Analysis of Two-Factor Experiments Based on Additive Models

An Example Two-Factor Experiment

Researchers were interested in studying the effects of 2 diets (low fiber, high fiber) and 3 drugs (D1, D2, D3) on weight gained by Yorkshire hogs. A total of 12 hogs were assigned to the 6 diet-drug combinations using a balanced and completely randomized experimental design. Hogs were housed in individual pens, injected with their assigned drugs once per week, and fed their assigned diets for a 6-week period. The amount of weight gained during the 6-week period was recorded for each hog.

As discussed in the previous set of slides, this experiment involves the two factors: Diet (with levels low fiber and high fiber) and Drug (with levels D1, D2, and D3).

The Additive Model

When factors do not interact, it makes sense to consider the *additive* model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad (i = 1, 2;\ j = 1, 2, 3;\ k = 1, 2) \quad \text{where}$$

$\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ are unknown real-valued parameters and

$$\epsilon_{111}, \epsilon_{112}, \epsilon_{121}, \epsilon_{122}, \epsilon_{131}, \epsilon_{132}, \epsilon_{211}, \epsilon_{212}, \epsilon_{221}, \epsilon_{222}, \epsilon_{231}, \epsilon_{232} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

for some unknown $\sigma^2 > 0$.

Table of Treatments and Means for the Additive Model

Treatment	Diet	Drug	Mean
1	1	1	$\mu + \alpha_1 + \beta_1$
2	1	2	$\mu + \alpha_1 + \beta_2$
3	1	3	$\mu + \alpha_1 + \beta_3$
4	2	1	$\mu + \alpha_2 + \beta_1$
5	2	2	$\mu + \alpha_2 + \beta_2$
6	2	3	$\mu + \alpha_2 + \beta_3$

Diet 1 = Low Fiber, Diet 2 = High Fiber

Drug 1 = D1, Drug 2 = D2, Drug 3 = D3

Additive Model in Matrix and Vector Form

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

Additive Model in Matrix and Vector Form

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 + \beta_1 \\ \mu + \alpha_1 + \beta_1 \\ \mu + \alpha_1 + \beta_2 \\ \mu + \alpha_1 + \beta_2 \\ \mu + \alpha_1 + \beta_3 \\ \mu + \alpha_1 + \beta_3 \\ \mu + \alpha_2 + \beta_1 \\ \mu + \alpha_2 + \beta_1 \\ \mu + \alpha_2 + \beta_2 \\ \mu + \alpha_2 + \beta_2 \\ \mu + \alpha_2 + \beta_3 \\ \mu + \alpha_2 + \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

Table of Cell Means for the Additive Model

	Drug 1	Drug 2	Drug 3
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$

All Interactions are Zero for the Additive Model

The simple effect of Diet is $\alpha_1 - \alpha_2$ for all levels of Drug.

The simple effect of Drug j vs. Drug j^* is $\beta_j - \beta_{j^*}$ regardless of Diet.

	Drug 1	Drug 2	Drug 3
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$

Marginal Means for the Additive Model

	Drug 1	Drug 2	Drug 3	
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$	$\mu + \alpha_1 + \bar{\beta}.$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \bar{\beta}.$
	$\mu + \bar{\alpha}. + \beta_1$	$\mu + \bar{\alpha}. + \beta_2$	$\mu + \bar{\alpha}. + \beta_3$	$\mu + \bar{\alpha}. + \bar{\beta}.$

Simple Effects = Main Effects for the Additive Model

The main effect of Diet is

$$(\mu + \alpha_1 + \bar{\beta}_.) - (\mu + \alpha_2 + \bar{\beta}_.) = \alpha_1 - \alpha_2$$

	Drug 1	Drug 2	Drug 3	
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$	$\mu + \alpha_1 + \bar{\beta}_.$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \bar{\beta}_.$
	$\mu + \bar{\alpha}_. + \beta_1$	$\mu + \bar{\alpha}_. + \beta_2$	$\mu + \bar{\alpha}_. + \beta_3$	$\mu + \bar{\alpha}_. + \bar{\beta}_.$

