Planejamento de Experimentos

Delineamentos Experimentais em Blocos

Marinho G Andrade Filho.

Delineamento em Blocos Completos

- O Delineamento em Blocos Completo Casualizados (DBCC) é utilizado quando as unidades experimentais não são homogêneas mas, podem ser agrupadas em grupos homogêneos chamados de blocos, contendo cada grupo, normalmente uma repetição de cada tratamento.
- Quando cada bloco contém pelo menos uma repetição de cada tratamento, diz-se que o experimento foi conduzido em um Delineamento em **Blocos Completos Casualizados** (o bloqueamento também é conhecido como controle local).

Delineamento em Blocos Completos

- *Controle Local (Blocos):* Esse princípio é utilizado quando as unidades experimentais apresentam alguma variação conhecida.
- A finalidade do controle local é dispor as unidades experimentais heterogêneas em sub-unidades homogêneas, reduzindo assim, o erro experimental.
- **Bloco completo:** é aquele que contém pelo menos uma repetição de cada tratamento.
- Os tratamentos agrupados em um bloco devem ter as condições experimentais homogêneas, ou seja, é permitido que se tenha variação entre blocos mas não dentro do bloco.

Delineamento em BC Casualizado

- Suponha que temos, em geral, a tratamentos que são comparados e b blocos.
- Considere apenas uma observação por tratamento em cada bloco.
- A ordem em que os tratamentos são executados dentro de cada blocos é determinado aleatoriamente.
- O modelo estatístico para esse delineamento é

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad i = 1, 2, \ldots, a \text{ e } j = 1, 2, \ldots, b,$$

em que μ é a média total, τ_i é o efeito do i-ésimo tratamento, β_j é o efeito do j-ésimo bloco e ϵ_{ij} é o erro aleatório.

Delineamento em BC Casualizado

- Tratamentos e blocos são considerados inicialmente como fatores fixos. Além disso, os efeitos de tratamentos e blocos são definidos como desvios da média total, tal que

$$\sum_{i=1}^{a} \tau_i = 0 \qquad \text{e} \qquad \sum_{j=1}^{b} \beta_j = 0.$$

Delineamento em BC Casualizado

Pode-se organizar os valores observados da seguinte forma:

Tratamento	Bloco 1	Bloco 2		Bloco b	Total	Média
1	y_{11}	y_{12}		y_{1b}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		
a	y_{a1}	y_{a2}		y_{ab}	$y_{a.}$	$\bar{y}_{a.}$
Total	$y_{.1}$	$y_{.2}$	$\cdots$	$y_{.b}$	$y_{..}$	-
Média	$\bar{y}_{.1}$	$\bar{y}_{.2}$		$\bar{y}_{.b}$	-	$\bar{y}_{..}$

Delineamento em BC Casualizado

- Considere:

$$y_{i.} = \sum_{j=1}^{b} y_{ij}, \qquad \bar{y}_{i.} = \frac{y_{i.}}{b}, \quad i = 1, 2, \ldots, a,$$

$$y_{.j} = \sum_{i=1}^{a} y_{ij}, \qquad \bar{y}_{.j} = \frac{y_{.j}}{a}, \quad j = 1, 2, \ldots, b,$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}, \qquad \bar{y}_{..} = \frac{y_{..}}{N},$$

em que $N = ab$ é o número total de observações.

Delineamento em BC Casualizado

- Estamos interessados em testar a igualdade das médias dos tratamentos. Assim, as hipóteses de interesse são:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_a \quad \times \quad H_1 : \mu_i \neq \mu_k, \ \exists \ (i, k), i \neq k.$$

- Uma vez que a média do i-ésimo tratamento é $\mu_i = \frac{1}{b} \sum_{j=1}^{b} (\mu + \tau_i + \beta_j) = \mu + \tau_i$, $i = 1, 2, \ldots, a$,

  uma maneira equivalente de escrever H_0 em termos dos efeitos de tratamentos é:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0 \quad \times \quad H_1 : \tau_i \neq 0, \ \exists \ i.$$

