Experimental Design
Note 5
Balanced Incomplete Block Design (BIBD)

Keunbaik Lee

Sungkyunkwan University

Review: Nuisance Factor I

Nuisance Factor (may be present in experiment)

- has effect on response but its effect is not of interest.
- If unknown $\rightarrow$ Protecting experiment through randomization.
- If known (measurable) but uncontrollable $\rightarrow$ Analysis of Covariance.
- If known and controllable $\rightarrow$ Blocking.

Example: Penicillin Experiment I

In this experiment, four penicillin manufacturing processes (A, B, C, and D) were being investigated. Yield was the response. It was known that an important raw material, corn steep liquor, was quite variable. The experiment and its results were given below:

	blend 1	blend 2	blend 3	blend 4	blend 5
A	89_1	84_4	81_2	87_1	79_3
B	88_3	77_2	87_1	92_3	81_4
C	97_2	92_3	87_4	89_2	80_1
D	94_4	79_1	85_3	84_4	88_2

- Blend is a nuisance factor, treated as a block factor;

Example: Penicillin Experiment II

- (Complete) Blocking: all the treatments are applied within each block, and they are compared within blocks.
- Advantage: Eliminate blend-to-blend (between-block) variation from experimental error variance when comparing treatments.
- Cost: degree of freedom.

Catalyst Experiment I

Four catalysts are being investigated in an experiment. The experimental procedure consists of selecting a batch raw material, loading the pilot plant, applying each catalyst in a separate run and observing the reaction time. The batches of raw material are considered as blocks, however each batch is only large enough to permit three catalysts to be run.

	Block (raw material)				
catalyst	1	2	3	4	$y_{i\cdot}$
1	73	74	-	71	218
2	-	75	67	72	214
3	73	75	68	-	216
4	75	-	72	75	222
$y_{\cdot j}$	221	224	207	218	$870 = y_{\cdot\cdot}$

Balanced Incomplete Block Design (BIBD) I

BIBD Properties

- There are a treatments and b blocks.
- Each block contains k (different) treatments.
- Each treatment appears in r blocks.
- Each pair of treatments appears together in λ blocks.

a, b, k, r, and λ are not independent.

- $N = ar = bk$, where N is the total number of runs;
- $\lambda(a-1) = r(k-1)$;
  - for any fixed treatment i_0
  - two different ways to count the total number of pairs including treatment i_0 in the experiment.
  - $a-1$ possible pairs, each appears in λ blocks, so $\lambda(a-1)$;

Balanced Incomplete Block Design (BIBD) II

  - treatment i_0 appears in r blocks. Within each block, there are $k-1$ pairs including i_0, so $r(k-1)$.
- $b \geq a$ (a brainteaser for math/stat students).
- Nonorthogonal design
- Reasoning for integer λ:
  - Each treatment is assigned to r blocks.
  - Each of those r blocks has $k-1$ remaining positions
  - Those $r(k-1)$ positions must be evenly shared among the remaining $a-1$ treatments.
- Analyses are based on Intra- and Inter-Block Information

Balanced Incomplete Block Design (BIBD) III

Example 1.

	block			$\mathcal{N}$		
treatment	1	2	3			
A	A	-	A	1	0	1
B	B	B	-	1	1	0
C	-	C	C	0	1	1

Incidence Matrix: $\mathcal{N} = (n_{ij})_{a\times b}$ where $n_{ij} = 1$, if treatment i is run in block j; $= 0$ otherwise.
In Example 1, $a = 3$, $b = 3$, $k = 2$, $r = 2$, $\lambda = 1$ where $\lambda =$the number of times each pair of treatments appers in the same block.

Balanced Incomplete Block Design (BIBD) IV

Example 2

	block						$\mathcal{N}$					
Treatment	1	2	3	4	5	6						
A	A	A	A	-	-	-	1	1	1	0	0	0
B	B	-	-	B	B	-	1	0	0	1	1	0
C	-	C	-	C	-	C	0	1	0	1	0	1
D	-	-	D	-	D	D	0	0	1	0	1	1

$a = 4,\ b = 6,\ k = 2,\ r = 3,$ and $\lambda = 1$.

Balanced Incomplete Block Design (BIBD) V

	block									
Treatment	1	2	3	4	5	6	7	8	9	10
A	A	A	A			A	A	A		
B	B	B		B		B			B	B
C	C			C	C		C	C	C	
D			D	D	D	D	D			D
E		E	E		E			E	E	E

$a = 5,\ b = 10,\ k = 3,\ r = 6,$ and $\lambda = 3.$

BIBD: Statistical Model I

- Statistical Model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

for $i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$.