Simple Effects = Main Effects for the Additive Model

The main effect for Drug 1 vs. Drug 2 is

$$(\mu + \bar{\alpha}. + \beta_1) - (\mu + \bar{\alpha}. + \beta_2) = \beta_1 - \beta_2$$

	Drug 1	Drug 2	Drug 3	
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$	$\mu + \alpha_1 + \bar{\beta}.$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \bar{\beta}.$
	$\mu + \bar{\alpha}. + \beta_1$	$\mu + \bar{\alpha}. + \beta_2$	$\mu + \bar{\alpha}. + \beta_3$	$\mu + \bar{\alpha}. + \bar{\beta}.$

Tests for Main Effects in the Additive Model

No Diet main effect $\iff \alpha_1 = \alpha_2$

No Drug main effects $\iff \beta_1 = \beta_2 = \beta_3$

	Drug 1	Drug 2	Drug 3	
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$	$\mu + \alpha_1 + \bar{\beta}.$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \bar{\beta}.$
	$\mu + \bar{\alpha}. + \beta_1$	$\mu + \bar{\alpha}. + \beta_2$	$\mu + \bar{\alpha}. + \beta_3$	$\mu + \bar{\alpha}. + \bar{\beta}.$

H_0: No Diet Main Effect $(\alpha_1 = \alpha_2)$

$$
\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix}
\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}
= \begin{bmatrix} 0 \end{bmatrix}
$$

H_0: No Drug Main Effects $(\beta_1 = \beta_2 = \beta_3)$

$$\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

LSMEANS for the Additive Model

LSMEANS are OLS estimators of the quantities in the margins below.

	Drug 1	Drug 2	Drug 3	
Diet 1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$	$\mu + \alpha_1 + \beta_3$	$\mu + \alpha_1 + \bar{\beta}.$
Diet 2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \bar{\beta}.$
	$\mu + \bar{\alpha}. + \beta_1$	$\mu + \bar{\alpha}. + \beta_2$	$\mu + \bar{\alpha}. + \beta_3$	$\mu + \bar{\alpha}. + \bar{\beta}.$

LSMEANS for the Additive Model (continued)

For example, the LSMEAN for Diet 1 is

$$\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \left[1, 1, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right] \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \hat{\mu} + \hat{\alpha}_1 + \frac{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3}{3},$$

where $\hat{\boldsymbol{\beta}}$ is any solution to the Normal Equations.

Although $\hat{\boldsymbol{\beta}}$ will depend on which of infinitely many solutions to the Normal Equations is used, $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ will be the same for all solutions.

SAS Code for the Additive Model

```
proc import datafile='C:\dietdrug.txt'
   dbms=TAB replace out=d;
run;

proc mixed;
  class diet drug;
  model weightgain=diet drug;
  lsmeans diet drug / cl;
  estimate 'diet effect' diet 1 -1 / cl;
  estimate 'drug 1 - drug 2' drug 1 -1  0;
  estimate 'drug 1 - drug 3' drug 1  0 -1;
  estimate 'drug 2 - drug 3' drug 0  1 -1;
  contrast 'drug main effects' drug 1 -1  0,
                               drug 1  0 -1;
run;
```

Fitting the Additive Model in R

```
d=read.delim("http://.../S510/dietdrug.txt")
```

```
d$diet=factor(d$diet)
```

```
d$drug=factor(d$drug)
```

```
o=lm(weightgain~diet+drug,data=d)
```

`coef(o)` is $\hat{\boldsymbol{\beta}}$.

`vcov(o)` is $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

`o$df` is $n - r$.

