Delineamento e Análise de Experimentos

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

Brasília - 2023

Exemplo

- Suponha que um pesquisador esteja estudando os efeitos de cinco formulações diferentes de um propulsor de foguete usado em sistemas de escape de tripulação na taxa de queima observada;
- Cada formulação é misturada a partir de um lote de matéria-prima que é grande o suficiente apenas para cinco formulações serem testadas;
- As formulações são preparadas por vários operadores

Exemplo

- Logo, parece que há dois fatores de "incômodo" a serem considerado no projeto: lotes de matéria-prima e operadores;
- O projeto apropriado para este problema consiste em testar a exatidão de cada formulação uma vez em cada lote de matéria-prima e que cada formulação seja preparada exatamente uma vez por cada um dos cinco operadores;
- O experimento adequado para esse problema é o Experimento em Quadrado Latino.

Os experimentos em Quadrado Latino consideram os seguintes princípios básicos da experimentação:

- Casualização;
- Repetição;
- Controle local em dois sentindos perpendiculares: chamados de linhas e colunas

Exemplos

- Experimentos envolvendo animais de pastejo em que pretende-se estudar o efeito de rações na produção de leite sobe regime de pastejo, com várias forrageiras:
  - De um lado controlam-se as várias forrageiras e de outro, os diferentes animais.
- Desgaste de pneus - deve-se controlar os tipos de carros e a posição em que o pneu se encontra.

Quadrado Latino

- A principal característica nesse tipo de experimento é que o número de linhas é igual ao número de colunas e igual ao número de tratamentos;

- Então, cada tratamento aparece apenas uma vez em cada linha e uma vez em cada coluna;

- Se tivermos p tratamentos, teremos p^2 parcelas.

Quadrado Latino

Tabela : Quadrado Latino antes da casualização

	Colunas				
Linhas	1	2	3	4	5
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

Quadrado Latino

- A casualização pode ser feita em uma ou duas etapas;
- Em uma etapa - faz-se o sorteio das linhas ou das colunas;
- Em duas etapas - primeiro faz-se o sorteio das linhas e em seguida das colunas. Ou vice e versa.

Quadrado Latino

Tabela : Quadrado Latino após casualização das colunas

	Colunas				
Linhas	3	5	1	2	4
1	C	E	A	B	D
2	D	A	B	C	E
3	E	B	C	D	A
4	A	C	D	E	B
5	B	D	E	A	C

Quadrado Latino

Tabela : Quadrado Latino após casualização das colunas e linhas

	Colunas				
Linhas	3	5	1	2	4
3	E	B	C	D	A
1	C	E	A	B	D
4	A	C	D	E	B
5	B	D	E	A	C
2	D	A	B	C	E

Modelo

O modelo do experimento em Quadrado Latino é definido por:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk}, \tag{1}$$

em que $i = 1, 2, ..., p$; $j = 1, 2, ..., p$; $k = 1, 2, ..., p$; y_{ijk} é a observação da i-ésima linha e k-ésima coluna para o j-ésimo tratamento; μ é a média geral; α_i é o efeito de linha; τ_j é o efeito dos tratamentos; β_k é o efeito de coluna e ε_{ijk} componente de erro aleatório com distribuição $N(0, \sigma^2)$.

Modelo

- As hipóteses de interesse para o modelo (1) são definidas por:

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

Modelo

A análise de variância pode ser estendida para o modelo (1) ao particionar a soma quadrados total de $N = p^2$ observações em componentes de linhas, colunas, tratamentos e erros, por exemplo

$$SQ_T = SQ_{Linha} + SQ_{Trat} + SQ_{Coluna} + SQ_{Res}; \qquad (2)$$

com respectivo grau de liberdade:

$$p^2 - 1 = (p-1) + (p-1) + (p-1) + (p-2)(p-1)$$

Modelo

- As somas de quadrados também podem ser escritas como:

$$SQ_T = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{\ldots}^2}{N}$$

$$SQ_{Trat} = \frac{1}{p} \sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{\ldots}^2}{N}$$

$$SQ_{Linha} = \frac{1}{p} \sum_{i=1}^{p} y_{i..}^2 - \frac{y_{\ldots}^2}{N}$$

$$SQ_{Coluna} = \frac{1}{p} \sum_{k=1}^{p} y_{..k}^2 - \frac{y_{\ldots}^2}{N}$$

$$SQ_{Res} = SQ_T - (SQ_{Linha} + SQ_{Trat} + SQ_{Coluna})$$

ANOVA

- Sob a suposição de $\varepsilon_{ijk} \sim N(0, \sigma^2)$, cada soma de quadrados do lado direito da equação (1) dividido por σ^2, são variáveis aleatórias independente com distribuição χ^2.
- 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 $F_{p-1,(p-2)(p-1)}$ se a hipótese nula for verdadeira.

