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.

- 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 hipótese 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:

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

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

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:

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

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.

Comos escolher o tamanho da amostra?

Tamanho da amostra

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

$$\begin{aligned} \beta &= P\{\text{não rejeitar} H_0 | H_0 \text{é falsa}\} \\ &= P\{F_0 < F_{\text{crítico}} | H_0 \text{é falsa}\} \end{aligned} \quad (7)$$

Tamanho da amostra

- Para avaliar a probabilidade do erro tipo II, definida na equação (7), é preciso conhecer a distribuição da estatística do teste F_0 se a hipótese nula for falsa;

- Porém, em experimentos fatorias temos mais de um teste F de interesse:

Caso 1 Teste F para fator A;
Caso 2 Teste F para fator B;
Caso 3 Teste F para interação AB;

Tamanho da amostra - Caso 1

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

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

- E esse parâmetro será igual a:

$$\delta_A = \frac{bn\sum_{i=1}^{a}\tau_i^2}{\sigma^2}, \tag{8}$$

- É possível observar que sob H_0, a equação (8) é igual a 0.

Tamanho da amostra - Caso 2

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

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

- E esse parâmetro será igual a:

$$\delta_B = \frac{an\sum_{j=1}^{b}\beta_j^2}{\sigma^2}, \tag{9}$$

- É possível observar que sob H_0, a equação (9) é igual a 0.

Tamanho da amostra - Caso 3

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

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

- E esse parâmetro será igual a:

$$\delta_{AB} = \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{\sigma^2}, \tag{10}$$

- É possível observar que sob H_0, a equação (10) é igual a 0.

É possível experimento fatorial de dois fatores SEM interação?

- Existem situações em que o pesquisador pode julgar que não é necessário o efeito de interação. Nesse caso o modelo com dois fatores sem interação é dado por:

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}, \tag{11}$$

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 e ε_{ijk} é o componente de erro aleatório.

- A análise de variância, teste de comparações de médias e análise de resíduo são semelhantes as análises realizadas para o modelo (1) que considera o termo de interação;

- Porém, deve-se ter cuidado em trabalhar com esse modelo, caso o efeito da interação seja significativo;

É possível incluir mais de dois fatores nos experimentos fatoriais?

Experimento Fatorial geral

- Os experimentos fatoriais podem ser estendidos para o caso geral;
- Em que terá a níveis do fator A, b níveis do fator B, c níveis do fator C, e assim por diante;
- Em geral, haverá $abc...n$ total de observações se houver n repetições do experimento;
- Sendo que, é importante observar, que devemos ter pelo menos $n = 2$ repetições para determinar a soma de quadrados devido ao erro se todas as interações possíveis forem incluídas no modelo.

Experimento Fatorial geral

- Se todos os fatores do experimento forem fixos, podemos facilmente formular e testar hipóteses sobre os principais efeitos e interações usando a ANOVA;

- Para um modelo de efeitos fixos, testes estatísticos para cada efeito principal e interação podem ser construídos dividindo o correspondente quadrado médio para o efeito ou interação pelo erro quadrado médio;

- O número de graus de liberdade para qualquer efeito principal é o número número de níveis do fator menos um;

- E o número de graus de liberdade para uma interação é o produto do número de graus de liberdade associados aos componentes individuais da interação.

Experimento Fatorial geral

O modelo de análise de variância para três fatores é definido por:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}, \quad (12)$$

em que $i = 1, 2, ..., a; j = 1, 2, ..., b; k = 1, 2, ..., c$; $l = 1, 2, ..., n$; 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; γ_k é o efeito do k-ésimo nível do fator C; $(\tau\beta)_{ij}$ é o efeito da interação entre τ_i e β_j; $(\tau\gamma)_{ik}$ é o efeito da interação entre τ_i e γ_k; $(\beta\gamma)_{jk}$ é o efeito da interação entre β_j e γ_k; $(\tau\beta\gamma)_{ijk}$ é o efeito da interação entre τ_i, β_j e γ_k; ε_{ijk} é o componente de erro aleatório.

Experimento Fatorial geral

Tabela 2: Tabela de Análise de Variância para três fatores

Fonte de Variação	SQ	g.l.	E[QM]	F
Fator A	SQ_A	a - 1	$\sigma^2 + \frac{bcn\sum_{i=1}^{a}\tau_i^2}{a-1}$	$\frac{QM_A}{QM_{Res}}$
Fator B	SQ_B	b - 1	$\sigma^2 + \frac{acn\sum_{j=1}^{b}\beta_j^2}{b-1}$	$\frac{QM_B}{QM_{Res}}$
Fator C	SQ_C	c - 1	$\sigma^2 + \frac{abn\sum_{k=1}^{c}\gamma_k^2}{c-1}$	$\frac{QM_C}{QM_{Res}}$
AB	SQ_{AB}	(a - 1)(b-1)	$\sigma^2 + \frac{cn\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{QM_{AB}}{QM_{Res}}$
AC	SQ_{AC}	(a - 1)(c-1)	$\sigma^2 + \frac{bn\sum_{i=1}^{a}\sum_{k=1}^{c}(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\frac{QM_{AC}}{QM_{Res}}$
BC	SQ_{BC}	(b - 1)(c-1)	$\sigma^2 + \frac{an\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\frac{QM_{BC}}{QM_{Res}}$
ABC	SQ_{ABC}	(a-1)(b-1)(c-1)	$\sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\frac{QM_{ABC}}{QM_{Res}}$
Resíduo	SQ_{Res}	abc(n-1)	σ^2	
Total	SQ_T	$abcn-1$		

Exemplo

Uma engarrafadora de refrigerantes está interessada em encher garrafas mais uniformes. A máquina de envase teoricamente enche cada garrafa para a altura correta do alvo, mas, na prática, há variação em torno deste alvo, e o engarrafador gostaria de entender as fontes dessa variabilidade e, eventualmente, reduzí-la.

Exemplo

O engenheiro do processo pode controlar três variáveis durante o processo de enchimento: a carbonatação percentual (A), a operação pressão de enchimento na enchedora (B), e as garrafas produzidas por minuto ou a velocidade da linha (C). A pressão e a velocidade são fácil de controlar, mas a porcentagem de carbonatação é mais difícil controlar durante a fabricação real porque varia com a temperatura do produto. Entretanto, para fins de experimento, o engenheiro pode controlar a carbonatação em três níveis: 10, 12 e 14 por cento.

Exemplo

Ela escolhe dois níveis para pressão (25 e 30 psi) e dois níveis de velocidade de linha (200 e 250 bpm). Ela decide executar duas repetições de um experimento fatorial com três fatores, com todas as 24 execuções feitas em ordem aleatória.
A variável resposta observada é a média do desvio da altura de preenchimento observada. Resultado positivo são alturas de preenchimento acima do alvo, enquanto resultado negativo são alturas de preenchimento abaixo do alvo.