# Analysis of Two-Factor Experiments Based on Cell Means 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 pigs. A total of 12 pigs were assigned to the 6 diet $\times$ drug combinations using a balanced and completely randomized experimental design. Pigs 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 pig.

# Factors and Levels

- This experiment involves 2 *factors*: Diet and Drug.
- The factor *Diet* has 2 *levels*: low fiber and high fiber.
- The factor *Drug* has 3 *levels*: D1, D2, and D3.

# Treatment Design vs. Experimental Design

- A combination of one level from each factor forms a *treatment*.
- The *treatment design* used in this experiment is known as a *full-factorial treatment design* because each possible combination of one level from each factor was applied to at least one experimental unit.
- The *experimental design* is a balanced *completely randomized design* (CRD) because all possible balanced assignments of the 12 pigs to the 6 treatment groups were equally likely.

# Table of Treatments

| Treatment | Diet | Drug |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 1 | 3 |
| 4 | 2 | 1 |
| 5 | 2 | 2 |
| 6 | 2 | 3 |

Diet 1 = Low Fiber, Diet 2 = High Fiber

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

# The Cell Means Model

For $i = 1, 2;\ j = 1, 2, 3;$ and $k = 1, 2;$ let $y_{ijk}$ denote the weight gain of the $k^{\text{th}}$ pig that received diet $i$ and drug $j$, and suppose

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \text{ where}$$

$\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}$ 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 Cell Means Model

| Treatment | Diet | Drug | Mean |
|---|---|---|---|
| 1 | 1 | 1 | $\mu_{11}$ |
| 2 | 1 | 2 | $\mu_{12}$ |
| 3 | 1 | 3 | $\mu_{13}$ |
| 4 | 2 | 1 | $\mu_{21}$ |
| 5 | 2 | 2 | $\mu_{22}$ |
| 6 | 2 | 3 | $\mu_{23}$ |

Diet 1 = Low Fiber, Diet 2 = High Fiber

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

# Table of Means for the Cell Means Model

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ |

# Same Model with Simpler Notation

For $t = 1, 2, 3, 4, 5, 6$ and $k = 1, 2$; let $y_{tk}$ denote the weight gain of the $k^{\text{th}}$ pig that received treatment $t$, and suppose

$$y_{tk} = \mu_t + \epsilon_{tk} \text{ where}$$

$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ are unknown real-valued parameters and

$$\epsilon_{11}, \epsilon_{12}, \epsilon_{21}, \epsilon_{22}, \epsilon_{31}, \epsilon_{32}, \epsilon_{41}, \epsilon_{42}, \epsilon_{51}, \epsilon_{52}, \epsilon_{61}, \epsilon_{62} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

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

# Table of Treatments and Means

| Treatment | Diet | Drug | Mean |
|---|---|---|---|
| 1 | 1 | 1 | $\mu_1$ |
| 2 | 1 | 2 | $\mu_2$ |
| 3 | 1 | 3 | $\mu_3$ |
| 4 | 2 | 1 | $\mu_4$ |
| 5 | 2 | 2 | $\mu_5$ |
| 6 | 2 | 3 | $\mu_6$ |

Diet 1 = Low Fiber, Diet 2 = High Fiber

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

# Table of Means for the Cell Means Model with Single Subscripting

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_1$ | $\mu_2$ | $\mu_3$ |
| Diet 2 | $\mu_4$ | $\mu_5$ | $\mu_6$ |

The different notations for specifying the cell means model are equally valid.

Double subscripting for treatments is typically used to better track the levels of each factor that make up each treatment.

# Cell Means Model in Matrix and Vector Form

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# Cell Means Model in Matrix and Vector Form