The Additive Model Matrix is Not Full-Column Rank

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

The Full-Rank Formulation Used by R

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}_R\boldsymbol{\beta}_R + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

The Full-Rank Formulation Used by R

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \mu + \beta_2 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu + \beta_3 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + \alpha_2 + \beta_2 \\ \mu + \alpha_2 + \beta_2 \\ \mu + \alpha_2 + \beta_3 \\ \mu + \alpha_2 + \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}_R\boldsymbol{\beta}_R + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

Table of Means for the R Full-Rank Formulation

	Drug 1	Drug 2	Drug 3	
Diet 1	μ	$\mu+\beta_2$	$\mu+\beta_3$	$\mu+\frac{\beta_2+\beta_3}{3}$
Diet 2	$\mu+\alpha_2$	$\mu+\alpha_2+\beta_2$	$\mu+\alpha_2+\beta_3$	$\mu+\alpha_2+\frac{\beta_2+\beta_3}{3}$
	$\mu+\frac{\alpha_2}{2}$	$\mu+\frac{\alpha_2}{2}+\beta_2$	$\mu+\frac{\alpha_2}{2}+\beta_3$	$\mu+\frac{\alpha_2}{2}+\frac{\beta_2+\beta_3}{3}$

Main Effects in the R Additive Model Formulation

No Diet main effect $\iff \alpha_2 = 0$

No Drug main effects $\iff \beta_2 = \beta_3 = 0$

	Drug 1	Drug 2	Drug 3	
Diet 1	μ	$\mu + \beta_2$	$\mu + \beta_3$	$\mu + \frac{\beta_2+\beta_3}{3}$
Diet 2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \frac{\beta_2+\beta_3}{3}$
	$\mu + \frac{\alpha_2}{2}$	$\mu + \frac{\alpha_2}{2} + \beta_2$	$\mu + \frac{\alpha_2}{2} + \beta_3$	$\mu + \frac{\alpha_2}{2} + \frac{\beta_2+\beta_3}{3}$

H_0: No Diet Main Effect ($\alpha_2 = 0$ in R)

$$\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

H_0: No Drug Main Effects ($\beta_2 = \beta_3 = 0$ in R)

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

LSMEANS for the R Formulation of the Additive Model

LSMEANS are still the OLS estimators of the quantities in the margins below.

	Drug 1	Drug 2	Drug 3	
Diet 1	μ	$\mu + \beta_2$	$\mu + \beta_3$	$\mu + \frac{\beta_2+\beta_3}{3}$
Diet 2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \frac{\beta_2+\beta_3}{3}$
	$\mu + \frac{\alpha_2}{2}$	$\mu + \frac{\alpha_2}{2} + \beta_2$	$\mu + \frac{\alpha_2}{2} + \beta_3$	$\mu + \frac{\alpha_2}{2} + \frac{\beta_2+\beta_3}{3}$

LSMEANS for the Additive Model (continued)

For example, the LSMEAN for Diet 1 is

$$\boldsymbol{c}'\hat{\boldsymbol{\beta}}_R = \left[1, 0, \frac{1}{3}, \frac{1}{3}\right] \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \hat{\mu} + \frac{\hat{\beta}_2 + \hat{\beta}_3}{3},$$

where $\hat{\boldsymbol{\beta}}_R = (\boldsymbol{X}_R'\boldsymbol{X}_R)^{-1}\boldsymbol{X}_R'\boldsymbol{y}$.

Fitting the Additive Model in R

```
> o=lm(weightgain~diet+drug,data=d)
> model.matrix(o)
   (Intercept) diet2 drug2 drug3
1            1     0     0     0
2            1     0     0     0
3            1     0     1     0
4            1     0     1     0
5            1     0     0     1
6            1     0     0     1
7            1     1     0     0
8            1     1     0     0
9            1     1     1     0
10           1     1     1     0
11           1     1     0     1
12           1     1     0     1
```

R: The $\hat{\beta}$ Vector

```
> #betahat vector:
>
> coef(o)
(Intercept)       diet2       drug2       drug3
  41.616667   -5.033333   -2.100000   -2.550000
```

