorial experiment). To perform the analysis in R we create 3 vectors: one containing the battery life readings, and the other two containing the corresponding levels of factor A, material type, and factor B, temperature:

```
> battery.life <- c(130, 155, 74, 180, 150, 188, 159, 126, 138, 110, 168, 160,
                    34, 40, 80, 75, 136, 122, 106, 115, 174, 120, 150, 139,
                    20, 70, 82, 58, 25, 70, 58, 45, 96, 104, 82, 60)
> material <- rep(rep(1:3, rep(4, 3)), 3)
> material.fac <- factor(material)
> temp <- rep(1:3, rep(12, 3))
> temp.fac <- factor(temp)
> battery <- data.frame(battery.life, temp.fac, material.fac)
```

Noting that ':' is used to construct the *interaction* term in R, then the analysis of variance can be performed as follows:

```
> battery.aov<-aov(battery.life~material.fac+temp.fac+material.fac:temp.fac, data=battery)
# equivalently
# battery.aov<-aov(battery.life~material.fac*temp.fac,data=battery,qr=T)
> summary(battery.aov)
                      Df Sum Sq Mean Sq F value   Pr(>F)
material.fac           2  10684    5342   7.911  0.00198 **
temp.fac               2  39119   19559  28.968 1.91e-07 ***
material.fac:temp.fac  4   9614    2403   3.560  0.01861 *
Residuals             27  18231     675
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Here, both F_A and F_B are significant at the 1% level, and F_{AB} is significant at the 2% level. Thus, variation in the battery life data can be accounted for by the effects of different materials, different temperatures, and interaction between the various materials and temperatures. The `model.tables` function in R is useful for examining such effects[5].

```
> model.tables(battery.aov)

Tables of effects

 material.fac
material.fac
      1       2       3
-22.361   2.806  19.556

 temp.fac
temp.fac
     1      2      3
 39.31   2.06 -41.36

 material.fac:temp.fac
            temp.fac
material.fac 1       2       3
           1  12.278 -27.972  15.694
           2   8.111   9.361 -17.472
           3 -20.389  18.611   1.778

> model.tables(battery.aov, type = "means")

Tables of means
Grand mean

105.5278

 material.fac
material.fac
     1      2      3
 83.17 108.33 125.08

 temp.fac
temp.fac
     1      2      3
144.83 107.58  64.17

 material.fac:temp.fac
            temp.fac
material.fac 1      2      3
           1 134.75  57.25  57.50
           2 155.75 119.75  49.50
           3 144.00 145.75  85.50
```

Note that since we have a significant interaction, we should pay some attention to the final *'table of means'* which illustrates the effect of both factors in combination. The following plots (`plot.design` and `interaction.plot`) are useful for examining the relationships between the factors. (The interaction plots shows clearly how the effect of each of `temp` and `material` are different according to the level of the other).

[5] under the 'sum-to-zero' constraints as previously described.

```
> plot.design(battery.life ~ material.fac + temp.fac)
> par(mfrow = c(1, 2))
> interaction.plot(material, temp, battery.life, main = "Interaction Plot")
> interaction.plot(temp, material, battery.life)
```

Before endorsing these conclusions, we should check the adequacy of the model: this can be done by inspection of the residuals.