Delineamento em BC Casualizado

- Podemos expressar a soma dos quadrados total como:

$$
\begin{aligned}
SQ_T &= \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\Big[(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})\Big]^2 \\
&= b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2,
\end{aligned}
$$

Expressando simbolicamente, temos:

$$SQ_T = SQ_{Trat} + SQ_{Bloco} + SQ_{Res}.$$

Delineamento em BC Casualizado

- As somas de quadrados pode ser calculadas por:

$$SQ_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - C,$$

$$SQ_{Trat} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - C$$

$$SQ_{Bloco} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - C$$

em que $C = \dfrac{y_{..}^2}{N}$. A SQ_{Res} é obtida pela subtração

$$SQ_{Res} = SQ_T - SQ_{Trat} - SQ_{Bloco}.$$

Delineamento em BC Casualizado

- Existem $N = ab$ observações; então, SQ_T tem $N - 1 = ab - 1$ *g.l*
- Existem a tratamentos e b blocos, então SQ_{Trat} tem $a - 1$ *g.l.* e SQ_{Bloco} tem $b - 1$ *g.l.*
- Então SQ_{Res} tem $ab - 1 - (a-1) - (b-1) = (a-1)(b-1)$ *g.l.*
- Os quadrados médios são calculados por:

$$QM_{Trat} = \frac{SQ_{Trat}}{a-1},$$

$$QM_{Bloco} = \frac{SQ_{Trat}}{b-1}$$

$$QM_{Res} = \frac{SQ_{Res}}{(a-1)(b-1)}$$

Delineamento em BC Casualizado

- Análise da variância para um delineamento em BCC.

Fonte de Variação	Soma de Quadrados	Graus de Liberdades	Quadrado Médio	F_0
Tratamentos	SQ_{Trat}	$a-1$	$QM_{Trat} = \frac{SQ_{Trat}}{a-1}$	$F_0 = \frac{QM_{Trat}}{QM_{Res}}$
Blocos	SQ_{Bloco}	$b-1$	$QM_{Bloco} = \frac{SQ_{Bloco}}{b-1}$	$F_b = \frac{QM_{bloco}}{QM_{Res}}$
Resíduo	SQ_{Res}	$(a-1)(b-1)$	$QM_{Res} = \frac{SQ_{Res}}{(a-1)(b-1)}$	
Total	SQ_T	$N-1$		-

- Para testar a igualdade das médias dos tratamentos, rejeita-ses $H_0 : \tau_i = 0$ se $F_0 > F_{\alpha,a-1;(a-1)(b-1)}$.
- Para testar o efeito dos blocos, rejeita-se $H_0 : \beta_j = 0$ se $F_b > F_{\alpha,b-1;(a-1)(b-1)}$.

Delineamento em BC Casualizado

- Apesar de ser bastante questionável (devido a aleatorização ter sido aplicada apenas nos tratamentos dentro dos blocos), este procedimento aproximado para investigar o efeito da variável bloco é razoável.
- O teste de Tukey para comparar as médias de dois tratamentos em um delineamento em blocos completos casualizado:

$$T_\alpha = q_\alpha(a; f)\sqrt{\frac{QM_{Res}}{b}},$$

em que b é o número de blocos e $q_\alpha(a; f)$ é tabelado em função do número de tratamento (a) e do número de *g.l.* do resíduo ($f = (a-1)(b-1)$).

Exemplo de Delineamento em BCC

Dados da produção de milho (em $Kg/100m^2$) segundo a variedade e o bloco.

Blocos	Variedades				Total
	A	B	C	D	
I	32	24	35	21	112
II	24	35	43	26	128
III	33	42	39	30	144
IV	38	36	43	39	156
V	33	38	55	34	160
Total	160	175	215	150	700

Deseja-se saber, ao nível de 5%, se as variedades estudadas tem em média a mesma produção nos diferentes blocos.

