Análise de Variância dois fatores

Modelo sem interação

Celma de Oliveira Ribeiro/ Linda Lee Ho

- Objetivo: testar a hipótese de que três lojas têm volumes médios de vendas iguais, considerando dois fatores: **Loja e desempenho dos empregados**

- **Exemplo:**

  Uma empresa tem cinco empregados estão igualmente familiarizados com três lojas da rede. Foram obtidos dados das vendas dos cinco empregados nas três lojas

***Pergunta:* a variação nas vendas é explicada apenas pelas lojas , ou também pode ser explicada pelo desempenho dos empregados?**

Celma de Oliveira Ribeiro/Linda Lee Ho

- Dados:

	Loja 1	Loja 2	Loja 3	Médias
Funcionário 1	53	61	51	55
Funcionário 2	47	55	51	51
Funcionário 3	46	52	49	49
Funcionário 4	50	58	54	54
Funcionário 5	49	54	50	51
Médias	49	56	51	52

Celma de Oliveira Ribeiro/Linda Lee Ho

- **Hipóteses:**

  normalidade,

  igualdade de variâncias,

  independência entre as observações

  Inexistência de interação entre os dois fatores.

- Ausência de interação entre os fatores:

  Em termos médios, a diferença entre dois quaisquer níveis do fator A não depende do nível do fator B, isto é, é igual para todos os níveis do fator B, e vice-versa.

  Os empregados estão igualmente familiarizados com todas as lojas e mantêm o mesmo comportamento em todas elas(em média, a diferença entre o desempenho do empregado i e do empregado j é igual para todas as lojas).

Celma de Oliveira Ribeiro/Linda Lee Ho

- Caso geral:

	B1	B2		BJ	
A1	x11	x12	....	x1J	$\bar{x}_{1\bullet}$
A2	x21	x22		x2J	$\bar{x}_{2\bullet}$
AI	xI1	xI2		XIJ	$\bar{x}_{I\bullet}$
	$\bar{x}_{\bullet 1}$	$\bar{x}_{\bullet 2}$		$\bar{x}_{\bullet J}$	

$$\bar{x}_{i\bullet} = \frac{\sum_{j=1}^{J} x_{ij}}{J} \qquad \bar{x}_{\bullet j} = \frac{\sum_{i=1}^{I} x_{ij}}{I} \qquad \bar{x}_{\bullet\bullet} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}}{IJ}$$

Celma de Oliveira Ribeiro/Linda Lee Ho

- Definição

$$SQT = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij} - \bar{x}_{\bullet\bullet}\right)^2 \qquad gl = I \times J - 1$$

$$SQA = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet}\right)^2 = J\sum_{i=1}^{I}\left(x_{i\bullet} - \bar{x}_{\bullet\bullet}\right)^2_{\bar{x}_{\bullet j}} \qquad gl = I - 1$$

$$SQB = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\bar{x}_{\bullet j} - \bar{x}_{\bullet\bullet}\right)^2 = I\sum_{j=1}^{J}\left(x_{\bullet j} - \bar{x}_{\bullet\bullet}\right)^2 \qquad gl = J - 1$$

$$SQE = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet}\right)^2 \qquad gl = (I - 1) \times (J - 1)$$

$$SQT = SQA + SQB + SQE$$

- Há três fontes de variação:
  1. variação devida ao fator Loja (medida por **SQA ou SQLoja);**
  2. variação devida ao fator Funcionário (medida por **SQB ou SSFunc);**
  3. variação não explicada pelo modelo (medida por **SQE),**

Celma de Oliveira Ribeiro/Linda Lee Ho

- Da mesma forma que no caso de um fator

Hipótese	**Valor da estatística do teste**	**Rejeição**
H_{0A} versus H_{1A}	$f_A = \frac{QMA}{QME}$	$f_A > F_{\alpha, I-1, (I-1)\times(J-1)}$
H_{0B} versus H_{1B}	$f_B = \frac{QMB}{QME}$	$f_B > F_{\alpha, J-1, (I-1)\times(J-1)}$

Celma de Oliveira Ribeiro/Linda Lee Ho

- Cálculos parciais (fator A = empregado = linhas)

$\bar{x}_{1\bullet}$	55
$\bar{x}_{2\bullet}$	51
$\bar{x}_{3\bullet}$	49
$\bar{x}_{4\bullet}$	54
$\bar{x}_{5\bullet}$	51

$$\bar{x}_{\bullet\bullet} = 52$$

$$SQA = J\sum_{i=1}^{I}(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2$$

$$SQA = 3\sum_{i=1}^{5}(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 = 72$$

