7. Analysis of Variance for Balanced Two-Factor Experiments

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-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.

A Model for the Data

always begin by writing out the model

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

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

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

are unknown real-valued parameters and

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

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

Model in Matrix and Vector Form

less than full rank model matrix

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

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

Same Model with Cell Means Parameterization

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

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

We could consider a sequence of progressively more complex models for the response mean that lead up to our full cell means model.

intercept only

1. $E(y_{ijk}) = \mu$ $+ \alpha_i + \beta_j + \gamma_{ij}$
   $$= 0$$

2. $E(y_{ijk}) = \mu + \alpha_i$

3. $E(y_{ijk}) = \mu + \alpha_i + \beta_j$

4. $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \iff E(y_{ijk}) = \mu_{ij}$

we could also start out by modeling drug first and then add diet

Matrices with Nested Column Spaces

$$
\boldsymbol{X}_1 =
\overset{\mu}{
\begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1
\end{bmatrix}}, \;
\boldsymbol{X}_2 =
\overset{\mu \;\; \alpha_1 \;\; \alpha_2}{
\begin{bmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1
\end{bmatrix}}, \;
\boldsymbol{X}_3 =
\overset{\mu \;\; \alpha_1 \;\; \alpha_2}{
\begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}} \; \&
$$

$$\beta_1 \;\; \beta_2 \;\; \beta_3$$

less than full rank

$$\boldsymbol{X}_4 = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{or}$$

γ_{ij}'s

Matrices with Nested Column Spaces

cell - means model matrix

$$\boldsymbol{X}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

→ full rank

ANOVA Table for Our Two-Factor Example

by how much does SSE decrease when adding diet into model?

end lecture 15 2-24-25

Source	Sum of Squares	DF
Diets|1	$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$2 - 1 = 1$
Drugs|1, Diets	$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$4 - 2 = 2$
Diets × Drugs|1, Diets, Drugs	$\boldsymbol{y}^\top(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{y}$	$6 - 4 = 2$
Error	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_4)\boldsymbol{y}$	$12 - 6 = 6$
C. Total	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$12 - 1 = 11$

given intercept & diet, how much more variability can we explain by adding drug?