Exemplo de Delineamento em BCC

- Temos: $a = 4$ (tratamentos) e $b = 5$ (blocos)

$$C = \frac{y_{..}^2}{N} = \frac{(700)^2}{20} = 24500$$

$$SQ_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - C = (32^2 + 24^2 + \ldots + 34^2) - 24500 = 1190$$

$$SQ_{Trat} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - C = \frac{160^2 + 175^2 + 215^2 + 150^2}{5} - 24500 = 490$$

$$SQ_{Bloco} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - C = \frac{112^2 + 128^2 + 144^2 + 156^2 + 160^2}{4} - 24500 = 400$$

$$SQ_{Res} = SQ_T - SQ_{Trat} - SQ_{Bloco} = 1190 - 490 - 400 = 300$$

Exemplo de Delineamento em BCC

Tabela da análise da variância.

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	F_0
Entre Tratamentos	490	3	163,33	6,53
Blocos	400	4	100,00	
Resíduos	300	12	25,00	
Total	1190	19		-

Sendo $F_0 = 6,53 > F_{0.05;3;12} = 3,49$, Rejeita-se H_0 com $\alpha = 5\%$, ou seja, as variedades estudadas **não** tem em média a mesma produção.

O que podemos dizer sobre o efeito dos blocos? ($F_{0.05;4;12} = 3,26$),

Teste de Tukey para BCC

A estatística de teste é:

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{QM_{Res}}{b}} = q_{0,05}(3; 12)\sqrt{\frac{25,00}{5}} = 3,77 \times 2,24 = 8,44$$

Considere as diferenças entre as médias:

Comparação	Tratamentos
$\bar{y}_{1.} - \bar{y}_{2.}$	$\|32 - 35\| = 3$
$\bar{y}_{1.} - \bar{y}_{3.}$	$\|32 - 43\| = 11(*)$
$\bar{y}_{1.} - \bar{y}_{4.}$	$\|32 - 30\| = 2$
$\bar{y}_{2.} - \bar{y}_{3.}$	$\|35 - 43\| = 8$
$\bar{y}_{2.} - \bar{y}_{4.}$	$\|35 - 30\| = 5$
$\bar{y}_{3.} - \bar{y}_{4.}$	$\|43 - 30\| = 13(*)$

(*) indica as diferenças significativas.

BCC: Estimação dos parâmetros do modelo

- O modelo estatístico linear com efeito fixo para os tratamentos e blocos, é:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad i = 1, 2, \ldots, a \text{ e } \quad j = 1, 2, \ldots, b.$$

- Utilizando o o método de mínimos quadrados, temos:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \mu - \tau_i - \beta_j)^2.$$

- Podemos encontrar uma estimativa para μ, τ_i e β_j ($i = 1, 2, \ldots, a$ e $j = 1, 2, \ldots, b$), calculando as derivadas parciais de $\sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2$ em relação a cada um dos parâmetros (μ, τ_i e β_j) e igualando a zero.

BCC: Estimação dos parâmetros do modelo

- Dessa forma, temos:

$$\frac{\partial \sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2}{\partial \mu} = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \mu - \tau_i - \beta_j) = 0$$

$$\frac{\partial \sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2}{\partial \tau_i} = \sum_{j=1}^{b} (y_{ij} - \mu - \tau_i - \beta_j) = 0, \quad i = 1, \ldots, a$$

$$\frac{\partial \sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2}{\partial \beta_j} = \sum_{i=1}^{a} (y_{ij} - \mu - \tau_i - \beta_j) = 0 \quad j = 1 \ldots, b.$$

BCC: Estimação dos parâmetros do modelo

- Aplicando os somatórios, chega-se ao seguinte sistema de equações normais:

$$N\mu + b\sum_{i=1}^{a}\tau_i + a\sum_{j=1}^{b}\beta_j = y_{..}$$

$$b\mu + b\tau_i + \sum_{j=1}^{b}\beta_j = y_{i.}, \quad i = 1, \ldots, a$$

$$a\mu + \sum_{i=1}^{a}\tau_i + a\beta_j = y_{.j}, \quad j = 1, \ldots, b$$

- Este sistema possui $a+b+1$ equações com $a+b+1$ incógnitas $(\mu, \tau_1, \tau_2, \ldots, \tau_a, \beta_1, \beta_2, \ldots, \beta_b)$.

