Análise de Variância com Dois Factores

Modelo sem interacção

Exemplo 2

Neste exemplo, ao testarmos a hipótese de as três lojas terem volumes médios de vendas iguais, estamos a testar se o **factor Loja** tem influência no volume de vendas.

Note que o volume de vendas deve também sofrer influência de outros factores.

Assim, a variação nas vendas pode estar relacionada não só com a loja, mas também com o desempenho do empregado. Vamos então introduzir no nosso estudo um segundo factor, o **factor Empregado**.

Exemplo 4

Admitamos então que o Sr Fernando tem cinco empregados que estão igualmente familiarizados com as três lojas. Os dados recolhidos das vendas dos cinco empregados nas três lojas (por conveniência, os mesmos apresentados anteriormente) são os seguintes:

		Factor Loja			Médias dos Empregados
		Loja 1	**Loja 2**	**Loja 3**	$\bar{x}_{\bullet j}$
Factor Empregado	**Emp 1**	53	61	51	55
	Emp 2	47	55	51	51
	Emp 3	46	52	49	49
	Emp 4	50	58	54	54
	Emp 5	49	54	50	51
	Médias das Lojas $\bar{x}_{i\bullet}$	49	56	51	$\bar{\bar{x}}=52$

O **factor Empregado** tem cinco níveis (**Emp1**, **Emp2**,..., **Emp5**) e o **factor Loja** tem três níveis (**Loja 1**, **Loja 2** e **Loja3**).

Os dados amostrais estão organizados de acordo com um esquema designado por **classificação cruzada**, uma vez que cada nível de um factor é cruzado com cada nível do outro factor.

Uma observação mais atenta da tabela anterior, mostra que há empregados que, aparentemente, apresentam melhores resultados do que outros.

Deste modo, é razoável pensar que **talvez as lojas não sejam assim tão diferentes umas das outras, no que diz respeito ao volume de vendas, pode é haver também diferenças no desempenho dos empregados**.

Podemos então perguntar:

A variação nas vendas é explicada apenas pelas lojas onde são efectuadas, ou será que também pode ser explicada pela performance dos empregados?

No **Exemplo 4** estamos perante um problema ao qual vamos aplicar um outro modelo da ANOVA, a **ANOVA com dois factores** (para o exemplo, factor Loja e factor Empregado), ainda sob os pressupostos de **normalidade**, **igualdade de variâncias**, **independência entre as observações** e assumindo adicionalmente que **não há interacção entre os dois factores**.

A **ausência de interacção** entre os factores significa que, ***em termos médios, a diferença entre dois quaisquer níveis do factor A não depende do nível do factor B, isto é, é igual para todos os níveis do factor B, e vice-versa***.

Para o **Exemplo 4** isto significa que os empregados estão igualmente familiarizados com todas as lojas e portanto mantêm o mesmo comportamento em todas elas.

Sendo assim, em média, a diferença entre o desempenho do ***empregado i*** e do ***empregado j*** é igual para todas as lojas.

Por outro lado, a diferença entre a ***loja i*** e a ***loja j*** é igual para todos os empregados.

De um modo geral, os dados com dois factores, o **Factor A** (ou **Factor Coluna**) com ***a*** níveis/grupos e o **Factor B** (**Factor Linha**) com ***b*** níveis, são apresentados numa tabela como a seguinte:

Factor A

Factor B	A_1	A_2	...	A_a	$\bar{x}_{\bullet j}$
B_1	x_{11}	x_{21}	...	x_{a1}	$\bar{x}_{\bullet 1}$
B_2	x_{12}	x_{22}	...	x_{a2}	$\bar{x}_{\bullet 2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
B_b	x_{1b}	x_{2b}	...	x_{ab}	$\bar{x}_{\bullet b}$
$\bar{x}_{i\bullet}$	$\bar{x}_{1\bullet}$	$\bar{x}_{2\bullet}$	...	$\bar{x}_{a\bullet}$	$\bar{\bar{x}}$

onde,

$$\bar{x}_{\bullet j}=\frac{1}{a}\sum_{i=1}^{a}x_{ij}\ , \qquad \bar{x}_{i\bullet}=\frac{1}{b}\sum_{j=1}^{b}x_{ij}$$

e

$$\bar{\bar{x}}=\frac{\sum_{j=1}^{b}\bar{x}_{\bullet j}+\sum_{i=1}^{a}\bar{x}_{i\bullet}}{a+b}=\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}}{ab}=\frac{\sum_{j=1}^{b}\bar{x}_{\bullet j}}{b}=\frac{\sum_{i=1}^{a}\bar{x}_{i\bullet}}{a}.$$

Recordemos que quando aplicamos a ANOVA com apenas um factor, as fontes de variação dos dados são duas:

- **variação entre os grupos** ou níveis do factor (**SS_A**)
- **variação que provem das flutuações aleatórias dentro dos grupos**, **SS_E**, e que fica por explicar (residual).