- A região crítica é a cauda superior da distribuição F, e rejeitamos H_0 se $F_0 > F_{\alpha,p-1,(p-2)(p-1)}$.

ANOVA

Tabela 1: Tabela de Análise de Variância - Quadrado Latino

Fonte de Variação	SQ	g.l.	QM	E[QM]	F
Tratamentos	SQ_{Trat}	p - 1	QM_{Trat}	$\sigma^2 + \frac{p\sum \tau_j^2}{p-1}$	QM_{Trat}/QM_{Res}
Linhas	SQ_{Linha}	p - 1	QM_{Linha}	$\sigma^2 + \frac{p\sum \alpha_i^2}{p-1}$	
Colunas	SQ_{Coluna}	p - 1	QM_{Coluna}	$\sigma^2 + \frac{p\sum \beta_k^2}{p-1}$	
Resíduo	SQ_{Res}	(p-2)(p-1)	QM_{Res}	σ^2	
Total	SQ_T	$p^2 - 1$			

ANOVA

- Como o número de tratamentos define o número de linhas e de colunas ele também define o número de repetições;
- A grande restrição dos experimentos em quadrado çatino é que para 2, 3 ou 4 tratamentos, teremos apenas 0, 2 ou 6 grau de liberdade para o resíduo, respectivamente;
- Por outro lado, com 9 ou mais tratamentos, o quadrado latino fica muito grande trazendo dificuldades na instalação;
- Então, os quadrados latinos mais utilizados são de $5 \times 5, 6 \times 6, 7 \times 7$ e 8×8.

Comparações de médias

- Caso seja verificado pela ANOVA a diferença entre as médias de tratamentos;
- Pode-se utilizar todas as técnicas de comparaçoes múltiplas estudadas anteriormente;
- Porém, deve-se realizar as modificações necessárias em relação ao número de repetições e número de graus de liberdade.

Estimação dos parâmetros

- Para encontrar os estimadores de mínimos quadrados de μ, α_i, τ_j e β_k, é necessário escrever a soma dos quadrados dos erros:

$$L = \sum_i \sum_{jk} \varepsilon_{ijk}^2 = \sum_i \sum_{jk} (y_{ijk} - \mu - \alpha_i - \tau_j - \beta_k)^2 \qquad (3)$$

e os valores de μ, α_i, τ_j e β_k, que minimizam a equação (3) são os estimadores de mínimos quadrados, $\hat{\mu}$,$\hat{\alpha}_i$ $\hat{\tau}_j$ e $\hat{\beta}_k$.

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

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

$$\frac{\partial L}{\partial \alpha_i} | \hat{\mu}, \hat{\alpha}_i, \hat{\tau}_j, \hat{\beta}_k = 0 \quad i = 1, 2, ..., p,$$

$$\frac{\partial L}{\partial \tau_j} | \hat{\mu}, \hat{\alpha}_i, \hat{\tau}_j, \hat{\beta}_k = 0 \quad i = 1, 2, ..., p,$$

e

$$\frac{\partial L}{\partial \beta_k} | \hat{\mu}, \hat{\alpha}_i, \hat{\tau}_j, \hat{\beta}_k = 0 \quad j = 1, 2, ..., p.$$

Estimação dos parâmetros

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

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

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

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

e

$$\hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{...}$$

Diagnóstico do Modelo

- As técnicas de diagnóstico utilizadas para os experimentos anteriores também devem ser utilizadas para verificar os pressupostos do modelo;
- Para o modelo (1), os resíduos são definidos por:

$$\begin{aligned} e_{ijk} &= y_{ijk} - \hat{y}_{ijk} \qquad (4) \\ &= y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...} \end{aligned}$$

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.
- 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: F, Bartlett e Levene podem ser utilizados com certa parcimônia.

Aditividade

- Teste de hipótese: Teste de aditividade de Tukey

Qual o tamanho da amostra?

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

Tamanho da amostra

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

$$\begin{aligned} \beta &= P\{\text{não rejeitar} H_0 | H_0 \text{é falsa}\} \qquad (5) \\ &= P\{F_0 < F_{\text{crítico}} | H_0 \text{é falsa}\} \end{aligned}$$

Tamanho da amostra

- Para avaliar a probabilidade do erro tipo II, definida na equação (5), 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 $p-1$ e $(p-2)(p-1)$ 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:

$$\delta = \frac{p\sum_{j=1}^{p}\tau_j^2}{\sigma^2}, \tag{6}$$

- É possível observar que sob H_0, a equação (6) 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.