R: $\widehat{\text{Var}}(\hat{\beta})$ and Error Degrees of Freedom

```
> #Estimated variance of betahat:
>
> vcov(o)

            (Intercept)         diet2         drug2         drug3
(Intercept)   0.8186111 -4.093056e-01 -6.139583e-01 -6.139583e-01
diet2        -0.4093056  8.186111e-01 -6.759159e-17 -6.759159e-17
drug2        -0.6139583 -6.759159e-17  1.227917e+00  6.139583e-01
drug3        -0.6139583 -6.759159e-17  6.139583e-01  1.227917e+00

> #The degrees of freedom for error:
>
> o$df
[1] 8
```

R: A Function for Point and Interval Estimation

```
> estimate=function(lmout,C,a=0.05)
+ {
+   b=coef(lmout)
+   V=vcov(lmout)
+   df=lmout$df
+   Cb=C%*%b
+   se=sqrt(diag(C%*%V%*%t(C)))
+   tval=qt(1-a/2,df)
+   low=Cb-tval*se
+   up=Cb+tval*se
+   m=cbind(C,Cb,se,low,up)
+   dimnames(m)[[2]]=c(paste("c",1:ncol(C),sep=""),
+                 "estimate","se",
+                 paste(100*(1-a),"% Conf.",sep=""),
+                 "limits")
+   m
+ }
```

R: Entering a C Matrix

```
> C=matrix(c(
+ 1, 0, 1/3, 1/3,
+ 1, 1, 1/3, 1/3,
+ 1, 1/2, 0, 0,
+ 1, 1/2, 1, 0,
+ 1, 1/2, 0, 1,
+ 0, -1, 0, 0,
+ 0, 0, -1,  0,
+ 0, 0,  0, -1,
+ 0, 0,  1, -1
+ ),ncol=4,byrow=T)
```

R: The C Matrix

```
> C
     [,1] [,2]       [,3]       [,4]
 [1,]    1  0.0  0.3333333  0.3333333
 [2,]    1  1.0  0.3333333  0.3333333
 [3,]    1  0.5  0.0000000  0.0000000
 [4,]    1  0.5  1.0000000  0.0000000
 [5,]    1  0.5  0.0000000  1.0000000
 [6,]    0 -1.0  0.0000000  0.0000000
 [7,]    0  0.0 -1.0000000  0.0000000
 [8,]    0  0.0  0.0000000 -1.0000000
 [9,]    0  0.0  1.0000000 -1.0000000
```

R: Interpreting the C Matrix

```
>
> #With this choice of C, you get estimates and
> #confidence intervals for the following:
>
> #Row 1: lsmean for diet 1
> #Row 2: lsmean for diet 2
> #Row 3: lsmean for drug 1
> #Row 4: lsmean for drug 2
> #Row 5: lsmean for drug 3
> #Row 6: diet 1 - diet 2 effect
> #Row 7: drug 1 - drug 2 effect
> #Row 8: drug 1 - drug 3 effect
> #Row 9: drug 2 - drug 3 effect
```

R: Results

```
> estimate(o,C)
     c1   c2         c3         c4  estimate        se     95% Conf.    limits
 [1,]  1  0.0  0.3333333  0.3333333 40.066667 0.6397699 38.591354577 41.541979
 [2,]  1  1.0  0.3333333  0.3333333 35.033333 0.6397699 33.558021243 36.508645
 [3,]  1  0.5  0.0000000  0.0000000 39.100000 0.7835549 37.293119084 40.906881
 [4,]  1  0.5  1.0000000  0.0000000 37.000000 0.7835549 35.193119084 38.806881
 [5,]  1  0.5  0.0000000  1.0000000 36.550000 0.7835549 34.743119084 38.356881
 [6,]  0 -1.0  0.0000000  0.0000000  5.033333 0.9047713  2.946926967  7.119740
 [7,]  0  0.0 -1.0000000  0.0000000  2.100000 1.1081140 -0.455315497  4.655315
 [8,]  0  0.0  0.0000000 -1.0000000  2.550000 1.1081140 -0.005315497  5.105315
 [9,]  0  0.0  1.0000000 -1.0000000  0.450000 1.1081140 -2.105315497  3.005315
```