Aplicando o modelo de ANOVA com dois factores, **esperamos reduzir a variação não explicada**, uma vez que esta pode provir da variação entre os grupos do segundo factor e essa passa a ser "contabilizada".

Para o **Exemplo 2** tínhamos,

Variação total nos dados, $SS_T = 224$,

Variação entre os níveis do factor Loja, $SS_A=130$,

Variação não explicada ou residual, $SS_E=94$,

onde,

$$SS_T=SS_A+SS_E$$

Introduzindo um segundo factor, o factor Empregado, esperamos, como já dissemos, reduzir a variação não explicada, pois, parte desta passa a ser explicada pela variação no desempenho dos empregados.

Passamos a ter então três fontes de variação:

- a variação devida ao factor Loja (medida por **SS_A** ou **SS_{Loja}**);
- a variação devida ao factor Empregado (medida por **SS_B** ou **SS_{Emp}**);
- a variação não explicada pelo modelo (medida por **SS_E**),

verificando-se agora,

$$SS_T = SS_A + SS_B + SS_E$$

Os cálculos são muito semelhantes aos efectuados na análise anterior, mas agora com mais um factor. Assim, consideramos:

- **Soma dos quadrados entre os grupos ou níveis do factor A**:

$$SS_A = b\sum_{i=1}^{a}(\bar{x}_{i\bullet} - \bar{\bar{x}})^2\,;$$

- **Soma dos quadrados entre os grupos ou níveis do factor B**:

$$SS_B = a\sum_{j=1}^{b}(\bar{x}_{\bullet j} - \bar{\bar{x}})^2\,.$$

Exemplo 4

Os cálculos destas medidas são resumidos nos quadros seguintes:

Factor Loja		
	$(\bar{x}_{i\bullet} - \bar{\bar{x}})$	$(\bar{x}_{i\bullet} - \bar{\bar{x}})^2$
	-3	9
	4	16
	-1	1
Totais	0	26

$SS_{Loja} = 5 \times 26 = 130$

Factor Empregado		
	$(\bar{x}_{\bullet j} - \bar{\bar{x}})$	$(\bar{x}_{\bullet j} - \bar{\bar{x}})^2$
	3	9
	-1	1
	-3	9
	2	4
	-1	1
Totais	0	24

$SS_{Emp} = 3 \times 24 = 72$

O seguinte passo é o cálculo da soma dos quadrados residual, **SS_E**, a variação não explicada pelo modelo:

$$\hat{x}_{ij} = \overline{\overline{x}} + (\overline{x}_{i\bullet} - \overline{\overline{x}}) + (\overline{x}_{\bullet j} - \overline{\overline{x}}), \qquad i = 1,...,a \qquad j = 1,...,b$$

Cada resíduo é dado por

$$x_{ij} - \hat{x}_{ij} = x_{ij} - \overline{x}_{i\bullet} - \overline{x}_{\bullet j} + \overline{\overline{x}}$$

e tem-se

$$\boxed{\mathbf{SS_E} = \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \overline{x}_{i\bullet} - \overline{x}_{\bullet j} + \overline{\overline{x}})^2}$$

Exemplo 4

$\hat{x}_{ij} = \overline{\overline{x}} + (\overline{x}_{i\bullet} - \overline{\overline{x}}) + (\overline{x}_{\bullet j} - \overline{\overline{x}})$		
52	59	54
48	55	50
46	53	48
51	58	53
48	55	50

$x_{ij} - \hat{x}_{ij} = x_{ij} - \overline{x}_{i\bullet} - \overline{x}_{\bullet j} + \overline{\overline{x}}$		
1	2	-3
-1	0	1
0	-1	1
-1	0	1
1	-1	0

$SS_E = 1^2 + 2^2 + (-3)^2 + (-1)^2 + ... + (-1)^2 + 0^2 = 22$

Comparando com o **Exemplo 2**, salienta-se que se **reduziu a variação não explicada pelo modelo de 94 para 22**. De facto, a variação não explicada no **Exemplo 2**, que valia **94**, está agora decomposta em duas parcelas, a **variação explicada pelo factor Empregado** (**72**) e a **variação residual** (**22**) - a variação que continua por explicar.

