Universidade de Brasília

Delineamento e Análise de Experimentos

Professora Juliana Betini Fachini Gomes
e-mail: jfachini@unb.br

Brasília - 2023

Modelo

O modelo do delineamento em blocos casualizados completo para a tratamentos e b blocos é definido por:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., b \quad (1)$$

em que μ é a média geral, τ_i é a média ou efeito dos tratamentos, β_j é o efeito de blocos e ε_{ij} componente de erro aleatório com distribuição $N(0, \sigma^2)$.

Modelo

- Usualmente, pensamos nos efeitos de tratamento e bloco como desvios da média geral, de modo que

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

- Também podemos escrever o modelo (31) como modelo de médias:

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., b$$

em que $\mu_{ij} = \mu + \tau_i + \beta_j$.

Modelo

- As hipóteses de interesse para o modelo (31) são definidas por:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0, \quad \text{(O efeito de tratamento é nulo)} \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

- Ou de forma equivalente:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = ... = \mu_a \\ H_1 : \exists \mu_i \neq \mu_j,\ i \neq j, \end{cases}$$

em que $\mu_i = (1/b)\sum_{j=1}^{b}(\mu + \tau_i + \beta_j) = \mu + \tau_i$.

Modelo

A análise de variância pode ser estendida para o modelo (31). Seja:

- $y_{i.} = \sum_{j=1}^{b} y_{ij}$ o total de todas as observações do i tratamento;
- $y_{.j} = \sum_{i=1}^{a} y_{ij}$ o total de todas as observações no j bloco;
- $y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}$ o total geral de todas as observações e $N = ab$ o número total de observações.

Modelo

Similarmente, tem-se as médias:

- $\bar{y}_{i.} = y_{i.}/b$ é a média das observações do i tratamento;
- $\bar{y}_{.j} = y_{.j}/a$ é a média das observações no j bloco;
- $\bar{y}_{..} = y_{..}/N$ é a média geral de todas as observações.

Modelo

- A SS_T pode ser escrita como:

$$
\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 &= b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j}-\bar{y}_{i.}+\bar{y}_{..})^2 \quad (2)\\
&= SQ_{Trat} + SQ_{Bloco} + SQ_{Res}
\end{aligned}
$$

ANOVA

- Relembrando o Teorema de Cochran temos:
  - Seja $Z_i \sim N(0,1)$ para $i = 1, 2, \ldots, \nu$ e

$$\sum_{i=1}^{\nu} Z_i^2 = Q_1 + Q_2 + \ldots + Q_s$$

em que $s \leq \nu$, e Q_i tem ν graus de liberdade $(i = 1, 2, \ldots, s)$. Então, $Q_1, Q_2, \ldots, Q_s$ são variáveis aleatórias independentes com distribuição qui-quadrado com $\nu_1, \nu_2, \ldots, \nu_s$ graus de liberdade, respectivamente, se e somente se

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_s$$

ANOVA

- Ao analisar a equação (2) é possível verificar que a soma dos graus de liberdade do lado direito da equação é igual ao grau de liberdade da SQ_{Total};
- E ao fazer as suposições de normalidade dos erros, tem-se que

$$\frac{SQ_{Trat}}{\sigma^2}, \quad \frac{SQ_{Bloco}}{\sigma^2} \quad \text{e} \quad \frac{SQ_{Res}}{\sigma^2}$$

  são variáveis aleatórias independentes com distribuição qui-quadrado;
- Cada soma de quadrados dividida por seus graus de liberdade é um quadrado médio.

ANOVA

- Então, a esperança dos quadrados médios, se tratamentos e blocos forem fixos, podem ser definidos por:

$$E(QM_{Trat}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(QM_{Bloco}) = \sigma^2 + \frac{a\sum_{j=1}^{a}\beta_j^2}{b-1}$$

$$E(QM_{Res}) = \sigma^2$$

ANOVA

- Sendo assim, para testar a igualdade das médias de tratamento, a estatística de teste é definida por:

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

  com distribuição $F_{a-1,(a-1)(b-1)}$ se a hipótese nula for verdadeira.

- A região crítica é a cauda superior da distribuição F, e rejeitamos H_0 se $F_0 > F_{\alpha,a-1,(a-1)(b-1)}$.