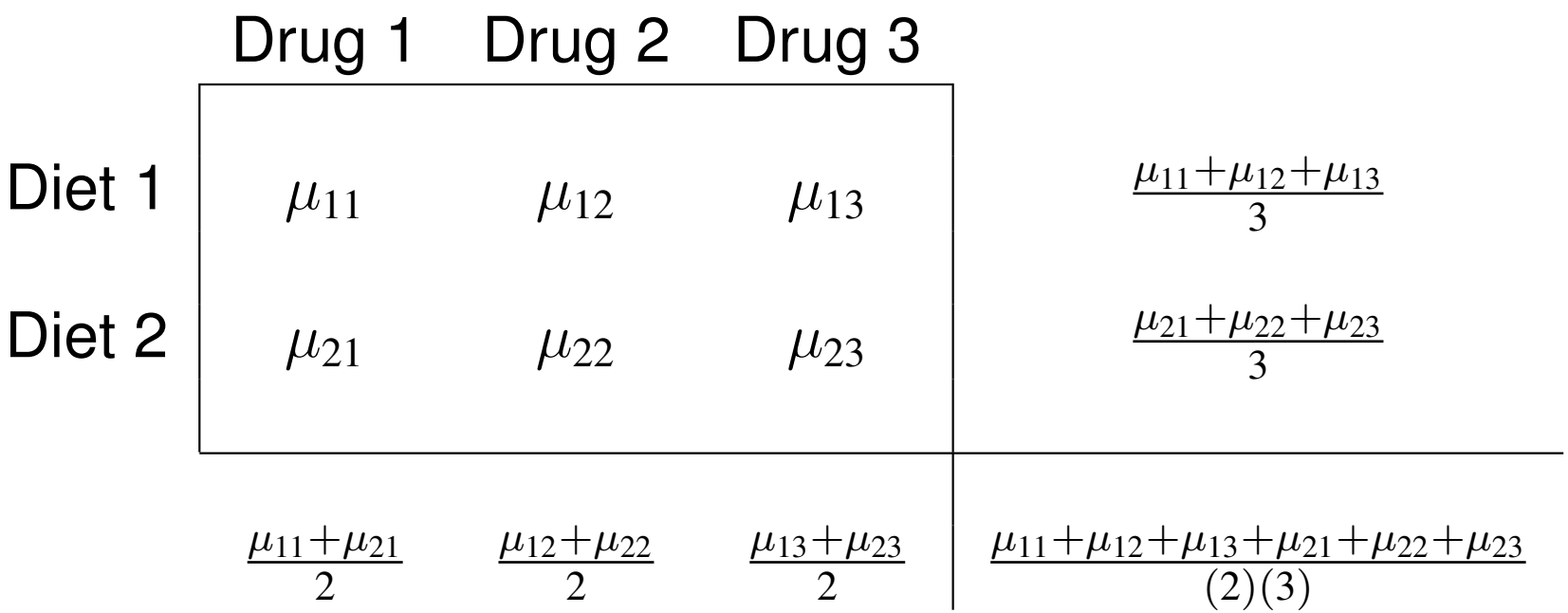
$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{11} \\ \mu_{12} \\ \mu_{12} \\ \mu_{13} \\ \mu_{13} \\ \mu_{21} \\ \mu_{21} \\ \mu_{22} \\ \mu_{22} \\ \mu_{23} \\ \mu_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# Table of Cell Means

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ |

# Marginal Means



# Marginal Means

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ | $\bar{\mu}_{1.}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ | $\bar{\mu}_{2.}$ |
| | $\bar{\mu}_{.1}$ | $\bar{\mu}_{.2}$ | $\bar{\mu}_{.3}$ | $\bar{\mu}_{..}$ |

$$\bar{\mu}_{i.} = \tfrac{1}{3}\sum_{j=1}^{3}\mu_{ij}$$

$$\bar{\mu}_{.j} = \frac{1}{2}\sum_{i=1}^{2}\mu_{ij} \qquad \bar{\mu}_{..} = \frac{1}{(2)(3)}\sum_{i=1}^{2}\sum_{j=1}^{3}\mu_{ij}$$

# Least Squares Means (LSMEANS) in SAS

SAS can be used to compute LSMEANS.

LSMEANS are simply OLS estimators of cell or marginal means.

Each LSMEAN has the form $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ for an appropriate vector $\boldsymbol{c}$.

For example, the LSMEAN for diet 1 is $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ with

$$\boldsymbol{c}' = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0\right] \text{ and } \hat{\boldsymbol{\beta}} = [\bar{y}_{11\cdot}, \bar{y}_{12\cdot}, \bar{y}_{13\cdot}, \bar{y}_{21\cdot}, \bar{y}_{22\cdot}, \bar{y}_{23\cdot}]'.$$

# LSMEANS (continued)

Note that the LSMEAN for Diet 1 is simply an average of the estimated means for treatments involving Diet 1.