Finalmente, a soma dos quadrados total, a que mede a variação total dos dados, que já foi calculada no **Exemplo 2**:

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{\bar{x}})^2=224$$

A **Tabela ANOVA com dois factores** tem o mesmo formato que a de um factor, e é construída do seguinte modo:

Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberd.	Variância (Soma Média dos Quadrados)	Razões F
Entre Grupos Factor A	$SS_A=b\sum_{i=1}^{a}(\bar{x}_{i\bullet}-\bar{\bar{x}})^2$	a-1	$MS_A=\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Entre Grupos Factor B	$SS_B=a\sum_{j=1}^{b}(\bar{x}_{\bullet j}-\bar{\bar{x}})^2$	b-1	$MS_B=\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Residual	$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{x}_{i\bullet}-\bar{x}_{\bullet j}+\bar{\bar{x}})^2$	(a-1)(b-1)	$MS_E=\frac{SS_E}{(a-1)(b-1)}$	
Total	$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij}-\bar{\bar{x}})^2$	ab-1		

Para o **Exemplo 4**, temos a seguinte **Tabela ANOVA**:

Fonte de Variação	Soma dos Quadrados (SS)	g.l.	Variância (Soma Média dos Quadrados)	Razões F
Entre grupos Lojas	$SS_{Loja}=130$	2	$MS_{Loja}=65$	$\frac{MS_{Loja}}{MS_E}=23.6$
Entre grupos Empregados	$SS_{Emp}=72$	4	$MS_{Emp}=18$	$\frac{MS_{Emp}}{MS_E}=6.5$
Residual	$SS_E=22$	8	$MS_E=2.75$	
Total	$SS_T=224$	14		

Testamos por um lado,

H_0: $\mu_1 = \mu_2 = \mu_3$ (os volumes médios de vendas são iguais nas três lojas)

H_1: $\mu_i \neq \mu_j$ para algum $i \neq j$ (existem pelo menos duas lojas com volumes médios de vendas diferentes)

Sob H_0, $$F = \frac{MS_{Loja}}{MS_E} \sim F^{a-1}_{(a-1)(b-1)}.$$

Tem-se:

- Quantil de probabilidade (1-0.05) da distribuição F_8^2: 4.46
- R.C.: [4.46, +∞[
- $F_{obs}=23.6 \in$ R.C., logo rejeitamos H_0, tal como na aplicação da ANOVA com apenas o factor Loja

Notemos, no entanto, que o valor **observado da estatística de teste F é neste caso maior** do que o obtido na análise anterior (**23.6**>**8.3**) - a variação não explicada é menor. **A rejeição de H_0 é neste caso ainda mais "forte".**

Por outro lado, também podemos testar

***H*₀**: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (os cinco emp.'s têm volumes médios de vendas iguais)

***H*₁**: $\mu_i \neq \mu_j$ para algum $i \neq j$ (existem pelo menos dois empregados com volumes médios de vendas diferentes)

Sob H_0, $F = \dfrac{MS_{Emp}}{MS_E} \sim F^{b-1}_{(a-1)(b-1)}$.

Tem-se:

- Quantil de probabilidade (1-0.05) da distribuição F_8^4: 3.84
- R.C.: [3.84, +∞[
- $F_{obs}=6.5 \in$ R.C., logo rejeitamos H_0.

Podemos concluir que os dados amostrais revelaram, 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 e, deste modo, tanto o **factor Loja como o factor Empregado afectam o volume de vendas**.

Comparações Múltiplas

Após a rejeição de H_0, tem sentido estudar quais os grupos que diferem entre si, em cada factor. O teste que vamos considerar é, uma vez mais, o **teste de Tuckey**.

Para o Factor A

A hipótese nula ***H*₀: $\mu_r = \mu_s$** (os grupos *r* e *s* do factor A têm médias iguais) é rejeitada se

$$|\bar{x}_{r\bullet} - \bar{x}_{s\bullet}| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{2}\left(\frac{1}{b}+\frac{1}{b}\right)} \quad \text{ou} \quad |\bar{x}_{r\bullet} - \bar{x}_{s\bullet}| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{b}}$$

onde,

- ✓ $S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da ***"Studentized Range"*** com (a, (a-1)(b-1)) graus de liberdade;
- ✓ $MS_E=\dfrac{SS_E}{(a-1)(b-1)}$;
- ✓ b é a dimensão das amostras de cada um dos grupos do factor A, neste caso coincidente com o número de grupos do factor B.