```
> library(MASS) #needed for stdres() function
> stres <- stdres(battery.aov)
> par(mfrow = c(2, 2))
> plot(fitted(battery.aov), stres, main = "Std. Res vs Fitted Values")
> abline(h=0) # adds a horizontal line through the origin
> plot(material, stres, main = "Std. Res vs Material")
> abline(h=0)
> plot(temp, stres, main = "Std. Res vs Temp")
> abline(h=0)
```

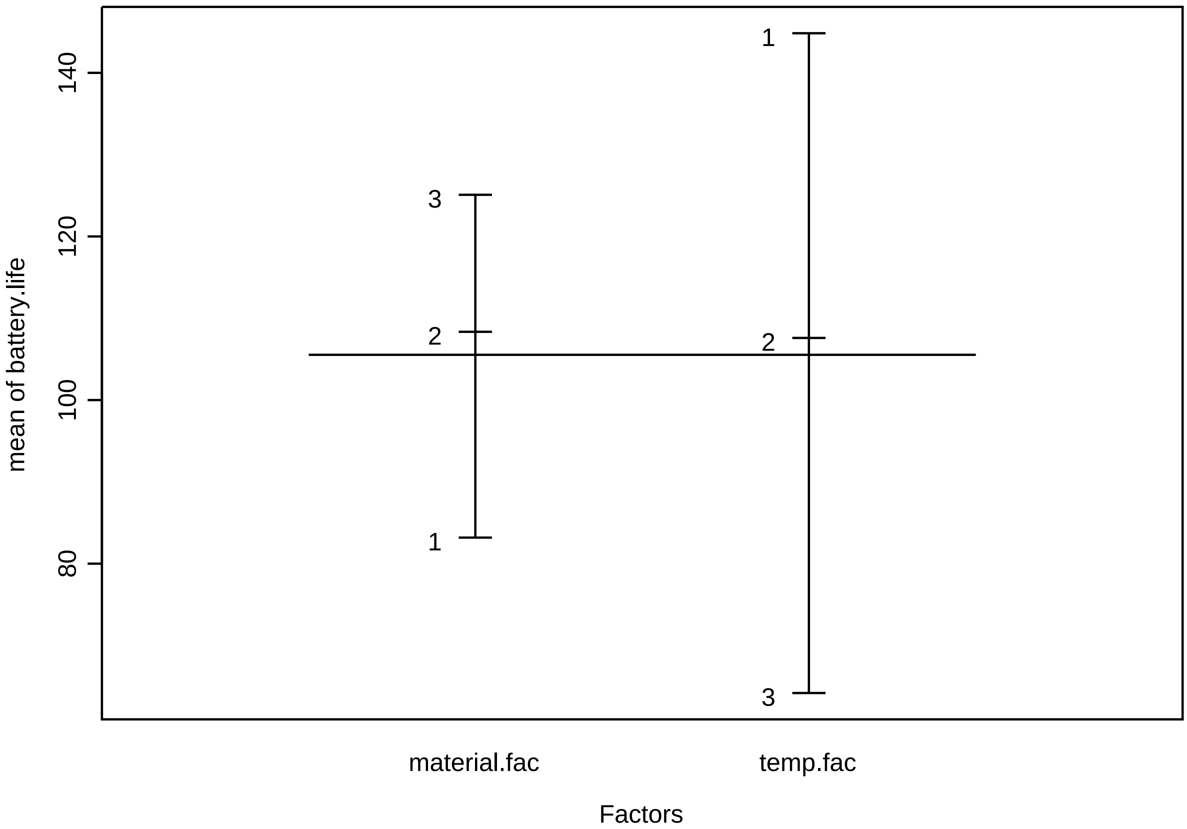
The first plot shows a mild tendency for the variability of the (standardized) residuals to increase with the size of the fitted values. This is not so significant to be of serious concern.

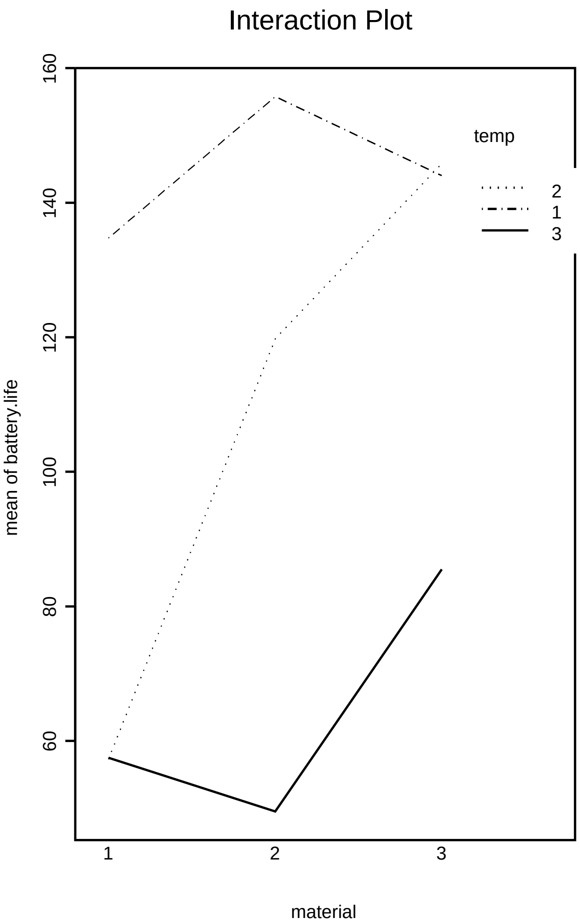
The variability of the residuals for material type M_1 is greater than those for M_2 and M_3; similarly the variability of the residuals for temperature 1 is higher than those for temperatures 2 and 3. Upon closer inspection, we see that rather high residuals are to be found in the cell corresponding to material type M_1 and temperature 1, these being -60.75 and 45.25 (standardized residuals -2.70 and 2.01):

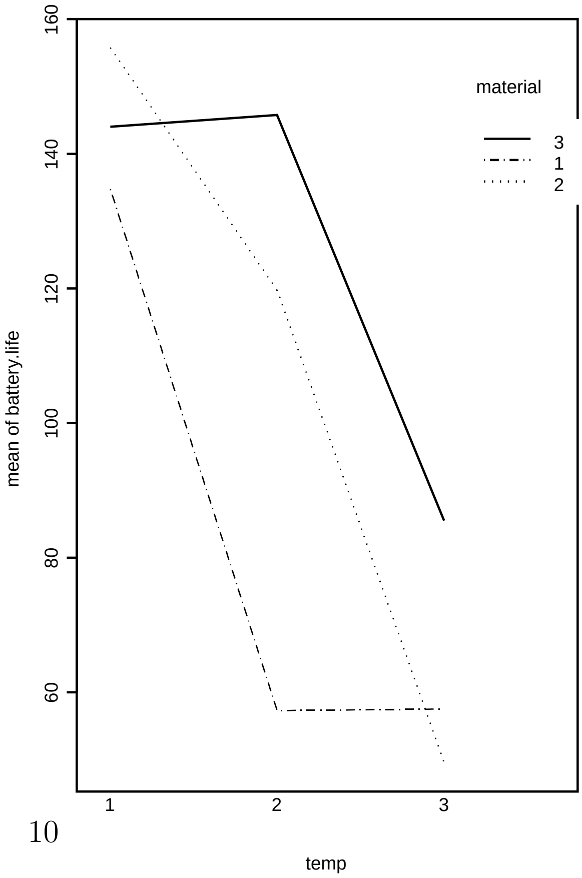
```
> residuals(battery.aov)
     1      2      3      4      5      6      7      8      9     10     11     12
 -4.75  20.25 -60.75  45.25  -5.75  32.25   3.25 -29.75  -6.00 -34.00  24.00  16.00
    13     14     15     16     17     18     19     20     21     22     23     24
-23.25 -17.25  22.75  17.75  16.25   2.25 -13.75  -4.75  28.25 -25.75   4.25  -6.75
    25     26     27     28     29     30     31     32     33     34     35     36
-37.50  12.50  24.50   0.50 -24.50  20.50   8.50  -4.50  10.50  18.50  -3.50 -25.50