When data are balanced, the LSMEAN for Diet 1 is also just the average of responses for all pigs that were fed Diet 1.

When data are unbalanced, the LSMEAN for Diet 1 may not equal the average of responses for all pigs that were fed Diet 1.

# Standard Error for an LSMEAN

A *standard error* is the estimated standard deviation of a statistic.

A *standard error* is usually found by estimating the variance of a statistic and then taking the square root of the estimate.

Because each LSMEAN has the form $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ for an appropriate vector $\boldsymbol{c}$, the standard error for an LSMEAN is given by

$$\sqrt{\widehat{\mathrm{Var}}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})} = \sqrt{\hat{\sigma}^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}.$$

# Simple Effects

A *simple effect* is the difference between cell means that differ in level for only one factor.

In our two-factor example, *simple effects* are differences between cell means within any row or within any column.

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ |

# Simple Effects (continued)

The *simple effect* of Diet for Drug 1 is $\mu_{11} - \mu_{21}$.

The *simple effect* of Drug 2 vs. Drug 3 for Diet 2 is $\mu_{22} - \mu_{23}$.

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ |

# Main Effects

A *main effect* is the difference between marginal means associated with two levels of a factor.

In our two-factor example, the *main effect* of Diet is $\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}$.

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ | $\bar{\mu}_{1\cdot}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ | $\bar{\mu}_{2\cdot}$ |
| | $\bar{\mu}_{\cdot 1}$ | $\bar{\mu}_{\cdot 2}$ | $\bar{\mu}_{\cdot 3}$ | $\bar{\mu}_{\cdot\cdot}$ |

# Main Effects (continued)

In our two-factor example, the *main effects* of Drug involve the differences $\bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 2}$, $\bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 3}$, and $\bar{\mu}_{\cdot 2} - \bar{\mu}_{\cdot 3}$.

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ | $\bar{\mu}_{1\cdot}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ | $\bar{\mu}_{2\cdot}$ |
| | $\bar{\mu}_{\cdot 1}$ | $\bar{\mu}_{\cdot 2}$ | $\bar{\mu}_{\cdot 3}$ | $\bar{\mu}_{\cdot\cdot}$ |

# Main Effects (continued)

If $\bar{\mu}_{1\cdot} = \bar{\mu}_{2\cdot}$, it would be customary to say, "There is no Diet main effect."

If $\bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2} = \bar{\mu}_{\cdot 3}$, it would be customary to say, "There are no Drug main effects."

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ | $\bar{\mu}_{1\cdot}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ | $\bar{\mu}_{2\cdot}$ |
| | $\bar{\mu}_{\cdot 1}$ | $\bar{\mu}_{\cdot 2}$ | $\bar{\mu}_{\cdot 3}$ | $\bar{\mu}_{\cdot\cdot}$ |

# Interaction Effects

The linear combination $\mu_{ij} - \mu_{ij^*} - \mu_{i^*j} + \mu_{i^*j^*}$ for $i \neq i^*$ and $j \neq j^*$ is an *interaction effect*.

For example,

$$\mu_{11}-\mu_{12}-\mu_{21}+\mu_{22} = (\mu_{11}-\mu_{12})-(\mu_{21}-\mu_{22}) = (\mu_{11}-\mu_{21})-(\mu_{12}-\mu_{22})$$

is an interaction effect.

| | Drug 1 | Drug 2 | Drug 3 |
|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ |
| Diet 2 | $\mu_{21}$ | $\mu_{22}$ | $\mu_{23}$ |

# Interaction Effects (continued)

When all interaction effects are zero, we may say there are “no interactions” between the factors or that the two factors do not “interact.”

When there are no interactions between factors, the simple effects of either factor are the same across all levels of the other factor.

For example, when there are no interactions between the factors Diet and Drug, the simple effect of Diet is the same for each level of Drug. Likewise, any simple effect of Drug is the same for both Diets.

# Testing for Non-Zero Effects

We can test whether simple effects, main effects, or interaction effects are zero vs. non-zero using tests of the form

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0} \text{ vs. } H_A : \boldsymbol{C}\boldsymbol{\beta} \neq \mathbf{0}.$$

The following slides give appropriate $\boldsymbol{C}$ matrices for several examples.