BCC: Estimação dos parâmetros do modelo

- Dada as restrições

$$\sum_{i=1}^{a} \tau_i = 0 \qquad \text{e} \qquad \sum_{j=1}^{b} \beta_j = 0,$$

A solução do sistema de equações normais fornece os estimadores de MQ dos parâmetros do modelo:

$$\hat{\mu} = \bar{y}_{..},$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \; i = 1, 2, \ldots, a$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}, \; j = 1, 2, \ldots, b,$$

BCC: Estimação dos parâmetros do modelo

- Os valores dos y_{ij} podem ser estimados por:

$$\begin{aligned}\hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}.\end{aligned}$$

- O erro $\epsilon_{ij} = y_{ij} - \mu - \tau_i - \beta_j$. pode ser estimados por:

$$\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij}.$$

$$\hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}.$$

Delineamento em BCC com n Repetições

Blocos	Tratamento 1	2	$\ldots$	a	Total	Média
1	y_{111}	y_{211}	$\cdots$	y_{a11}	$y_{.1.}$	$\bar{y}_{.1.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	y_{11n}	y_{21n}	$\cdots$	y_{a1n}		
2	y_{121}	y_{221}	$\cdots$	y_{a21}	$y_{.2.}$	$\bar{y}_{.2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	y_{12n}	y_{22n}	$\cdots$	y_{a2n}		
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
b	y_{1b1}	y_{2b1}	$\cdots$	y_{ab1}	$y_{.b.}$	$\bar{y}_{.b.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	y_{1bn}	y_{2bn}	$\cdots$	y_{abn}		
Total	$y_{1..}$	$y_{2..}$	$\cdots$	$y_{a..}$	$y_{...}$	-
Média	$\overline{y}_{1..}$	$\overline{y}_{2..}$	$\cdots$	$\overline{y}_{a..}$	-	$\bar{y}_{...}$

Delineamento em BCC com n Repetições

- Desta forma, considere:

  $y_{i..}$ a soma das observações do i-ésimo tratamento,
  $y_{.j.}$ a soma das observações do j-ésimo bloco,
  $y_{...}$ a soma de todas as observações,
  $\bar{y}_{i..}$ a média sobre as observações do i-ésimo tratamento,
  $\bar{y}_{.j.}$ a média sobre as observações do i-ésimo tratamento e
  $\bar{y}_{...}$ a média de todas as observações.

Delineamento em BCC com n Repetições

- Expressando-os simbolicamente temos:

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}, \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn}, \; i = 1, 2, \ldots, a,$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}, \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{an}, \; j = 1, 2, \ldots, b,$$

$$y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}, \qquad \bar{y}_{...} = \frac{y_{...}}{N},$$

em que $N = abn$ é o número total de observações.

Delineamento em BCC com n Repetições

- A soma dos quadrados total é dada por:

$$
\begin{aligned}
SQ_T &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})\right]^2 \\
&= bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2,
\end{aligned}
$$

- a SQ_T pode ser decomposta como::

$$SQ_T = SQ_{Trat} + SQ_{Bloco} + SQ_{Res}.$$

Delineamento em BCC com n Repetições

- Existem $N = abn$ total de observações; assim, as somas de quadrados pode ser obtidas por:

$$SQ_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - C, \quad \text{tem} \quad N-1 \quad g.l.$$

$$SQ_{Trat} = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - C \quad \text{tem} \quad a-1 \quad g.l.$$

$$SQ_{Bloco} = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - C \quad \text{tem} \quad b-1 \quad g.l.$$

- em que $C = \dfrac{y_{...}^2}{N}$.