ANOVA

- É possível testar efeito de bloco ($H_0 : \beta_j = 0$)?
- Qual a estatística do teste?
- É possível seguir a mesma ideia da estatística de teste para efeitro de tratamento. Então, tem-se que:

$$F_0 = \frac{QM_{Bloco}}{QM_{Res}}$$

- No entanto, lembre-se de que a aleatorização foi aplicada apenas para tratamentos em cada blocos; ou seja, os blocos representam uma restrição à aleatorização;
- Que efeito isso tem na estatística do teste?

ANOVA

Tabela 1: Tabela de Análise de Variância

Fonte de Variação	SQ	g.l.	QM	F
Tratamentos	SQ_{Trat}	a - 1	QM_{Trat}	QM_{Trat}/QM_{Res}
Bloco	SQ_{Bloco}	b - 1	QM_{Bloco}	
Resíduo	SQ_{Res}	(a-1)(b-1)	QM_{Res}	
Total	SQ_T	N - 1		

Comparações de médias

- Caso seja verificado pela ANOVA a diferença entre as médias de tratamentos;
- Pode-se utilizar todas as técnicas de comparaçoes múltiplas aprendidas em Delineamento Inteiramente Casualizados;
- Porém, deve-se substituir o número de repetições (n) por número de blocos (b);
- E o número de graus de liberdade do resíduo de $(a(n-1))$ por $((a-1)(b-1)$.

Qual o tamanho da amostra ou o número de blocos?

- Em qualquer problema de planejamento experimental, uma decisão crítica é a escolha do tamanho da amostra, isto é, determinar o número de blocos a serem executadas, ao considerar o delineamento em blocos casualizados;

- Aumentar o número de blocos aumenta o número de repetições e o número de graus de liberdade do erro, tornando o experimento mais sensível.

Tamanho da amostra

- A probabilidade do erro tipo II do modelo de efeitos fixos é definida por:

$$\begin{aligned} \beta &= P\{\text{não rejeitar} H_0 | H_0 \text{é falsa}\} \\ &= P\{F_0 < F_{\text{crítico}} | H_0 \text{é falsa}\} \end{aligned} \tag{3}$$

Tamanho da amostra

- Para avaliar a probabilidade do erro tipo II, definida na equação (3), é preciso conhecer a distribuição da estatística do teste F_0 se a hipótese nula for falsa;

- Pode-se mostrar que, se H_0 for falsa, a estatística $F_0 = QM_{Trat}/QM_{Res}$ tem distribuição F não central com $a-1$ e $(a-1)(b-1)$ graus de liberdade e parâmetro de não centralidade δ;

- Se $\delta = 0$, a distribuição F não central torna-se a distribuição F usual (central).

Tamanho da amostra

- O parâmetro de não centralidade da distribuição F pode ser obtido ao calcular:

$$E\left[\frac{SQ_{Trat}}{\sigma^2}\right],$$

- E esse parâmetro será igual a:

$$\delta = \frac{b\sum_{i=1}^{a}\tau_i^2}{\sigma^2}, \tag{4}$$

- É possível observar que sob H_0, a equação (4) é igual a 0.

Tamanho da amostra

- O pesquisador deve especificar os valores de τ e σ^2;
- A estimativa de σ^2 pode estar disponível a partir de uma experiência anterior, um experimento anterior ou uma estimativa de julgamento.
- Ao calcular δ, β e o poder do teste $(1-\beta)$ para diferentes valores de n, é possível encontrar o tamanho da amostra.

Estimação dos parâmetros

- Para encontrar os estimadores de mínimos quadrados de μ, τ_i e β_j, é necessário escrever a soma dos quadrados dos erros:

$$L = \sum_{i=1}^{a} \sum_{j=1}^{b} \varepsilon_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \mu - \tau_i - \beta_j)^2 \tag{5}$$

e os valores de μ, τ_i e β_j que minimizam a equação (5) são os estimadores de mínimos quadrados, $\hat{\mu}$, $\hat{\tau}_i$ e $\hat{\beta}_j$.