- Cálculos parciais (fator B = lojas = colunas)

$\bar{x}_{\bullet 1}$	49
$\bar{x}_{\bullet 2}$	56
$\bar{x}_{\bullet 3}$	51

$$SQB == I\sum_{j=1}^{J}\left(x_{\bullet j} - \bar{x}_{\bullet\bullet}\right)^2$$

$$SQB = 5\sum_{j=1}^{J}\left(x_{\bullet j} - \bar{x}_{\bullet\bullet}\right)^2 = 130$$

$$\bar{x}_{\bullet\bullet} = 52$$

Celma de Oliveira Ribeiro/Linda Lee Ho

- Cálculo do erro total

$$SQT = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij} - \bar{x}_{\bullet\bullet}\right)^2 = 224$$

(IJ-1)*Variância amostral

- Resultado

ANOVA

Fonte da variação	*SQ*	*gl*	*MQ*	*F*	*valor-P*	*F crítico*
Linhas	72	4	18	6,545	0,012	3,83785
Colunas	130	2	65	23,64	4E-04	4,45897
Erro	22	8	2,75			
Total	224	14				

Celma de Oliveira Ribeiro/Linda Lee Ho

- Testamos

$$\begin{cases} H_0 : \mu^F{}_1 = \mu^F{}_2 = \mu^F{}_3 = \mu^F{}_4 = \mu^F{}_5 \\ H_1 : \exists(i, j) \quad \mu^F{}_i \neq \mu^F{}_j \end{cases}$$

$$\begin{cases} H_0 : \mu^L{}_1 = \mu^L{}_2 = \mu^L{}_3 \\ H_1 : \exists(i, j) \quad \mu^L{}_i \neq \mu^L{}_j \end{cases}$$

- Podemos concluir ao nível de significância de 5%, não só que as lojas são significativamente diferentes, mas também que existem diferenças entre os empregados, no que diz respeito ao volume de vendas semanais. Ou seja tanto o **fator Loja como o fator Funcionário afetam o volume de vendas.**

- O modelo analisa: $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

- As hipóteses sendo testadas são:

$$H_{0A} : \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0 \qquad H_{0B} : \beta_1 = \beta_2 = \ldots = \beta_J = 0$$

$$H_{1A} : \exists\, i : \alpha_i \neq 0 \qquad H_{1B} : \exists\, i : \beta_i \neq 0$$

Exemplo

Os dados a seguir comparam diferentes marcas de caneta e quatro tipos de tratamento de lavagem em relação à capacidade de remover manchas em um certo tipo de tecido.

A variável resposta indica a mudança da cor, e quanto menor mais manchas foram removidas.

- Qual a conclusão ao nível de 5% de significância?

	1	2	3	4
1	0,97	0,48	0,48	0,46
2	0,77	0,14	0,22	0,25
3	0,67	0,39	0,57	0,19

Marca da caneta

Tipo de lavagem

Celma de Oliveira Ribeiro/Linda Lee Ho

- Resultado

RESUMO	*Contagem*	*Soma*	*Média*	*Variância*
Linha 1	4	2,39	0,5975	0,061758
Linha 2	4	1,38	0,345	0,082433
Linha 3	4	1,82	0,455	0,044633
Coluna 1	3	2,41	0,803333	0,023333
Coluna 2	3	1,01	0,336667	0,031033
Coluna 3	3	1,27	0,423333	0,033033
Coluna 4	3	0,9	0,3	0,0201

A alteração de cor
não parece depender do tipo
de caneta

ANOVA

Fonte da varia	*SQ*	*gl*	*MQ*	*F*	*valor-P*	*F crítico*
Linhas	0,128217	2	0,064108	4,432303	0,065765	5,143253
Colunas	0,479692	3	0,159897	11,05493	0,007399	4,757063
Erro	0,086783	6	0,014464			
Total	0,694692	11				

A alteração de cor
depende do tipo
de lavagem

Teste de Tukey

Estabelece a diferença mínima entre médias amostrais que permite identificar com significância estatística que as médias são diferentes

$$DMS = Q_{\alpha, K, (I-1)(J-1)} \sqrt{\frac{QME}{N}}$$

K = número de níveis sendo testado naquele fator

Tukey deve ser adotado quando há interesse em todas as possíveis comparações de médias duas a duas