# $H_0$: No simple effect of Diet for Drug 1 $(\mu_{11} = \mu_{21})$

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

# $H_0$: No simple effect of Drug 2 vs. Drug 3 for Diet 2 $(\mu_{22} = \mu_{23})$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

# $H_0$: No Diet Main Effect $(\bar{\mu}_{1\cdot} = \bar{\mu}_{2\cdot})$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

# $H_0$: No Drug Main Effects $(\bar{\mu}_{.1} = \bar{\mu}_{.2} = \bar{\mu}_{.3})$

$$\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

# $H_0$: No Drug Main Effects $(\bar{\mu}_{.1} = \bar{\mu}_{.2} = \bar{\mu}_{.3})$

Here is an alternative specification that will yield the same test.

$$\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

# $H_0$: No Drug Main Effects $(\bar{\mu}_{.1} = \bar{\mu}_{.2} = \bar{\mu}_{.3})$

Here is another alternative specification that will yield the same test.

$$\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

# $H_0$: No Diet-by-Drug Interactions

$(\mu_{ij} - \mu_{ij^*} - \mu_{i^*j} + \mu_{i^*j^*} = 0$ for all $i \neq i^*$ and $j \neq j^*)$

$$\begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

# SAS Code

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

proc print data=d;
run;
```

# SAS Output

| Obs | diet | drug | weightgain |
|---|---|---|---|
| 1 | 1 | 1 | 41.3 |
| 2 | 1 | 1 | 43.7 |
| 3 | 1 | 2 | 40.9 |
| 4 | 1 | 2 | 39.2 |
| 5 | 1 | 3 | 37.4 |
| 6 | 1 | 3 | 37.9 |
| 7 | 2 | 1 | 36.8 |
| 8 | 2 | 1 | 34.6 |
| 9 | 2 | 2 | 33.6 |
| 10 | 2 | 2 | 34.3 |
| 11 | 2 | 3 | 35.8 |
| 12 | 2 | 3 | 35.1 |

# SAS Code

```
proc mixed;
  class diet drug;
  model weightgain=diet*drug / noint;
  lsmeans diet*drug;
```

# SAS Output

Least Squares Means

| Effect | diet | drug | Estimate | Standard Error | DF | t Value |
|---|---|---|---|---|---|---|
| diet*drug | 1 | 1 | 42.5000 | 0.7832 | 6 | 54.27 |
| diet*drug | 1 | 2 | 40.0500 | 0.7832 | 6 | 51.14 |
| diet*drug | 1 | 3 | 37.6500 | 0.7832 | 6 | 48.07 |
| diet*drug | 2 | 1 | 35.7000 | 0.7832 | 6 | 45.58 |
| diet*drug | 2 | 2 | 33.9500 | 0.7832 | 6 | 43.35 |
| diet*drug | 2 | 3 | 35.4500 | 0.7832 | 6 | 45.27 |

# SAS Code (proc mixed statement continued)

```
estimate 'lsmean for diet 1'
         diet*drug 1 1 1 0 0 0 /divisor=3;

estimate 'simple effect of diet for drug 1'
         diet*drug 1 0 0 -1 0 0;

estimate 'simple effect drug 2 vs. drug 3 for diet 2'
         diet*drug 0 0 0 0 1 -1;

estimate 'diet main effect'
         diet*drug 1 1 1 -1 -1 -1 / divisor=3;
```

# SAS Output

Estimates

| Label | Estimate | Standard Error | DF | t Value |
|---|---|---|---|---|
| lsmean for diet 1 | 40.0667 | 0.4522 | 6 | 88.61 |
| simple effect of diet for drug 1 | 6.8000 | 1.1075 | 6 | 6.14 |
| simple effect drug 2 vs. drug 3 for diet 2 | -1.5000 | 1.1075 | 6 | -1.35 |
| diet main effect | 5.0333 | 0.6394 | 6 | 7.87 |

Estimates

| Label | Pr > \|t\| |
|---|---|
| lsmean for diet 1 | <.0001 |
| simple effect of diet for drug 1 | 0.0009 |
| simple effect of drug 2 vs. drug 3 for diet 2 | 0.2244 |
| diet main effect | 0.0002 |

# SAS Code (proc mixed statement continued)