Exemplo 4

Factor Loja,

Para α=0.05, tem-se $4.04\sqrt{\frac{2.75}{5}}$=2.996 e

$$|\bar{x}_{1\bullet} - \bar{x}_{2\bullet}| = |49-56| = 7 > 2.996$$

$$|\bar{x}_{1\bullet} - \bar{x}_{3\bullet}| = |49-51| = 2 < 2.996$$

$$|\bar{x}_{2\bullet} - \bar{x}_{3\bullet}| = |56-51| = 5 > 2.996.$$

Confirmamos assim o resultado obtido anteriormente, i.e., que a loja 2 (grupo 2) difere significativamente das lojas 1 e 3, no que diz respeito ao volume médio de vendas.

Para o Factor B

A hipótese nula $\boldsymbol{H_0: \mu_r = \mu_s}$ (os grupos *r* e *s* do factor B têm médias iguais) é rejeitada se

$$|\bar{x}_{\bullet r} - \bar{x}_{\bullet s}| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{a}}$$

onde,

- ✓ $S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da ***"Studentized Range"*** com (b, (a-1)(b-1)) graus de liberdade;
- ✓ a é a dimensão das amostras de cada um dos grupos do factor B, neste caso coincidente com o número de grupos do factor A.

Exemplo 4

Factor Empregado,

Para α=0.05, tem-se $4.89\sqrt{\frac{2.75}{3}}$=4.682 e

$$\begin{array}{ll}
\left|\bar{x}_{\bullet 1}-\bar{x}_{\bullet 2}\right|=|55\text{-}51|=4 & \left|\bar{x}_{\bullet 2}-\bar{x}_{\bullet 4}\right|=|51\text{-}54|=3 \\
\left|\bar{x}_{\bullet 1}-\bar{x}_{\bullet 3}\right|=|55\text{-}49|=6>4.682 & \left|\bar{x}_{\bullet 2}-\bar{x}_{\bullet 5}\right|=|51\text{-}51|=0 \\
\left|\bar{x}_{\bullet 1}-\bar{x}_{\bullet 4}\right|=|55\text{-}54|=1 & \left|\bar{x}_{\bullet 3}-\bar{x}_{\bullet 4}\right|=|49\text{-}54|=5>4.682 \\
\left|\bar{x}_{\bullet 1}-\bar{x}_{\bullet 5}\right|=|55\text{-}51|=4 & \left|\bar{x}_{\bullet 3}-\bar{x}_{\bullet 5}\right|=|49\text{-}51|=2 \\
\left|\bar{x}_{\bullet 2}-\bar{x}_{\bullet 3}\right|=|51\text{-}49|=2 & \left|\bar{x}_{\bullet 4}-\bar{x}_{\bullet 5}\right|=|54\text{-}51|=3
\end{array}$$

Há evidência de que o empregado 3 tem um volume médio de vendas diferente dos empregados 1 e 4. Observando as médias amostrais, podemos verificar que essa diferença é favorável aos empregados 1 e 4.

Modelo com interacção

O modelo de ANOVA com dois factores que apresentámos não contempla a interacção entre os dois factores.

De facto, alguma da variação existente nos dados pode ter ainda origem na interacção entre os dois factores, e esta deve de ser pesada na análise.

No entanto, para levar a cabo esta análise são necessárias mais observações por célula, dando origem a uma estrutura de dados mais complexa.

Num modelo onde se considera a interacção entre os dois factores, o Factor A e o Factor B, os dados são em geral apresentados numa tabela como a que se segue.

