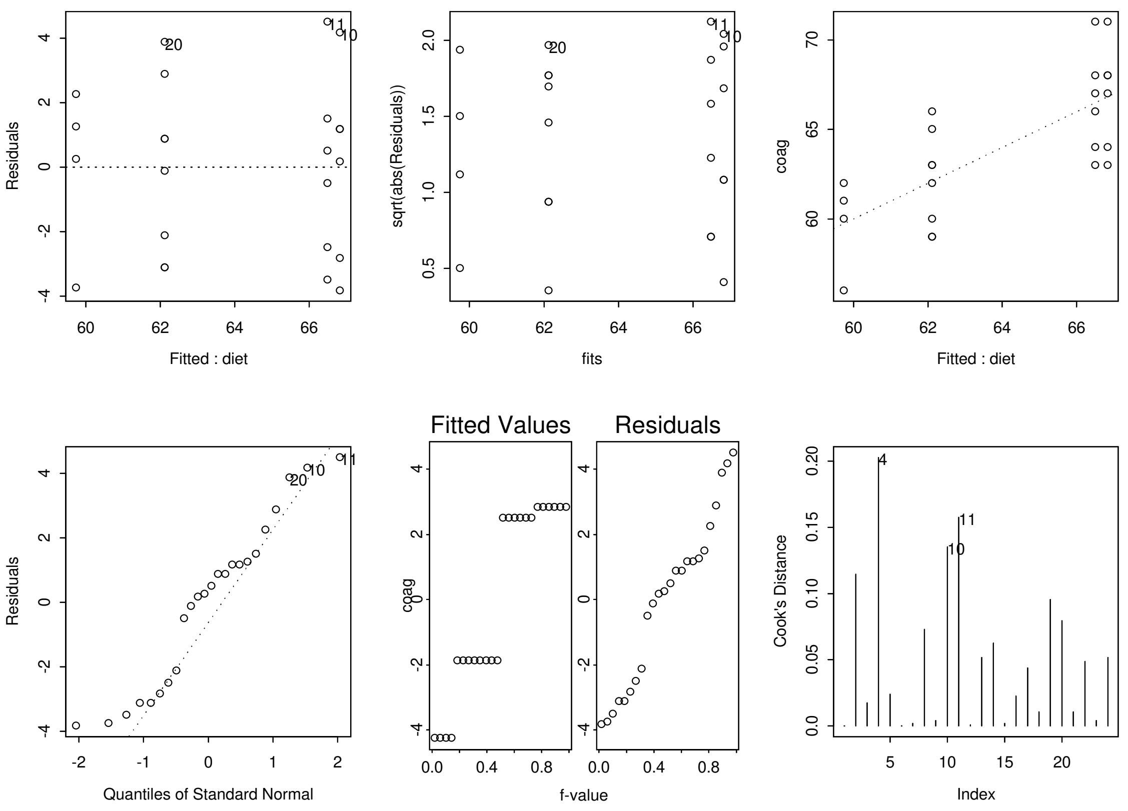
Notes for Tuesday January 27, 1998

1 Diagnostics and Remediations for ANOVA models

1.1 Graphical inspection of residuals

The diagnostic plots that one gets from plotting an aov or lm object in SPLUS give visual checks for most of the problems one could encounter in linear regression and ANOVA modeling:



Working from top left to lower right, these plots tell us:

- *First two plots: Equal error variance in all cells?* Fitted vs. residuals and fitted vs. sqrt(|residuals|).
- *Third plot: Nonlinearity (linear models) or non-equal error variances?*

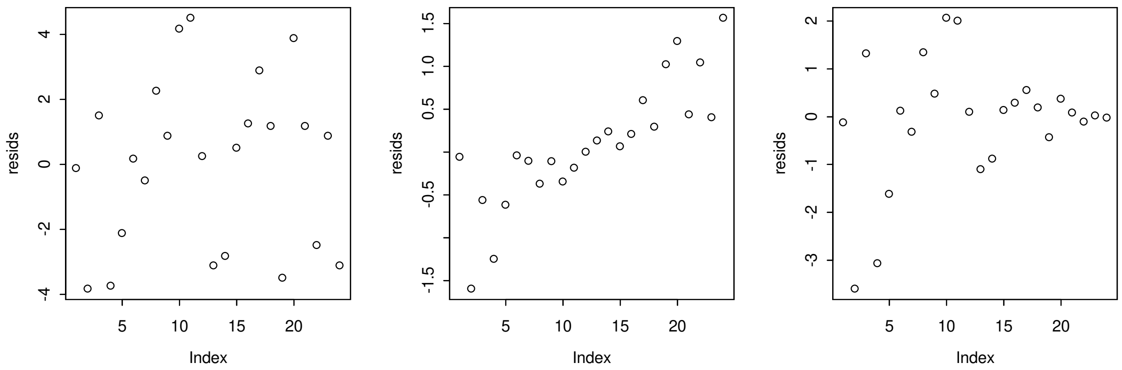
Fitted vs. observed.