- Os valores apropriados seriam as soluções para as equações simultâneas:

$$\frac{\partial L}{\partial \mu} | \hat{\mu}, \hat{\tau}_i, \hat{\beta}_j = 0,$$

$$\frac{\partial L}{\partial \tau_i} | \hat{\mu}, \hat{\tau}_i, \hat{\beta}_j = 0 \quad i = 1, 2, ..., a.$$

e

$$\frac{\partial L}{\partial \beta_j} | \hat{\mu}, \hat{\tau}_i, \hat{\beta}_j = 0 \quad j = 1, 2, ..., b.$$

Estimação dos parâmetros

- Ao aplicar as restrições: $\sum_{i=1}^{a} \hat{\tau}_i = 0$ e $\sum_{j=1}^{b} \hat{\beta}_j = 0$, mostre que a solução para o sistema de equações normais é:

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

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

e

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

Diagnóstico

- É fundamental verificar a adequabilidade do modelo antes de tirar conclusões sobre o processo inferencial.
- Verificar as características do modelo.
- Para avaliar a adequabilidade do modelo serão usados:
  - Métodos gráficos
  - Testes estatísticos

Resíduo

- Violações das suposições básicas e adequação do modelo podem ser facilmente investigadas através da análise dos resíduos.
- O resíduo no i-ésimo tratamento do j-ésimo bloco é definido por:

$$e_{ij} = y_{ij} - \hat{y}_{ij}, \tag{6}$$

em que $\hat{y}_{ij}$ é o valor estimado obtido por:

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..} \end{aligned} \tag{7}$$

Resíduo

- Para o modelo (31), assume-se que o erros, ε_{ij} são variáveis aleatórias independentes, identicamente distribuídos com distribuição Normal com média 0 e variância σ^2.

- Se o modelo (31) for apropriado para os dados em estudo, os resíduos observados, e_{ij}, devem refletir as propriedades assumidas para o erro do modelo, ε_{ij}.

Suposição de Normalidade

- Pode ser inicialmente verificada ao construir um histograma dos resíduos. Se a suposição sobre os erros for satisfeita, este gráfico deve se parecer com uma amostra de uma distribuição normal centrada em zero;
- Outro recurso gráfico importante é o **Gráfico de Probabilidade Normal dos resíduos** - cada resíduo é colocado em um gráfico versus seu valor esperado sob normalidade.
- Se o resultado gráfico for aproximadamente linear, há evidências da suposição de normalidade.
- Se o resultado gráfico afasta substancialmente da linearidade, não há evidências de que a distribuição do erro é normal.

Suposição de Normalidade

- As vezes é útil padronizar os resíduos para a análise de resíduos.
- Como o desvio padrão dos erros, ε_{ij}, é σ, que pode ser estimado por $\hat{\sigma} = \sqrt{\text{MSRes}}$. É natural considerar a seguinte padronização:

$$d_{ij} = \frac{e_{ij}}{\sqrt{\text{MSRes}}}. \tag{8}$$

- Se $\varepsilon_{ij} \sim N(0, \sigma^2)$, os resíduos padronizados (8) deve ser aproximadamente normal com média zero e variância constante;
- Teste de hipótese também pode ser utilizado, como o teste de Shapiro-Wilk.

Suposição de independência

- O objetivo é verificar se existe qualquer correlação entre os erros que são próximos um dos outros em uma sequência;
- Quando os erros são independentes espera-se que os resíduos versus a ordem das observações flutuem em um padrão mais ou menos aleatório em torno da linha de referência Zero.

Variância constante

- Gráficos dos resíduos versus valores ajustados podem ser usados para verificar se a variância dos erros é constante. Pois, os sinais dos resíduos não são importantes para examinar a constância da variância do erro;
- Testes de hipóteses.

Variância constante

Para verificar as hipóteses:

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 = ... = \sigma_a = 0 \\ H_1 : \exists \sigma_i \neq 0 \end{cases}$$

pode-se utilizar os seguintes testes de hipóteses:

- F;
- Bartlett, e
- Levene

Aditividade do modelo em blocos casualizados

Outro aspecto do delineamento em blocos casualizados é supor que o modelo linear

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., b$$

é completamente aditivo.

Para verificar esse pressuposto deve-se:

1. Ajustar um modelo de regressão com as variáveis explicativas: Tratamento e Bloco;

2. Ajustar um modelo de regressão com as variáveis explicativas: Tratamento e Bloco, e o efeito de Interação entre Tratamento e Bloco;

3. Comparar os dois modelos e verificar se o parâmetro de Interação não é siginificativo $\Rightarrow$ o pressuposto de aditividade é atendido.