R: Results

```
> estimate(o,C)[,-(1:4)]
      estimate        se    95% Conf.    limits
 [1,] 40.066667 0.6397699 38.591354577 41.541979
 [2,] 35.033333 0.6397699 33.558021243 36.508645
 [3,] 39.100000 0.7835549 37.293119084 40.906881
 [4,] 37.000000 0.7835549 35.193119084 38.806881
 [5,] 36.550000 0.7835549 34.743119084 38.356881
 [6,]  5.033333 0.9047713  2.946926967  7.119740
 [7,]  2.100000 1.1081140 -0.455315497  4.655315
 [8,]  2.550000 1.1081140 -0.005315497  5.105315
 [9,]  0.450000 1.1081140 -2.105315497  3.005315
```

R: Function for Testing $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$

```
> test=function(lmout,C,d=0){
+    b=coef(lmout)
+    V=vcov(lmout)
+    dfn=nrow(C)
+    dfd=lmout$df
+    Cb.d=C%*%b-d
+    Fstat=drop(
           t(Cb.d)%*%solve(C%*%V%*%t(C))%*%Cb.d/dfn)
+    pvalue=1-pf(Fstat,dfn,dfd)
+    list(Fstat=Fstat,pvalue=pvalue)
+ }
```

Fitting the Additive Model in R

```
> #Test for diet main effect:
> C=matrix(c(
+ 0, 1, 0, 0
+ ), nrow=1, byrow=T)
> C
     [,1] [,2] [,3] [,4]
[1,]    0    1    0    0
> test(o,C)
$Fstat
[1] 30.94808

$pvalue
[1] 0.0005327312
```

Fitting the Additive Model in R

```
> #Test for drug main effects:
> C=matrix(c(
+ 0, 0, -1,  0,
+ 0, 0,  0, -1
+ ), nrow=2, byrow=T)
> C
     [,1] [,2] [,3] [,4]
[1,]    0    0   -1    0
[2,]    0    0    0   -1
> test(o,C)
$Fstat
[1] 3.017306

$pvalue
[1] 0.1055743
```

Another Example Use of the Additive Model

$$
\begin{array}{cc|ccc|}
 & & & \text{movie} & \\
 & & 1 & 2 & 3 \\
\hline
 & 1 & 4 & 1 & ? \\
\text{customer} & 2 & ? & 3 & 5 \\
 & 3 & ? & ? & 3 \\
 & 4 & 3 & 1 & ? \\
\hline
\end{array}
$$

Can we guess ratings for customer/movie combinations not in the dataset?

y_{ij} = customer i's rating of movie j

Which movie is best?

$$y_{ij} = \mu + c_i + m_j + \epsilon_{ij}$$

The Linear Model in Vector and Matrix Form

$$
\begin{bmatrix} 4 \\ 1 \\ 3 \\ 5 \\ 3 \\ 3 \\ 1 \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix} \mu \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix}
+
\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \end{bmatrix}
$$

$$
\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}
$$

Can we estimate means for missing data?

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} \mu + c_1 + m_1 \\ \mu + c_1 + m_2 \\ \mu + c_2 + m_2 \\ \mu + c_2 + m_3 \\ \mu + c_3 + m_3 \\ \mu + c_4 + m_1 \\ \mu + c_4 + m_2 \end{bmatrix}$$

Can we estimate

$\mu + c_1 + m_3$?

$\mu + c_2 + m_1$?

$\mu + c_3 + m_1$?

$\mu + c_3 + m_2$?

$\mu + c_4 + m_3$?

$m_1 - m_2$ is estimable because

$$[1, -1, 0, 0, 0, 0, 0]E(\boldsymbol{y}) = [1, -1, 0, 0, 0, 0, 0]\boldsymbol{X\beta} = m_1 - m_2.$$

Can we estimate means for missing data?

$$X\beta = \begin{bmatrix} \mu + c_1 + m_1 \\ \mu + c_1 + m_2 \\ \mu + c_2 + m_2 \\ \mu + c_2 + m_3 \\ \mu + c_3 + m_3 \\ \mu + c_4 + m_1 \\ \mu + c_4 + m_2 \end{bmatrix}$$

Can we estimate
$\mu + c_1 + m_3$?
$\mu + c_2 + m_1$?
$\mu + c_3 + m_1$?
$\mu + c_3 + m_2$?
$\mu + c_4 + m_3$?