- *Fourth plot: Normal errors?* Normal Quantile plot of residuals.
- *Fifth (double) plot: High r^2?* Normal quantile plots of fitted values, compared with normal quantile plot of residuals.
- *Sixth plot: high leverage or influence?* Cook's distances.

An additional plot can be helpful: In the very first plot, "fitted vs residuals", it is sometimes helpful to plot residuals not against the fitted values, but rather against some potentially important aspect of the data collection process, such as

- Time order observations were taken in;
- Spatial order (e.g. in an agricultural experiment, distance from the ag. station);
- etc.

What we are looking for are trends and/or serial correlations, in these plots. For example, here are three plots of residuals in time order.



- The plot on the left is a fine-looking residual plot.
- The middle plot shows a distressing time-trend in the residuals; it is probably best to begin working with the experimenter to find out what caused the trend, and recollect data (if possible) that does not have the trend. If this is not possible, then perhaps a time-series model for the residuals is needed.

- The plot on the right shows great variability in the residual variance, across time. Again, working with the experimenter to find the reason and correct it is the best course of action. Second best is to try some sort of transformation to fix the problem.

1.2 Variance-stabilizing transformations and Box-Cox transformations

Referring back to a "fitted vs residuals" plot can be helpful in selecting the right transformation, when unequal variances are discovered in a residual plot:

- *If σ_i^2 is proportional to μ_i,* try

$$y^* = \sqrt{y} \quad \text{or} \quad y^* - = \sqrt{y} + \sqrt{y+1}$$

- *If σ_i is proportional to μ_i,* try

$$y^* = \log y$$

- *If σ_i is proportional to μ_i^2,* try

$$y^* = 1/y$$

- *If y is a proportion,* so that $\sigma_i^2 = [\pi_i(1-\pi_i)/n_i$, try

$$y^* = 2 \arcsin \sqrt{y}$$

The idea behind each of these is the following. If $\sigma_{y_{ij}}^2 = g(\mu_i)$, then by taking variances on both sides of the Taylor approximation

$$h(y) - h(\mu) \approx h'(\mu)(y - \mu)$$

one can show that the variance after a transformation $y^* = h(y)$ is $\sigma_{y_{ij}^*}^2 \approx [h'(\mu_i)]^2 g(\mu_i)$. We wish to

$$\text{choose } h(\cdot) \text{ so that } [h'(\mu_i)]^2 g(\mu_i) = k, \text{ a constant,}$$

i.e. the transformed random variable y^* has constant variance k. This is a differential equation; the solutions, for $g(\mu) = \mu$, $\sqrt{(\mu)}$, etc., are listed above.

Finally,

- It might also be desirable to estimate, by hand or using Venables and Ripley's `boxcox()` function, the Box-Cox transformation

$$y^* = (y^\lambda - 1)/\lambda$$

  for the problem. A power transformation, such as the Box-Cox transformation, is also a standard way to try to correct non-normality in residuals.

After any of these transformations, the model should be re-fit and new diagnostics should be carried out, to ensure that the transformation helped and not hindered the solution of the problem.

1.3 Formal tests for equality of variances

There also exist formal tests of equality of variances in the cells of an ANOVA model. In general, visual inspection as in the previous section is adquate for many purposes.

As an example, we present the *Bartlett test*, which works for all ANOVA models, whether the number of observations per cell is constant or not; however it is fairly sensitive to non-normality in the residuals. When the number of observations per cell is constant, the *Hartley test* can be used.

The Bartlett Test. Let s_i^2 denote the sample variance of the residuals in the i^{th} cell of the data, $i = 1, \ldots, k$, each with degrees of freedom $df_i = n_i - 1$; and total degrees of freedom $df_T = \sum df_i$. Then the MSE can be expressed as an arithmetic mean

$$MSE = SSE/df_T = \frac{1}{df_T} \sum_{i=1}^{k} df_i s_i^2$$

On the other hand, the geometric mean is

$$GMSE = [(s_1^2)^{df_1} (s_2^2)^{df_2} \cdots (s_k^2)^{df_k}]^{1/df_T}$$

and the arithmetic-geometric mean inequality states that

$$GMSE \leq MSE.$$

Bartlett showed that if the y_{ij} are all independent $N(\mu_i, \sigma^2)$ (same variance in all cells) then

$$B = \frac{df_T}{C}(\log MSE - \log GMSE) = \frac{1}{C}\left[df_T \log MSE - \sum_{i=1}^{k} \log s_i^2\right]$$

is approximately χ^2_{k-1}, where

$$C = 1 + \frac{1}{3(k-1)}\left[\left(\sum_{i=1}^{k} \frac{1}{df_i}\right) - \frac{1}{df_T}\right].$$