```
contrast 'simple effect of diet for drug 1'
         diet*drug 1 0 0 -1 0 0;
contrast 'simple effect drug 2 vs. drug 3 for diet 2'
         diet*drug 0 0 0 0 1 -1;
contrast 'diet main effect'
         diet*drug 1 1 1 -1 -1 -1;
contrast 'drug main effects'
         diet*drug 1 -1  0  1 -1  0,
         diet*drug 1  0 -1  1  0 -1;
```

# SAS Code (proc mixed statement continued)

```
contrast 'drug main effects also'
         diet*drug 1 -1  0  1 -1  0,
         diet*drug 0  1 -1  0  1 -1;
contrast 'drug main effects also2'
         diet*drug 2 -2  0  2 -2  0,
         diet*drug 1  1 -2  1  1 -2;
contrast 'diet-by-drug interactions'
         diet*drug 1 -1  0 -1  1  0,
         diet*drug 1  0 -1 -1  0  1;
run;
```

# SAS Output

Contrasts

| Label | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| simple effect of diet for drug 1 | 1 | 6 | 37.70 | 0.0009 |
| simple effect drug 2 vs. drug 3 for diet 2 | 1 | 6 | 1.83 | 0.2244 |
| diet main effect | 1 | 6 | 61.96 | 0.0002 |
| drug main effects | 2 | 6 | 6.04 | 0.0365 |
| drug main effects also | 2 | 6 | 6.04 | 0.0365 |
| drug main effects also2 | 2 | 6 | 6.04 | 0.0365 |
| diet-by-drug interactions | 2 | 6 | 5.01 | 0.0526 |

# Fitting the Cell Means Model in R

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

d$diet=factor(d$diet)

d$drug=factor(d$drug)

o=lm(weightgain~0+drug:diet,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 $n - r$.

# An Alternative Parameterization

An alternative parameterization of the cell-means model is

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

$$\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{22}, \gamma_{23}$$

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

| Treatment | Diet | Drug | Mean |
|---|---|---|---|
| 1 | 1 | 1 | $\mu + \alpha_1 + \beta_1 + \gamma_{11}$ |
| 2 | 1 | 2 | $\mu + \alpha_1 + \beta_2 + \gamma_{12}$ |
| 3 | 1 | 3 | $\mu + \alpha_1 + \beta_3 + \gamma_{13}$ |
| 4 | 2 | 1 | $\mu + \alpha_2 + \beta_1 + \gamma_{21}$ |
| 5 | 2 | 2 | $\mu + \alpha_2 + \beta_2 + \gamma_{22}$ |
| 6 | 2 | 3 | $\mu + \alpha_2 + \beta_3 + \gamma_{23}$ |

Diet 1 = Low Fiber, Diet 2 = High Fiber

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

# Model in Matrix and Vector Form

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# Model in Matrix and Vector Form

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 + \beta_1 + \gamma_{11} \\ \mu + \alpha_1 + \beta_1 + \gamma_{11} \\ \mu + \alpha_1 + \beta_2 + \gamma_{12} \\ \mu + \alpha_1 + \beta_2 + \gamma_{12} \\ \mu + \alpha_1 + \beta_3 + \gamma_{13} \\ \mu + \alpha_1 + \beta_3 + \gamma_{13} \\ \mu + \alpha_2 + \beta_1 + \gamma_{21} \\ \mu + \alpha_2 + \beta_1 + \gamma_{21} \\ \mu + \alpha_2 + \beta_2 + \gamma_{22} \\ \mu + \alpha_2 + \beta_2 + \gamma_{22} \\ \mu + \alpha_2 + \beta_3 + \gamma_{23} \\ \mu + \alpha_2 + \beta_3 + \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# Table of Cell and Marginal Means

Any linear combination of the entries in this table is estimable.

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu + \alpha_1 + \beta_1 + \gamma_{11}$ | $\mu + \alpha_1 + \beta_2 + \gamma_{12}$ | $\mu + \alpha_1 + \beta_3 + \gamma_{13}$ | $\mu + \alpha_1 + \bar{\beta}. + \bar{\gamma}_1.$ |
| Diet 2 | $\mu + \alpha_2 + \beta_1 + \gamma_{21}$ | $\mu + \alpha_2 + \beta_2 + \gamma_{22}$ | $\mu + \alpha_2 + \beta_3 + \gamma_{23}$ | $\mu + \alpha_2 + \bar{\beta}. + \bar{\gamma}_2.$ |
| | $\mu + \bar{\alpha}. + \beta_1 + \bar{\gamma}_{.1}$ | $\mu + \bar{\alpha}. + \beta_2 + \bar{\gamma}_{.2}$ | $\mu + \bar{\alpha}. + \beta_3 + \bar{\gamma}_{.3}$ | $\mu + \bar{\alpha}. + \bar{\beta}. + \bar{\gamma}..$ |

# Example Estimable Functions

- Simple effect of Diet for Drug 1: $\alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{21}$
- Simple effect of Drug 1 vs. Drug 3 for Diet 2:
  $\beta_1 - \beta_3 + \gamma_{21} - \gamma_{23}$
- Main effect of Diet: $\alpha_1 - \alpha_2 + \bar{\gamma}_{1\cdot} - \bar{\gamma}_{2\cdot}$
- Interaction effect involving Diets 1 and 2 and Drugs 1 and 3:
  $$\begin{aligned}&[(\mu + \alpha_1 + \beta_1 + \gamma_{11}) - (\mu + \alpha_2 + \beta_1 + \gamma_{21})] \\ -&[(\mu + \alpha_1 + \beta_3 + \gamma_{13}) - (\mu + \alpha_2 + \beta_3 + \gamma_{23})] \\ &= \gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23}\end{aligned}$$

# Estimation and Testing

As before, estimation or testing involves finding an appropriate matrix $\boldsymbol{C}$ to estimate $\boldsymbol{C\beta}$ or test $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$.

For example, the Diet main effect is estimated by $\boldsymbol{C}\hat{\boldsymbol{\beta}}$, where

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} & \hat{\alpha}_1 & \hat{\alpha}_2 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 & \hat{\gamma}_{11} & \hat{\gamma}_{12} & \hat{\gamma}_{13} & \hat{\gamma}_{21} & \hat{\gamma}_{22} & \hat{\gamma}_{23} \end{bmatrix}'$$

is any solution to the Normal Equations involving the new matrix $\boldsymbol{X}$.

# Estimation and Testing (continued)

As another example, the test for Drug main effects can be carried out as a test of $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}$, where

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & \frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

so that

$$\boldsymbol{C}\boldsymbol{\beta} = \begin{bmatrix} \beta_1 - \beta_2 + \bar{\gamma}_{\cdot 1} - \bar{\gamma}_{\cdot 2} \\ \beta_1 - \beta_3 + \bar{\gamma}_{\cdot 1} - \bar{\gamma}_{\cdot 3} \end{bmatrix}.$$

# SAS Code

