Delineamento e Análise de Experimentos

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

Brasília - 2023

Quando temos dois fatores, exemplo, cultivar e tipo de solo, como devemos proceder?

Experimentos Fatorias

- O caso mais simples de experimentos fatoriais envolvem apenas dois fatores: fator A com a níveis e fator B com b níveis;

- Totalizando ab combinações de tratatmentos;

- Cada replicação do experimento contém todas as combinações de tratamento ab. No geral, existem n réplicas.

Exemplo

Um engenheiro está projetando uma bateria para uso em um dispositivo que será submetido a algumas variações extremas de temperatura. O único parâmetro de projeto que ele pode selecionar neste ponto é o material da placa para a bateria, e ele tem três escolhas possíveis. Quando o dispositivo é fabricado e enviado para o campo, o engenheiro não tem controle sobre os extremos de temperatura que o dispositivo irá encontrar, e ele sabe por experiência que essa temperatura provavelmente afetará a vida útil efetiva da bateria. No entanto, a temperatura pode ser controlada no laboratório de desenvolvimento do produto para fins de teste.

Exemplo

O engenheiro decide testar todos os três materiais da placa em três níveis de temperatura: 15,70 e 125° F. Porque esses níveis de temperatura são consistentes com o uso final do produto no ambiente.

Como existem dois fatores em três níveis, esse experimento é chamado de 3^2 projeto fatorial

Quatro baterias são testadas em cada combinação de material de placa e temperatura, e todos os 36 testes são executados em ordem aleatória. O experimento e o resultado observado de duração da bateria são fornecidos na Tabela 1.

Exemplo

Tabela 1: Resultado de duração da bateria para 3 tipos de materiais e 3 diferentes temperaturas

	Temperatura					
Material	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Exemplo

Neste problema, o engenheiro deseja responder às seguintes perguntas:

1. Quais efeitos o tipo de material e a temperatura têm na vida útil da bateria?
2. Existe uma escolha de material que proporcione uma vida uniformemente longa, independentemente da temperatura?

- Seja y_{ijk} a resposta observada quando o fator A está no i-ésimo nível ($i = 1, 2, \ldots, a$) e o fator B está no j-ésimo nível ($j = 1, 2, \ldots, b$) para a k-ésima repetição ($k = 1, 2, \ldots, n$);

- As *abn* observações são selecionadas aleatoriamente segundo um experimento inteiramente casualizado.

O modelo de efeitos é definido por:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \tag{1}$$

em que $i = 1, 2, ..., a; j = 1, 2, ..., b; k = 1, 2, ..., n$; μ é a média geral, τ_i é o efeito do i-ésimo nível do fator A; β_j é o efeito do j-ésimo nível do fator B; $(\tau\beta)_{ij}$ é o efeito da interação entre τ_i e β_j e ε_{ijk} é o componente de erro aleatório.

$\rightarrow$ Ambos os fatores são assumidos ser fixos, o efeito de cada tratamento é um desvio da média geral, então $\sum_{i=1}^{a} \tau_i = 0$ e $\sum_{j=1}^{b} \beta_j = 0$;

$\rightarrow$ Similarmente os efeitos da interação são fixos e são definidos como $\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$.

As hipóteses de interesse são definidas por:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0, \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

$$\begin{cases} H_0 : \beta_1 = \beta_2 = ... = \beta_b = 0, \\ H_1 : \exists \beta_j \neq 0 \end{cases}$$

$$\begin{cases} H_0 : (\tau\beta)_{ij} = 0, \quad \text{para todo } i, j \\ H_1 : \exists (\tau\beta)_{ij} \neq 0 \end{cases}$$

ANOVA

- A soma de quadrados total corrigida SS_T pode ser escrita como:

$$
\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{\ldots})^2 &= bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{\ldots})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{\ldots})^2 \\
&+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{\ldots})^2 \\
&+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2 \qquad (2)
\end{aligned}
$$

ANOVA

A equação (2) também pode ser escrita como:

$$SQ_T = SQ_A + SQ_B + SQ_{AB} + SQ_{Res}, \tag{3}$$

com graus de liberdade dado por:

$$abn - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1) \tag{4}$$

ANOVA

As somas de quadrados também podem ser escritas por:

$$SQ_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$SQ_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$SQ_B = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

ANOVA

É conveniente obter SQ_{AB} em duas etapas. Primeiro, calculamos a soma dos quadrados entre os totais da célula ab, que é chamada de soma dos quadrados devido aos "subtotais":

$$SQ_{Subtotal} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn}$$

Essa soma de quadrados também contém as SQ_A e SQ_B . Portanto, o segundo passo é calcular SQ_{AB} como:

$$SQ_{AB} = SQ_{Subtotal} - SQ_A - SQ_B$$

ANOVA

E a soma de quadrados do resíduo é obtida por:

$$SQ_{Res} = SQ_T - SQ_{AB} - SQ_A - SQ_B$$

ou

$$SQ_{Res} = SQ_T - SQ_{Subtotal}$$

ANOVA

Tabela 1: Tabela de Análise de Variância para dois fatores

Fonte de Variação	SQ	g.l.	QM	E[QM]	F
Fator A	SQ_A	a - 1	QM_A	$\sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$	$\frac{QM_A}{QM_{Res}}$
Fator B	SQ_B	b - 1	QM_B	$\sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$	$\frac{QM_B}{QM_{Res}}$
Interação AB	SQ_{AB}	(a - 1)(b-1)	QM_{AB}	$\sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{QM_{AB}}{QM_{Res}}$
Resíduo	SQ_{Res}	ab(n-1)	QM_{Res}	σ^2	
Total	SQ_T	$abn-1$			

Comparações de médias

- Caso seja verificado pela ANOVA a diferença entre as médias de tratamentos;
- Pode-se utilizar todas as técnicas de comparaçoes múltiplas 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, τ_i, β_j, e $(\tau\beta)_{ij}$ é necessário escrever a soma dos quadrados dos erros:

$$L = \sum_i \sum_j \sum_k \varepsilon_{ijk}^2 = \sum_i \sum_j \sum_k (y_{ijk} - \mu - \tau_i - \beta_j - (\tau\beta)_{ij})^2 \qquad (5)$$

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

Estimação dos parâmetros

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

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

$$\frac{\partial L}{\partial \tau_i}|\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\hat{\tau\beta})_{ij} = 0,$$

$$\frac{\partial L}{\partial \beta_j}|\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\hat{\tau\beta})_{ij} = 0,$$

e

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

Estimação dos parâmetros

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

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

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

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

e

$$(\hat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

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:

$$\begin{aligned} e_{ijk} &= y_{ijk} - \hat{y}_{ijk} \\ &= y_{ijk} - \bar{y}_{ij.} \end{aligned} \tag{6}$$

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.