This provides a test (large $B \Rightarrow$ reject H_0) of the hypothesis

$$\begin{aligned} H_0 : &\quad \sigma_1^2 = \cdots = \sigma_k^2, \ vs. \\ H_1 : &\quad \text{not all } \sigma_i^2 \text{ are equal.} \end{aligned}$$

Here is a short program in SPLUS to implement Bartlett's test, for one-way ANOVA.

```
bartlett.test _ function(y,x) {
#
# bartlett
#
# y is the response
# x is the factor
#
        ys _ split(y,x)
        vars _ sapply(ys,var)
        df _ sapply(ys,length)-1
        dfT _ sum(df)
        MSE _ sum(df*vars)/dfT
        k _ length(unique(x))
        const _ 1 + 1/(3*(k-1)) * ( sum(1/df) - 1/dfT )
        B _ (dfT*log(MSE) - sum(df*log(vars))) / const
        return(B=B,pval=1-pchisq(B,k-1))
}
```

Bartlett's test works for more complicated ANOVA problems also, but the SPLUS code above would have to be modified somehow to do this.

The Hartley test The Hartley test is based on the ratio

$$H = \frac{\max(\sigma_1^2, \ldots, \sigma_k^2)}{\min(\sigma_1^2, \ldots, \sigma_k^2)}$$

and simply compares this to the distribution of H under the null hypothesis

$$H_0 : \quad \sigma_1^2 = \cdots = \sigma_k^2$$

The distribution of H under H_0 has been tabulated, but it is also easy to simulate. Here is a short SPLUS program that implements the Hartley test, using a simulated p-value.

```
hartley.test _ function(y,x,reps=1000) {
#
# hartley.test
#
# y is the response
# x is the factor
#
# The simulation part of this is pretty slow!
#
        ys _ split(y,x)
        vars _ sapply(ys,var)
        ns _ sapply(ys,length)
        if (sum(abs(diff(ns)))!=0)
                stop("Cell sample sizes must be equal!")
        n _ ns[1]
        k _ length(unique(x))
        H _ max(vars)/min(vars)
        cnt _ 0
        for(i in 1:reps) {
                vars _ sapply(split(rnorm(n*k),x),var)
                if (max(vars)/min(vars)>H)
                        cnt _ cnt+1
```

```
        }
        return(H=H,pval=cnt/reps,reps=reps)
}
```

Like the SPLUS code for the Bartlett test, this is really only set up for one-way ANOVA; it would be fairly easy to adapt to other ANOVA models as well. Also, since the p-value is computed using simulation, this test could be adapted to unequal sample sizes in the cells (this adaptation would be much harder to do if we had to explicitly tabulate the distribution of H for "all" combinations of cell sample sizes).

2 Regression Equations for ANOVA; Reparametrization

We have seen that the "dummy variables" linear regression version of an ANOVA model is always overparametrized. To see how ANOVA reparametrizations overcome this, we now examine the one-way ANOVA model from a linear regression point of view.

For simplicity, consider a one-way ANOVA problem for a factor with three levels, a, b and c, and two observations per cell. Let

$$X_1 = \begin{cases} 1 & \text{at level } a \\ 0 & \text{else} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{at level } b \\ 0 & \text{else} \end{cases} \qquad X_3 = \begin{cases} 1 & \text{at level } c \\ 0 & \text{else} \end{cases}$$

so the linear regression model is

$$y_{ij} = \theta_0 + X_{1j}\theta_1 + X_{2j}\theta_2 + X_{3j}\theta_3 + \epsilon_{ij} \tag{1}$$

or in matrix form

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = X\theta + \epsilon = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

Basically we want to know how to introduce a linear constraint, to get a new set of parameters that can be estimated uniquely (technical term: the new parameters are *identifiable*).