```
proc mixed;
  class diet drug;
  model weightgain=diet drug diet*drug;
  lsmeans diet drug diet*drug;
```

# SAS Output

Least Squares Means

| Effect | diet | drug | Estimate | Standard Error | DF | t Value |
|---|---|---|---|---|---|---|
| diet | 1 | | 40.0667 | 0.4522 | 6 | 88.61 |
| diet | 2 | | 35.0333 | 0.4522 | 6 | 77.48 |
| drug | | 1 | 39.1000 | 0.5538 | 6 | 70.61 |
| drug | | 2 | 37.0000 | 0.5538 | 6 | 66.81 |
| drug | | 3 | 36.5500 | 0.5538 | 6 | 66.00 |
| diet*drug | 1 | 1 | 42.5000 | 0.7832 | 6 | 54.27 |
| diet*drug | 1 | 2 | 40.0500 | 0.7832 | 6 | 51.14 |
| diet*drug | 1 | 3 | 37.6500 | 0.7832 | 6 | 48.07 |
| diet*drug | 2 | 1 | 35.7000 | 0.7832 | 6 | 45.58 |
| diet*drug | 2 | 2 | 33.9500 | 0.7832 | 6 | 43.35 |
| diet*drug | 2 | 3 | 35.4500 | 0.7832 | 6 | 45.27 |

# SAS Code (proc mixed statement continued)