Likewise, $m_2 - m_3$ is estimable because

$$[0, 0, 1, -1, 0, 0, 0]E(\boldsymbol{y}) = [0, 0, 1, -1, 0, 0, 0]\boldsymbol{X\beta} = m_2 - m_3.$$

We can estimate all pairwise differences between movie effects.

Because $m_1 - m_2$ and $m_2 - m_3$ are estimable, we can also estimate

$$(m_1 - m_2) + (m_2 - m_3) = m_1 - m_3.$$

This follows because any linear combination of estimable functions is also estimable.

We can estimate the mean underlying the rating for any combination of customer and movie.

It follows that any linear combination of the form

$$\mu + c_i + m_j$$

can be estimated $\forall\ i = 1, 2, 3, 4$ and $j = 1, 2, 3$ because

$$\mu + c_i + m_j = (\mu + c_i + m_{j^*}) + (m_j - m_{j^*})$$

$$\forall\ i = 1, 2, 3, 4 \text{ and } j, j^* = 1, 2, 3.$$

Movie LSMEANS

If our goal is to compare movies to see which is most highly rated, we can accomplish that by estimating the pairwise differences between movie effects.

However, if we want to retain information about the mean rating rather than the difference between mean ratings, it is natural to consider estimating the average (across *all* customers) of the mean rating for each movie.

Movie LSMEANS

This average for the jth movie is

$$\frac{1}{4}\sum_{i=1}^{4}(\mu + c_i + m_j) = \mu + \frac{1}{4}\sum_{i=1}^{4} c_i + m_j = \mu + \bar{c}_{\cdot} + m_j.$$

This average is estimable for each movie in our example because it is a linear combination of estimable functions.

Estimates of $\mu + \bar{c}_{\cdot} + m_j \ \forall \ j = 1, 2, 3$ are movie LSMEANS.

Suppose we consider a different model.

		movie 1	movie 2	movie 3
customer	1	4	1	?
	2	?	3	5
	3	?	?	3
	4	3	1	?

Can we guess ratings for customer/movie combinations not in the dataset?

y_{ij} = customer i's rating of movie j

Which movie is best?

$$y_{ij} = \mu_{ij} + \epsilon_{ij}$$

The Linear Model in Matrix and Vector Form

$$
\begin{bmatrix} 4 \\ 1 \\ 3 \\ 5 \\ 3 \\ 3 \\ 1 \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{41} \\ \mu_{42} \\ \mu_{43} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \end{bmatrix}
$$

Can we estimate the means that underly the missing table entries? No.

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{22} \\ \mu_{23} \\ \mu_{33} \\ \mu_{41} \\ \mu_{42} \end{bmatrix}$$

Can we estimate

μ_{13}?
μ_{21}?
μ_{31}?
μ_{32}?
μ_{43}?

None of the means underlying missing table entries are estimable under the cell means model.

Movie Ratings Example in R

```
> y=c(4,1,3,5,3,3,1)
>
> X=matrix(c(
+ 1,1,0,0,0,1,0,0,
+ 1,1,0,0,0,0,1,0,
+ 1,0,1,0,0,0,1,0,
+ 1,0,1,0,0,0,0,1,
+ 1,0,0,1,0,0,0,1,
+ 1,0,0,0,1,1,0,0,
+ 1,0,0,0,1,0,1,0
+ ),byrow=T,nrow=7)
```