> stres
          1           2           3           4           5           6           7
-0.21107782  0.89985806 -2.69957417  2.01079393 -0.25551525  1.43310728  0.14442166
          8           9          10          11          12          13          14
-1.32201369 -0.26662461 -1.51087279  1.06649844  0.71099896 -1.03317036 -0.76654575
         15          16          17          18          19          20          21
 1.01095165  0.78876447  0.72210832  0.09998423 -0.61101473 -0.21107782  1.25535754
         22          23          24          25          26          27          28
-1.14426395  0.18885910 -0.29995269 -1.66640381  0.55546794  1.08871716  0.02221872
         29          30          31          32          33          34          35
-1.08871716  0.91096742  0.37771820 -0.19996846  0.46659307  0.82209255 -0.15553102
         36
-1.13315459
```

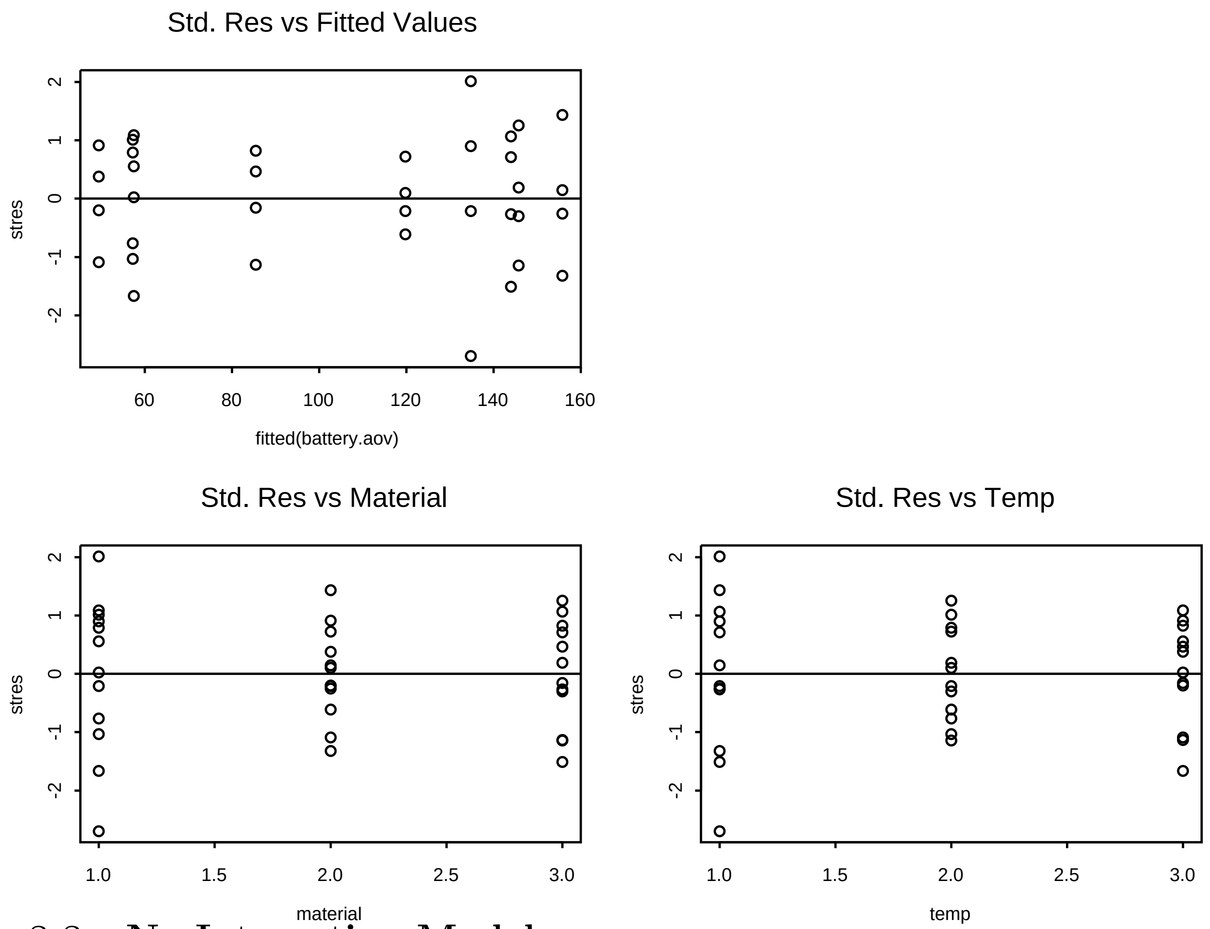