```
estimate 'lsmean for diet 1'
          intercept 3 diet 3 0 drug 1 1 1
diet*drug 1 1 1 0 0 0 /divisor=3;

estimate 'simple effect of diet for drug 1'
          diet 1 -1 diet*drug 1 0 0 -1 0 0;

estimate 'simple effect drug 2 vs. drug 3 for diet 2'
          drug 0 1 -1 diet*drug 0 0 0 0 1 -1;

estimate 'diet main effect'
        diet 3 -3 diet*drug 1 1 1 -1 -1 -1/divisor=3;
```

# SAS Output

Estimates

| Label | Estimate | Standard Error | DF | t Value |
|---|---|---|---|---|
| lsmean for diet 1 | 40.0667 | 0.4522 | 6 | 88.61 |
| simple effect of diet for drug 1 | 6.8000 | 1.1075 | 6 | 6.14 |
| simple effect drug 2 vs. drug 3 for diet 2 | -1.5000 | 1.1075 | 6 | -1.35 |
| diet main effect | 5.0333 | 0.6394 | 6 | 7.87 |

Estimates

| Label | Pr > \|t\| |
|---|---|
| lsmean for diet 1 | <.0001 |
| simple effect of diet for drug 1 | 0.0009 |
| simple effect of drug 2 vs. drug 3 for diet 2 | 0.2244 |
| diet main effect | 0.0002 |

# SAS Code (proc mixed statement continued)