- additive model (without interaction).
- Not all y_{ij} exist because of incompleteness.
- Usual treatment and block restrictions: $\sum_i \tau_i = 0$; $\sum_j \beta_j = 0$.
- Nonorthogonality of treatments and blocks.

Use Type III Sums of Squres and lsmeans

Model Estimates I

- Least squares estimates for parameters

$$\hat{\mu} = \frac{y_{..}}{N}; \ \hat{\tau}_i = \frac{kQ_i}{\lambda a}; \ \hat{\beta}_j = \frac{rQ_j^{'}}{\lambda b}$$

where

$$Q_i = y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j}; \quad Q_j^{'} = y_{.j} - \frac{1}{r}\sum_i n_{ij} y_{i.}$$

$$\begin{aligned} var(Q_i) &= var(y_{i.}) + var\left(\frac{1}{k}\sum_j n_{ij} y_{.j}\right) - 2cov\left(y_{i.}, \frac{1}{k}\sum_j n_{ij} y_{.j}\right) \\ &= r\sigma^2 + \frac{r}{k^2}k\sigma^2 - \frac{2}{k}r\sigma^2 = \frac{(k-1)r}{k}\sigma^2 \end{aligned}$$

- $var(\hat{\tau}_i) = \left(\frac{k}{\lambda a}\right)^2 var(Q_i) = \left(\frac{k}{\lambda a}\right)^2 \frac{(k-1)r}{k}\sigma^2 = \frac{k(a-1)}{\lambda a^2}\sigma^2;$
  $var(\hat{\tau}_i - \hat{\tau}_j) = \frac{2k\sigma^2}{\lambda a};$

Model Estimates I

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Block	SS_{Block}	$b-1$	MS_{Block}	
Treatment(Adj.)	$SS_{Treatment(Adj.)}$	$a-1$	$MS_{Treatment(Adj.)}$	$MS_{Treatment(Adj.)}/MSE$
Error	SSE	$N-a-b+1$	MSE	
Total	SS_T	$N-1$		

where $SS_T = \sum_i \sum_j y_{ij}^2 - y_{\cdot\cdot}^2/N$; $SS_{Block} = \frac{1}{k}\sum_j y_{\cdot j}^2 - y_{\cdot\cdot}^2/N$

- $SS_{Treatment}$ needs adjustment for incompleteness

$$Q_i = y_{i\cdot} - \frac{1}{k}\sum_j n_{ij} y_{\cdot j}$$

where $n_{ij} = 1$ if trt i in block j; $= 0$ otherwise.

- trt i's total minus trt i's block averages.

Model Estimates II

- $\sum_i Q_i = 0$

$$SS_{Treatment(adjusted)} = k\sum_i Q_i^2/\lambda a = \frac{\lambda a}{k}\sum_i \hat{\tau}_i^2$$

- *SSE* by subtraction
- If $F_0 > F_{\alpha,a-1,N-a-b+1}$, then reject H_0.

Mean Tests and Contrasts

- Must compute adjusted means (lsmeans)
- Adjusted mean is $\hat{\mu} + \hat{\tau}_i$
- Standard error of adjusted mean is $\sqrt{MSE\left(\frac{k(a-1)}{\lambda a^2} + \frac{1}{N}\right)}$
- Contrasts based on adjusted treatment totals
  For a contrast: $\sum_i c_i \mu_i$
  Its estimate: $\sum_i c_i \hat{\tau}_i = \frac{k}{\lambda a} \sum_i c_i Q_i$
  Contrast sum of squares:

$$SS_C = \frac{k\left(\sum_{i=1}^{a} c_i Q_i\right)^2}{\lambda a \sum_{i=1}^{a} c_i^2}$$

Pairwise Comparison

- Pairwise comparison $\tau_i - \tau_j$:
  - Bonferroni:

$$CD = t_{\alpha/2m, ar-a-b+1} \sqrt{MSE \frac{2k}{\lambda a}}$$

  - Tukey:

$$CD = \frac{q_\alpha(a, ar-a-b+1)}{\sqrt{2}} \sqrt{MSE \frac{2k}{\lambda a}}$$

In SAS, there are two procedures: **PROC ANOVA** and **PROC GLM**. PROC ANOVA is for balanced data and PROC GLM is for both balanced or unbalanced data.
See BIBD.SAS.