2.1 Matrix algebra details

Note: *This is a bit more mathematically formal than the usual development of introducing constraints to make the parameters identifiable, but I wanted to be able to explain the "Helmert parametrization" in SPLUS, and this seems to require fooling around like this.*

We want to consider reparametrizations of the model $y = X\theta + \epsilon$, that is, we want to consider new sets of parameters

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = C \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \tag{2}$$

where

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}.$$

Our goal is to choose a new design matrix R so that

$$y = X\theta + \epsilon = Z\beta + \epsilon$$

i.e. the two regression problems are exactly the same (more formally: X and Z have the same column space). Note that, by (2), $R\beta = ZC\theta$, so it must be that

$$X \;=\; ZC.$$

Now, we would like to solve this by "multiplying on the right by C^{-1}", but unfortunately C is not square, so it doesn't have an inverse. Instead, notice what happens if we multiply on the right by $C'(CC')^{-1}$ (this is called a "generalized inverse"):

$$\begin{aligned} X &= ZC \\ X \cdot C'(CC')^{-1} &= ZC \cdot C'(CC')^{-1} \\ X \cdot C'(CC')^{-1} &= Z \end{aligned}$$

so now we have a way of finding Z from X and C, namely $Z = XC'(CC')^{-1}$.

To summarize, our goal in understanding reparametrization is:

> **Step 1.** Figure out what "constraint(s)" are being imposed on the θ's to get rid of the linear dependency;
>
> **Step 2.** Figure out how to write that as something like $\beta = C\theta$;
>
> **Step 3.** Compute the design matrix for the β's as $Z = XC'(CC')^{-1}$.
>
> **Step 4.** The new regression is then $y = Z\beta + \epsilon$.

Two notes before going on:

- In many simple problems, you can intuitively get from **Step 1** to **Step 4** without fooling around with the C matrix in **Steps 2** and **3**.
- After **Step 4** it is often helpful to write the columns of Z in terms of the 0/1 dummy variables X_1, X_2, X_3, and re-expressing the cell means μ_1, μ_2 and μ_3 in terms of the new parameters.

2.2 Some illustrations

Cell means model (delete the first column). This is easy. Setting $\theta_0 = 0$ is the same as deleting the first column in the X matrix; the matrix form of the regression becomes

this becomes

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

and (1) becomes

$$y_{ij} = X_{1j}\theta_1 + X_{2j}\theta_2 + X_{3j}\theta_3 + \epsilon_{ij}$$

and it is easy to see in the resulting model, $\mu_1 = \theta_1$, $\mu_2 = \theta_2$ and $\mu_3 = \theta_3$, the cell means model.

Let's see how we would do this by setting up the C matrix and calculating $Z = XC'(CC')^{-1}$:

We want $\theta_0 = 0$ and

$$\begin{aligned} \mu_1 &= \theta_1 \\ \mu_2 &= \theta_2 \\ \mu_3 &= \theta_3 \end{aligned}$$

or

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = C\theta$$

Hence

$$Z = XC'(CC')^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

which is the same design matrix that we got "by eye" above.

Delete the last column. Some computer packages set $\theta_3 = 0$ instead of $\theta_0 = 0$. If we set $\theta_3 = 0$, this is like deleting the last column in the matrix formulation.

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

and (1) becomes

$$y_{ij} = \theta_0 + X_{1j}\theta_1 + X_{2j}\theta_2 + \epsilon_{ij}$$

We can also do this by setting up C and computing $Z = XC'(CC')^{-1}$. We want to set $\theta_3 = 0$ and to define

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = C\theta$$

It then follows that

$$Z = XC'(CC')^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

as expected. What do the parameters mean here?

$$\begin{aligned} \mu_1 &= \gamma_0 + \gamma_1 \\ \mu_2 &= \gamma_0 + \gamma_2 \\ \mu_3 &= \gamma_0. \end{aligned}$$

Hence, $\gamma_0 = \mu_3$ is the mean of the last cell, and γ_1 and γ_2 are the differences between the third cell and the first and second (respectively).

Grand mean plus treatment effects. If we set the constraint

$$\theta_1 + \theta_2 + \theta_3 = 0$$

then we can express $\theta_3 = -\theta_1 - \theta_2$ and (1) becomes

$$y_{ij} = \theta_0 + X_{1j}\theta_1 + X_{2j}\theta_2 - X_{3j}(\theta_1 + \theta_2) + \epsilon_{ij}$$

or equivalently

$$y_{ij} = \theta_0 + (X_{1j} - X_{3j})\theta_1 + (X_{2j} - X_{3j})\theta_2 + \epsilon_{ij} \tag{3}$$

and the matrix form becomes

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

Now let's try setting up C and $Z = XC'(CC')^{-1}$:

Step 1. This parametrization can be thought of as doing two things: (a) we set $\theta_0 = 0$; and (b) we define:

$$\begin{aligned} \mu &= (\theta_1 + \theta_2 + \theta_3)/3 \\ \alpha_1 &= \theta_1 - \mu = \frac{2}{3}\theta_1 - \frac{1}{3}\theta_2 - \frac{1}{3}\theta_3 \\ \alpha_2 &= \theta_2 - \mu = -\frac{1}{3}\theta_1 + \frac{2}{3}\theta_2 - \frac{1}{3}\theta_3 \\ \alpha_3 &= \theta_3 - \mu = -\frac{1}{3}\theta_1 - \frac{1}{3}\theta_2 + \frac{2}{3}\theta_3 \end{aligned}$$

(as a check, note that $\alpha_1 + \alpha_2 + \alpha_3 = 0$ as desired).

Step 2. This means that $\alpha_3 = -\alpha_1 - \alpha_2$, so we can omit it from the reparametrization, and so C is the matrix in the transformation

$$\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = C\theta$$

Step 3. Therefore

$$Z = XC'(CC')^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

so the new model is

$$y = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \epsilon$$

Step 4. Now if we define

$$Z_1 = X_1 - X_3 = \begin{cases} 1 & \text{at level } a \\ 0 & \text{at level } b \\ -1 & \text{at level } c \end{cases} \qquad Z_2 = X_2 - X_3 = \begin{cases} 0 & \text{at level } a \\ 1 & \text{at level } b \\ -1 & \text{at level } c \end{cases}$$

we may express the model as

$$y_{ij} = \mu + Z_{1j}\alpha_1 + Z_{2j}\alpha_2 + \epsilon_{ij} \;=\; \mu + (X_1 - X_3)\alpha_1 + (X_2 - X_3)\alpha_2 + \epsilon_{ij}$$

In this model, the cell means are

$$\begin{aligned} \mu_1 &= \mu + \alpha_1 \\ \mu_2 &= \mu + \alpha_2 \\ \mu_3 &= \mu + \alpha_3 \end{aligned}$$

so that the α's really are offsets from the grand mean μ to each cell mean.

SPLUS's default "Helmert contrasts". The default parametrization for ANOVA models in SPLUS is in terms of the "Helmert contrasts." It is easiest to explain how to set this model up directly in terms of C and $Z = XC'(CC')^{-1}$:

Step 1. Like the "grand mean plus treatment effects" parametrization we do two things: (a) we set $\theta_0 = 0$; and (b) we define:

$$\begin{aligned} \mu &= (\theta_1 + \theta_2 + \theta_3)/3 \\ \beta_1 &= \{-\theta_1 + \theta_2\}/2 \\ \beta_2 &= \{-(\theta_1 + \theta_2)/2 + \theta_3\}/3 \\ \beta_3 &= \{-(\theta_1 + \theta_2 + \theta_3)/3 + \theta_4\}/4 \end{aligned}$$

(Here, $\theta_4 \equiv 0$ [since there is no fourth category] and is just included to show you the pattern, should you ever have to construct Helmert contrasts for more than three categories).

Step 2. Again, β_3 is expressible in terms of the other parameters ($\beta_3 = 3\mu$), so we omit it from the reparametrization, and C is the matrix in the transformation

$$\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = C\theta$$

Step 3. Therefore

$$Z = XC'(CC')^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix}$$

so the new model is

$$y = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon$$

Step 4. Now if we define

$$Z_1 = X_2 - X_1 = \begin{cases} -1 & \text{at level } a \\ 1 & \text{at level } b \\ 0 & \text{at level } c \end{cases} \qquad Z_2 = 2X_3 - X_2 - X_1 = \begin{cases} -1 & \text{at level } a \\ -1 & \text{at level } b \\ 2 & \text{at level } c \end{cases}$$

we may express the model as

$$y_{ij} = \mu + Z_{1j}\beta_1 + Z_{2j}\beta_2 + \epsilon_{ij} \;=\; \mu + (X_2 - X_1)\beta_1 + (2X_3 - X_2 - X_1)\beta_2 + \epsilon_{ij}$$

In this model, the cell means are

$$\begin{aligned}
\mu_1 &= \mu - \beta_1 - \beta_2 \\
\mu_2 &= \mu + \beta_1 - \beta_2 \\
\mu_3 &= \mu + 2\beta_2
\end{aligned}$$

Note that it is easy to read off the "treatment effects" also:

$$\begin{aligned}
\alpha_1 &= -\beta_1 - \beta_2 \\
\alpha_2 &= \quad \beta_1 - \beta_2 \\
\alpha_3 &= 2\beta_2
\end{aligned}$$