```
contrast 'simple effect of diet for drug 1'
         diet 1 -1 diet*drug 1 0 0 -1 0 0;
contrast 'simple effect drug 2 vs. drug 3 for diet 2'
         drug 0 1 -1 diet*drug 0 0 0 0 1 -1;
contrast 'diet main effect'
         diet 3 -3 diet*drug 1 1 1 -1 -1 -1;
contrast 'drug main effect'
         drug 2 -2  0 diet*drug 1 -1  0  1 -1  0,
         drug 2  0 -2 diet*drug 1  0 -1  1  0 -1;
contrast 'diet-by-drug interactions'
         diet*drug 1 -1  0 -1  1  0,
         diet*drug 1  0 -1 -1  0  1;
run;
```

# SAS Output

Contrasts

| Label | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| simple effect of diet for drug 1 | 1 | 6 | 37.70 | 0.0009 |
| simple effect drug 2 vs. drug 3 for diet 2 | 1 | 6 | 1.83 | 0.2244 |
| diet main effect | 1 | 6 | 61.96 | 0.0002 |
| drug main effect | 2 | 6 | 6.04 | 0.0365 |
| diet-by-drug interactions | 2 | 6 | 5.01 | 0.0526 |

# The model matrix $\boldsymbol{X}$ is not full-column rank.

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# The Alternative Parameterization in R

The `lm` function in R uses a full-rank model matrix.

The code

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

will fit the following full-rank version of the cell means model.

# The Full-Rank Formulation Used by R

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# The Full-Rank Formulation Used by R

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \mu + \beta_2 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu + \beta_3 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + \alpha_2 + \beta_2 + \gamma_{22} \\ \mu + \alpha_2 + \beta_2 + \gamma_{22} \\ \mu + \alpha_2 + \beta_3 + \gamma_{23} \\ \mu + \alpha_2 + \beta_3 + \gamma_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \end{bmatrix}$$

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

# Table of Means for the R Full-Rank Formulation

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu$ | $\mu + \beta_2$ | $\mu + \beta_3$ | $\mu + \frac{\beta_2+\beta_3}{3}$ |
| Diet 2 | $\mu + \alpha_2$ | $\mu + \alpha_2 + \beta_2 + \gamma_{22}$ | $\mu + \alpha_2 + \beta_3 + \gamma_{23}$ | $\mu + \alpha_2 + \frac{\beta_2+\beta_3}{3} + \frac{\gamma_{22}+\gamma_{23}}{3}$ |
| | $\mu + \frac{\alpha_2}{2}$ | $\mu + \frac{\alpha_2}{2} + \beta_2 + \frac{\gamma_{22}}{2}$ | $\mu + \frac{\alpha_2}{2} + \beta_3 + \frac{\gamma_{23}}{2}$ | $\mu + \frac{\alpha_2}{2} + \frac{\beta_2+\beta_3}{3} + \frac{\gamma_{22}+\gamma_{23}}{6}$ |

# Parameter Interpretation for R Full-Rank Formulation

- $\mu$ is the mean for the Diet 1 Drug 1 treatment.
- $\alpha_2$ is the simple effect of Diet 2 vs. Diet 1 for Drug 1.
- $\beta_2$ is the simple effect of Drug 2 vs. Drug 1 for Diet 1.
- $\beta_3$ is the simple effect of Drug 3 vs. Drug 1 for Diet 1.
- $\gamma_{22}$ is an interaction effect involving Diets 1 and 2 and Drugs 1 and 2. It tells how much bigger the Diet 2 $-$ Diet 1 mean difference is for Drug 2 than it is for Drug 1.
- $\gamma_{23}$ is an interaction effect involving Diets 1 and 2 and Drugs 1 and 3. It tells how much bigger the Diet 2 $-$ Diet 1 mean difference is for Drug 3 than it is for Drug 1.

# Estimation and Testing in R

As before, estimation or testing involves finding an appropriate matrix $\boldsymbol{C}$ to estimate $\boldsymbol{C\beta}$ or test $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$.

For example, the Diet main effect is estimated by

$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_R = \begin{bmatrix} 0 & -1 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\gamma}_{22} \\ \hat{\gamma}_{23} \end{bmatrix} = -\hat{\alpha}_2 - \frac{\hat{\gamma}_{22} + \hat{\gamma}_{23}}{3},$$

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

# Estimation and Testing in R (continued)

As another example, the test for Drug main effects can be carried out as a test of $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}$, where

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 1 & 0 & \frac{1}{2} & 0 \\ \\ 0 & 0 & 0 & 1 & 0 & \frac{1}{2} \end{bmatrix}$$

so that

$$\boldsymbol{C}\boldsymbol{\beta} = \begin{bmatrix} \beta_2 + \frac{\gamma_{22}}{2} \\ \\ \beta_3 + \frac{\gamma_{23}}{2} \end{bmatrix}.$$

# Tests Based on Reduced vs. Full Model Comparison

Any of the tests we have discussed could alternatively be carried out using a statistic of the form

$$F = \frac{\boldsymbol{y}'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{y} / \left[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)\right]}{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y} / \left[n - \text{rank}(\boldsymbol{X})\right]}$$

for an appropriate reduced model matrix $\boldsymbol{X}_0$.

Finding $\boldsymbol{X}_0$ is straightforward for tests involving simple effects.

The next set of slides shows example $\boldsymbol{X}_0$ matrices for testing the null hypothesis of no interactions.

It is possible to conduct the tests for main effects we have discussed using a reduced vs. full models comparison, but specification of the reduced model ($\boldsymbol{X}_0$) is not so easy.

# Testing for Main Effects When Factors Interact

Some statisticians argue against testing for main effects when there are interactions between factors.

Others believe that, depending on the scientific questions of interest, any contrasts of treatment means may be worth examining.

Be aware that "no main effects" does not necessarily mean "no effects."

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11} = 42$ | $\mu_{12} = 40$ | $\mu_{13} = 38$ | $\bar{\mu}_{1\cdot} = 40$ |
| Diet 2 | $\mu_{21} = 38$ | $\mu_{22} = 40$ | $\mu_{23} = 42$ | $\bar{\mu}_{2\cdot} = 40$ |
| | $\bar{\mu}_{\cdot 1} = 40$ | $\bar{\mu}_{\cdot 2} = 40$ | $\bar{\mu}_{\cdot 3} = 40$ | $\bar{\mu}_{\cdot\cdot} = 40$ |

# Unbalanced Data and Missing Cells

Although we have focused on a balanced two-factor experiment with 2 experimental units per treatment, the techniques presented in these slides work the same way whether data are balanced or not, as long as each treatment has a response for at least one experimental unit and some treatments have more than one.

If there are no experimental units for one or more treatments, then the treatment design may not be a full-factorial treatment design, and we may have a *missing cell* or *missing cells*.

# Example with Missing Cells

| | Drug 1 | Drug 2 | Drug 3 | |
|---|---|---|---|---|
| Diet 1 | $\mu_{11}$ | $\mu_{12}$ | $\mu_{13}$ | $\bar{\mu}_{1\cdot}$ |
| Diet 2 | $\mu_{21}$ | - | $\mu_{23}$ | - |
| | $\bar{\mu}_{\cdot 1}$ | - | $\bar{\mu}_{\cdot 3}$ | - |