Computation of P_X

```
> XX=t(X)%*%X
>
> library(MASS)
>
> XXgi=ginv(XX)
>
> Px=X%*%XXgi%*%t(X)
>
> #Px has entries like -1.387779e-16
```

```
> round(Px,2)
      [,1]  [,2] [,3] [,4] [,5]  [,6]  [,7]
[1,]  0.75  0.25    0    0    0  0.25 -0.25
[2,]  0.25  0.75    0    0    0 -0.25  0.25
[3,]  0.00  0.00    1    0    0  0.00  0.00
[4,]  0.00  0.00    0    1    0  0.00  0.00
[5,]  0.00  0.00    0    0    1  0.00  0.00
[6,]  0.25 -0.25    0    0    0  0.75  0.25
[7,] -0.25  0.25    0    0    0  0.25  0.75
```

```
> fractions(Px)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  3/4  1/4    0    0    0  1/4 -1/4
[2,]  1/4  3/4    0    0    0 -1/4  1/4
[3,]    0    0    1    0    0    0    0
[4,]    0    0    0    1    0    0    0
[5,]    0    0    0    0    1    0    0
[6,]  1/4 -1/4    0    0    0  3/4  1/4
[7,] -1/4  1/4    0    0    0  1/4  3/4
```

Computing $P_X y = \hat{y}$

```
> yhat=Px%*%y
>
> yhat
     [,1]
[1,] 3.75
[2,] 1.25
[3,] 3.00
[4,] 5.00
[5,] 3.00
[6,] 3.25
[7,] 0.75
```

One Solution to the Normal Equations

```
> bhat=XXgi%*%t(X)%*%y
>
> bhat
            [,1]
[1,]  1.89473684
[2,]  0.22368421
[3,]  1.97368421
[4,] -0.02631579
[5,] -0.27631579
[6,]  1.63157895
[7,] -0.86842105
[8,]  1.13157895
```

C Matrix for Estimating $\mu + c_i + m_j \;\; \forall \; i, j$

```
> C=matrix(c(
+ 1,1,0,0,0,1,0,0,
+ 1,1,0,0,0,0,1,0,
+ 1,1,0,0,0,0,0,1,
+ 1,0,1,0,0,1,0,0,
+ 1,0,1,0,0,0,1,0,
+ 1,0,1,0,0,0,0,1,
+ 1,0,0,1,0,1,0,0,
+ 1,0,0,1,0,0,1,0,
+ 1,0,0,1,0,0,0,1,
+ 1,0,0,0,1,1,0,0,
+ 1,0,0,0,1,0,1,0,
+ 1,0,0,0,1,0,0,1
+ ),byrow=T,nrow=12)
```

OLS Estimates of $\mu + c_i + m_j \ \forall\ i, j$

```
> Cbhat=C%*%bhat
> Cbhat
      [,1]
 [1,] 3.75
 [2,] 1.25
 [3,] 3.25
 [4,] 5.50
 [5,] 3.00
 [6,] 5.00
 [7,] 3.50
 [8,] 1.00
 [9,] 3.00
[10,] 3.25
[11,] 0.75
[12,] 2.75
```

OLS Estimates of $\mu + c_i + m_j$ and $\mu + \bar{c}_\cdot + m_j \; \forall \; i, j$

```
> M=matrix(Cbhat,nrow=4,byrow=T)
>
> M
     [,1] [,2] [,3]
[1,] 3.75 1.25 3.25
[2,] 5.50 3.00 5.00
[3,] 3.50 1.00 3.00
[4,] 3.25 0.75 2.75
>
> apply(M,2,mean)
[1] 4.0 1.5 3.5
```

OLS Estimates of $m_j - m_{j^*} \;\; \forall \, j \neq j^*$

```
> C=matrix(c(
+ 0,0,0,0,0,1,-1,0,
+ 0,0,0,0,0,1,0,-1,
+ 0,0,0,0,0,0,1,-1
+ ),byrow=T,nrow=3)
>
> Cbhat=C%*%bhat
> Cbhat
     [,1]
[1,]  2.5
[2,]  0.5
[3,] -2.0
```

Response Weights for Estimation of $m_j - m_{j^*} \quad \forall\, j \neq j^*$

```
> round(C%*%XXgi%*%t(X),2)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  0.5 -0.5    0    0    0  0.5 -0.5
[2,]  0.5 -0.5    1   -1    0  0.5 -0.5
[3,]  0.0  0.0    1   -1    0  0.0  0.0
```

		movie		
		1	2	3
customer	1	4	1	?
	2	?	3	5
	3	?	?	3
	4	3	1	?

