4. Analysis of Two-Factor Experiments Based on Cell-Means Models

An Example Two-Factor Experiment

response y

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.

treatment

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

$i = 1, 2$ diet

$j = 1, 2, 3$ drug

$k = 1, 2$ replication

Diet 1 = Low Fiber, Diet 2 = High Fiber

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

The Cell-Means Model

diet   *drug*   *pig*

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

$$
\begin{aligned}
y_{ijk} &= \mu_{ij} + \epsilon_{ijk}, \\
\epsilon_{ijk} &\overset{iid}{\sim} \mathcal{N}(0, \sigma^2),
\end{aligned}
$$

GMM NE

where $\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23} \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$ are unknown parameters.

treatment means

Table of Treatments and Means for Cell-Means Model

Treatment	Diet	Drug	Mean
1	1	1	μ_{11}
2	1	2	μ_{12}
3	1	3	μ_{13}
4	2	1	μ_{21}
5	2	2	μ_{22}
6	2	3	μ_{23}

Diet 1 = Low Fiber,
Diet 2 = High Fiber

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

an estimate of μ_{11}
would be

$$\frac{y_{111} + y_{112}}{2}$$

Cell means
table

	Drug 1	Drug 2	Drug 3
Diet 1	μ_{11}	μ_{12}	μ_{13}
Diet 2	μ_{21}	μ_{22}	μ_{23}

$$\frac{y_{231} + y_{232}}{2}$$

is an estimate of μ_{23}

Cell-Means Model in Matrix and Vector Form

one column for each treatment combination ij

responses y_{ijk}

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

$\boldsymbol{y}$ $\boldsymbol{X}$ $\boldsymbol{\beta}$

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

Each Cell Mean is Estimable

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{11} \\ \mu_{12} \\ \mu_{12} \\ \mu_{13} \\ \mu_{13} \\ \mu_{21} \\ \mu_{21} \\ \mu_{22} \\ \mu_{22} \\ \mu_{23} \\ \mu_{23} \end{bmatrix}$$

$E(y)$ for each of the 12 pigs

For the General Linear Model, the parameter vector $\boldsymbol{\beta}$ is estimable whenever $\boldsymbol{X}$ has full-column rank, i.e., whenever $\mathrm{rank}(\boldsymbol{X}) = p$.

Finding the OLS Estimator of β

$$\hat{\beta} = (X^\top X)^- X^\top y$$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \implies X^\top X = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$(a X)^- = (a I)^-$$

$$\Downarrow = \frac{1}{a} I^-$$

$$\implies X^\top X = 2 I_{6 \times 6}$$

unique inverse

$$\implies \boxed{(X^\top X)^{-1}} = \frac{1}{2} I_{6 \times 6}$$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \; y = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{bmatrix} \implies \boxed{X^\top y} = \begin{bmatrix} y_{11\cdot} \\ y_{12\cdot} \\ y_{13\cdot} \\ y_{21\cdot} \\ y_{22\cdot} \\ y_{23\cdot} \end{bmatrix}$$

$y_{11\cdot} = \boxed{y_{111} + y_{112}}$

Summing over $k = 1, 2$

dividing by 2

$$\therefore \; \widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top y = \frac{1}{2} I X^\top y = \left(\frac{1}{2}\right) X^\top y = \begin{bmatrix} \bar{y}_{11\cdot} \\ \bar{y}_{12\cdot} \\ \bar{y}_{13\cdot} \\ \bar{y}_{21\cdot} \\ \bar{y}_{22\cdot} \\ \bar{y}_{23\cdot} \end{bmatrix}$$

Table of Cell Means

μ_{ij}

$$\hat{\mu}_{11} = \overline{y}_{11.}$$

We can estimate each μ_{ij} by averaging all available replicates of treatment ij

	Drug 1	Drug 2	Drug 3
Diet 1	μ_{11}	μ_{12}	μ_{13}
Diet 2	μ_{21}	μ_{22}	μ_{23}

$\frac{1}{3}(\mu_{11} + \mu_{12} + \mu_{13}) = \overline{\mu}_{1.}$ // averag. over drug

$\overline{\mu}_{2.}$ — marginal means associated with diet

$\frac{1}{2}(\mu_{11} + \mu_{21})$ = $\overline{\mu}_{.1}$ $\overline{\mu}_{.2}$ $\overline{\mu}_{.3}$

marginal means / associated with drug

overall mean $\overline{\mu}_{..}$

associated = $\frac{1}{6}(\mu_{11} + \mu_{12} + \mu_{13} + \mu_{21} + \mu_{22} + \mu_{23})$

Marginal Means

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

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

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

Least Squares Means (LSMEANS) in SAS

Order of parameters in β determines set-up of $c^\top$ or C

$$\beta = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix} \Rightarrow \hat{\beta} \begin{pmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \vdots \\ \bar{y}_{23.} \end{pmatrix}$$

SAS can be used to compute LSMEANS.

LSMEANS are simply OLS estimators of cell or marginal means.

Each LSMEAN has the form $c^\top \widehat{\beta}$ for an appropriate vector c.

is estimator for $\bar{\mu}_{1.}$

For example, the LSMEAN for diet 1 is $\boxed{c^\top \widehat{\beta}}$ with

diet 2 means

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

$= \dfrac{\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.}}{3}$ estimator of marginal mean $\bar{\mu}_{1.}$

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

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 $c^\top \widehat{\beta}$ for an appropriate vector c, the standard error for an LSMEAN is given by

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

estimated - usually comes
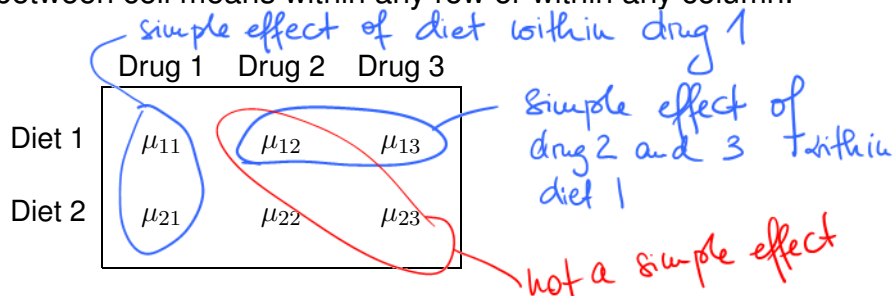from output (SAS, R) = MSE

effects are functions of cell- or marginal means

- Simple Effects

- Main Effects

- Interactions

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	μ_{11}	μ_{12}	μ_{13}
Diet 2	μ_{21}	μ_{22}	μ_{23}

(handwritten annotations:)

simple effect of diet within drug 1

simple effect of drug 2 and 3 within diet 1

not a simple effect

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	μ_{11}	μ_{12}	μ_{13}
Diet 2	μ_{21}	μ_{22}	μ_{23}

$\bar{y}_{11\cdot} - \bar{y}_{21\cdot}$ = estimated

simple effect

lecture 6

02/03/25