In practice, we could attempt to find an explanation for these high residuals, which may have resulted from an error in recording the battery life, faulty equipment etc., and then remove/replace that portion of the data if necessary.

Plots to examine effects of factors on a response









8.8 No Interaction Model

We could associate a linear statistical model with our experimental design that has no interaction term. In this case, the model equation becomes

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk} \quad i=1,\ldots,a, \quad j=1,\ldots,b, \quad k=1,\ldots,n$$

One can derive parameter estimates in much the same way as for the interaction model to discover that this time round, the fitted values are

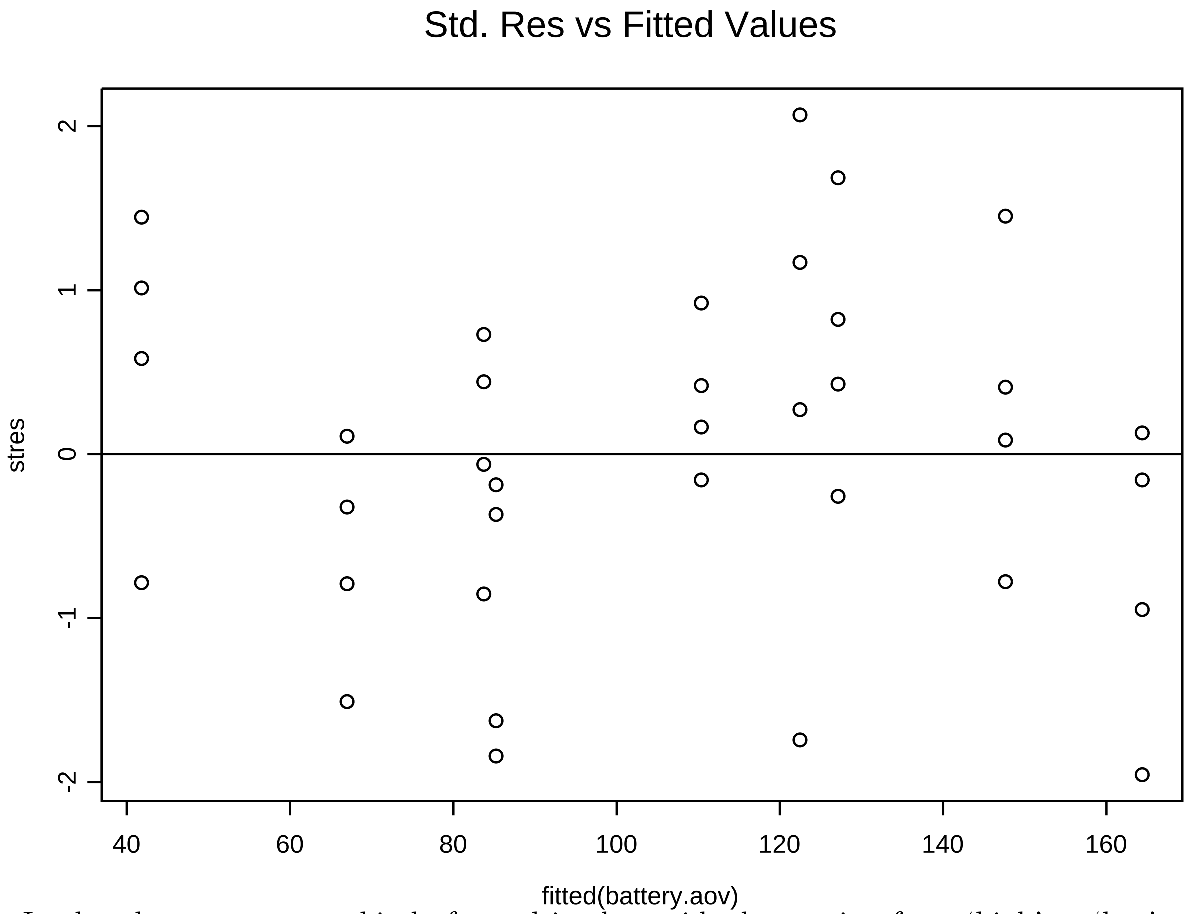
$$\hat{y}_{ijk} = \overline{y}_{i..} + \overline{y}_{.j.} - \overline{y}_{...} \quad i=1,\ldots,a, \quad j=1,\ldots,b, \quad k=1,\ldots,n$$

Analysis of Variance is performed as follows:

```
> battery.aov <- aov(battery.life ~ material.fac + temp.fac, data = battery)
> summary(battery.aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
material.fac  2  10684    5342   5.947  0.00651 **
temp.fac      2  39119   19559  21.776 1.24e-06 ***
Residuals    31  27845     898
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Model still indicates significant differences due to material type and temperature. However, we must check whether the model fits adequately or not.

```
> stres <- stdres(battery.aov)

> plot(fitted(battery.aov), stres)
> abline(h=0)
```



In the plot we see some kind of trend in the residuals, moving from 'high' to 'low', to 'high', then back to 'low' again, as the fitted values increase. Based on this *residual analysis*, we reject this *no-interaction* model.

Appendix

Computational Issues in the Two-way Design

In order to facilitate the coding of a procedure that performs two-way ANOVA outlined in Section 8.5 (customized in a particular computer language/package), we express the *sums-of-squares* in terms of the (sub-)totals arising in the table.

First define $C_f = y_{...}^2/abn$. Then

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - C_f$$
$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - C_f$$
$$SS_B = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - C_f$$

We can deal with the computation of SS_{AB} in two stages. First define

$$SS_{Subtotals} = \tfrac{1}{n}\textstyle\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - C_f$$

Then, secondly, we can obtain SS_{AB} by subtraction, as it can be shown that

$$SS_{AB} = SS_{Subtotals} - SS_A - SS_B.$$

Finally, SS_R can be obtained via subtraction through the identity (8.23):

$$SS_R = SS_T - SS_A - SS_B - SS_{AB}$$

which conveniently reduces down to

$$SS_R = SS_T - SS_{Subtotals}.$$