(Best to make sense of rows 1 and 3 first and then row 2 follows.)

Alternative Analysis Using the R Full-Rank $\boldsymbol{X}$ Matrix

```
> customer=factor(c(1,1,2,2,3,4,4))
> movie=factor(c(1,2,2,3,3,1,2))
> d=data.frame(customer,movie,y)
>
> d
  customer movie y
1        1     1 4
2        1     2 1
3        2     2 3
4        2     3 5
5        3     3 3
6        4     1 3
7        4     2 1
```

The R Full-Rank $\boldsymbol{X}$ Matrix

```
> o=lm(y~customer+movie,data=d)
> model.matrix(o)

  (Intercept) customer2 customer3 customer4 movie2 movie3
1           1         0         0         0      0      0
2           1         0         0         0      1      0
3           1         1         0         0      1      0
4           1         1         0         0      0      1
5           1         0         1         0      0      1
6           1         0         0         1      0      0
7           1         0         0         1      1      0
```

$\hat{\beta}$, $\hat{y}$, and $y - \hat{y}$

```
> coef(o)
(Intercept)  customer2  customer3  customer4    movie2    movie3
       3.75       1.75      -0.25      -0.50     -2.50     -0.50

> fitted(o)
   1    2    3    4    5    6    7
3.75 1.25 3.00 5.00 3.00 3.25 0.75

> resid(o)
            1             2     3     4     5
 2.500000e-01 -2.500000e-01     0     0     0
            6             7
-2.500000e-01  2.500000e-01
```

$\hat{\beta}$, $\hat{y}$, and $y - \hat{y}$

```
> o$coe
(Intercept)  customer2  customer3  customer4   movie2   movie3
       3.75       1.75      -0.25      -0.50    -2.50    -0.50

> o$fit
   1    2    3    4    5    6    7
3.75 1.25 3.00 5.00 3.00 3.25 0.75

> o$res
            1             2     3     4     5
 2.500000e-01 -2.500000e-01     0     0     0
            6             7
-2.500000e-01  2.500000e-01
```

OLS Estimates of $m_j - m_{j^*} \;\; \forall\, j \neq j^*$

```
> -o$coe[5]
movie2
   2.5
> -o$coe[6]
movie3
   0.5
> o$coe[5]-o$coe[6]
movie2
    -2
```

OLS Estimates of $m_j - m_{j^*} \;\; \forall\, j \neq j^*$

```
> C=matrix(c(
+ 0,0,0,0,-1,0,
+ 0,0,0,0,0,-1,
+ 0,0,0,0,1,-1
+ ),byrow=T,nrow=3)
>
> C%*%o$coe
     [,1]
[1,]  2.5
[2,]  0.5
[3,] -2.0
```

C Matrix for Estimating $\mu + c_i + m_j \;\; \forall \; i, j$

```
> C=matrix(c(
+ 1,0,0,0,0,0,
+ 1,0,0,0,1,0,
+ 1,0,0,0,0,1,
+ 1,1,0,0,0,0,
+ 1,1,0,0,1,0,
+ 1,1,0,0,0,1,
+ 1,0,1,0,0,0,
+ 1,0,1,0,1,0,
+ 1,0,1,0,0,1,
+ 1,0,0,1,0,0,
+ 1,0,0,1,1,0,
+ 1,0,0,1,0,1
+ ),byrow=T,nrow=12)
```

OLS Estimates of $\mu + c_i + m_j \;\; \forall \; i, j$

```
> matrix(C%*%o$coe,nrow=4,byrow=T)
     [,1] [,2] [,3]
[1,] 3.75 1.25 3.25
[2,] 5.50 3.00 5.00
[3,] 3.50 1.00 3.00
[4,] 3.25 0.75 2.75
```