Delineamento em BCC com n Repetições

- A soma dos quadrados dos erro pode ser calculada por diferença.

$$SQ_{Res} = SQ_T - SQ_{Trat} - SQ_{Bloco}.$$

- SQ_{Res} tem $abn - 1 - (a-1) - (b-1) = N - a - b + 1$ *g.l.*.
- O quadrado médio dos tratamentos, dos blocos e dos resíduos são dados por:

$$QM_{Trat} = \frac{SQ_{Trat}}{a-1},$$

$$QM_{Bloco} = \frac{SQ_{Trat}}{b-1}$$

$$QM_{Res} = \frac{SQ_{Res}}{N-a-b+1}$$

Delineamento em BCC com n Repetições

- A tabela da ANOVA para o BCC com n Repetições:

Fonte de Variação	Soma de Quadrados	Graus de Liberdades	Quadrado Médio	F_0
Tratamentos	SQ_{Trat}	$a-1$	$QM_{Trat} = \frac{SQ_{Trat}}{a-1}$	
Blocos	SQ_{Bloco}	$b-1$	$QM_{Bloco} = \frac{SQ_{Bloco}}{b-1}$	$F_0 = \frac{QM_{Trat}}{QM_{Res}}$
Resíduo	SQ_{Res}	$N-a-b+1$	$QM_{Res} = \frac{SQ_{Res}}{N-a-b+1}$	
Total	SQ_T	$N-1$		-

Para testar $H_0 : \mu_1 = \mu_2 = \ldots = \mu_a$, ou equivalentemente, $H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$, utilizamos a estatística de teste

$$F_0 = \frac{QM_{Trat}}{QM_{Res}},$$

Rejeita-se H_0 se $F_0 > F_{\alpha;a-1;N-a-b+1}$.

Exemplo: delineamento em BCC com repetições

- Considere a tabela com as notas dos alunos em um teste segundo a fonte de informação (tratamento) e a faixa de idade (bloco):

Bloco	Tratamento				Total
	A	B	C	D	
	65	56	58	38	
I	69	49	65	30	648
	73	54	57	34	
	72	73	76	71	
II	79	77	69	65	864
	80	69	71	62	
Total	438	378	396	300	1512

- Deseja-se saber, ao nível de 5%, se a nota média dos alunos são iguais para todas as fontes de informação, independente da idade.

Exemplo: delineamento em BCC com repetições

$a = 4$ (níveis do fator), $b = 2$ (blocos), $n = 3$ (réplicas)

$$C = \frac{y_{\ldots}^2}{N} = \frac{(1512)^2}{24} = 95256$$

$$SQ_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - C = (65^2 + 69^2 + 73^2 + \ldots + 65^2 + 62^2) - 95256 = 4382$$

$$SQ_{Trat} = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - C = \frac{438^2 + 378^2 + 396^2 + 300^2}{6} - 95256 = 1668$$

$$SQ_{Bloco} = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - C = \frac{648^2 + 864^2}{12} - 95256 = 1944$$

$$SQ_{Res} = SQ_T - SQ_{Trat} - SQ_{Bloco} = 4382 - 1668 - 1944 = 770$$

Exemplo: delineamento em BCC com repetiçõe

Tabela da ANOVA

Fonte de Variação	Soma de Quadrados	Graus de Liberdades	Quadrado Médio	F_0
Tratamentos	1668	3	556	
Blocos	1944	1	1944	$F_0 = 13,72$
Resíduo	770	19	$40,526$	
Total	4382	23		-

- Sendo $F_0 = 13,72 > F_{0.05;3;19} = 3,13$, há indícios para rejeitar H_0, ou seja, a nota média dos alunos não é a mesma para todas as fontes de informação
- A nota média dos alunos depende da idade? $(F_{0.05;1;19} = 4,38)$