Factor A

Factor B		**A₁**		**A₂**			**Aₐ**		$\bar{x}_{\bullet j \bullet}$
	B₁	$x_{111},\ldots,x_{11n}$	$\bar{x}_{11\bullet}$	$x_{211},\ldots,x_{21n}$	$\bar{x}_{21\bullet}$	…	$x_{a11},\ldots,x_{a1n}$	$\bar{x}_{a1\bullet}$	$\bar{x}_{\bullet 1 \bullet}$
	B₂	$x_{121},\ldots,x_{12n}$	$\bar{x}_{12\bullet}$	$x_{221},\ldots,x_{22n}$	$\bar{x}_{22\bullet}$	…	$x_{a21},\ldots,x_{a2n}$	$\bar{x}_{a2\bullet}$	$\bar{x}_{\bullet 2 \bullet}$
	⋮	⋮		⋮			⋮		⋮
	B_b	$x_{1b1},\ldots,x_{1bn}$	$\bar{x}_{1b\bullet}$	$x_{2b1},\ldots,x_{2bn}$	$\bar{x}_{2b\bullet}$	…	$x_{ab1},\ldots,x_{abn}$	$\bar{x}_{ab\bullet}$	$\bar{x}_{\bullet b \bullet}$
	$\bar{x}_{i\bullet\bullet}$	$\bar{x}_{1\bullet\bullet}$		$\bar{x}_{1\bullet\bullet}$		…	$\bar{x}_{a\bullet\bullet}$		$\bar{\bar{x}}$

onde,

- $\bar{x}_{\bullet j \bullet} = \dfrac{1}{na}\sum_{i=1}^{a}\sum_{k=1}^{n} x_{ijk}$
- $\bar{x}_{i \bullet\bullet} = \dfrac{1}{nb}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}$
- $\bar{x}_{ij\bullet} = \dfrac{1}{n}\sum_{k=1}^{n} x_{ijk}$
- $\bar{\bar{x}} = \dfrac{1}{nab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}$

Note que, cada célula, isto é, cada combinação possível entre níveis do factor A com níveis do factor B, contém ***n*** observações, sendo portanto o número total de observações igual a ***nab***.

Exemplo 5

Retomemos os Exemplo 4, mas agora admitindo a possibilidade de existência de interacção entre o factor Loja e o factor Empregado.

Vamos então aplicar o modelo de análise de variância com interacção, o que nos obriga a ter mais do que uma observação por cada combinação Loja-Empregado.

Assim, suponhamos que os dados recolhidos pelo Sr. Fernando foram os seguintes (consideramos apenas três empregados para facilitar os cálculos):

Loja

		1		**2**		**3**		$\bar{x}_{\bullet j \bullet}$
Empregado	**1**	53, 52, 54	53	53, 56, 56	55	52, 56, 54	54	**54**
	2	41, 46, 45	44	48, 51, 51	50	48, 48, 45	47	**47**
	3	51, 54, 54	53	54, 56, 52	54	48, 51, 48	49	**52**
	$\bar{x}_{i \bullet \bullet}$	**50**		**53**		**50**		**51**

Pretende-se testar:

1. H_0^1: os volumes médios de vendas são iguais nas três lojas

H_1^1: existem pelo menos duas lojas com volumes médios de vendas diferentes

2. H_0^2: os três empregados têm volumes médios de vendas iguais

H_1^2: existem pelo menos dois empregados com volumes médios de vendas diferentes

3. H_0^3: não existe interacção entre o factor Loja e o factor Empregado

H_1^3: existe interacção entre o factor Loja e o factor Empregado

Num modelo com interacção a variação total dos dados é decomposta em quatro parcelas:

- ★ a variação devida ao factor A (**SS_A**);
- ★ a variação devida ao factor B (**SS_B**);
- ★ a variação devida à interacção (**SS_I**);
- ★ a variação residual (**SS_E**) que é a variação não explicada pelo modelo.

Mais uma vez os cálculos a efectuar são muito semelhantes aos das análises anteriores:

$$SS_A = nb\sum_{i=1}^{a}(\bar{x}_{i\bullet\bullet}-\bar{\bar{x}})^2 \qquad SS_B = na\sum_{j=1}^{b}(\bar{x}_{\bullet j\bullet}-\bar{\bar{x}})^2$$

$$SS_I = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij\bullet}-\bar{x}_{i\bullet\bullet}-\bar{x}_{\bullet j\bullet}+\bar{\bar{x}})^2 \qquad SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij\bullet})^2$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{\bar{x}})^2$$

com, $SS_T = SS_A + SS_B + SS_I + SS_E$

A **Tabela ANOVA para o modelo com interacção** é a seguinte:

Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberdade	Variância (Soma Média Quadrados)	Razões F
Factor A	$SS_A = nb\sum_{i=1}^{a}(\bar{x}_{i\bullet\bullet}-\bar{\bar{x}})^2$	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Factor B	$SS_B = na\sum_{j=1}^{b}(\bar{x}_{\bullet j\bullet}-\bar{\bar{x}})^2$	b-1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Interacção	$SS_I = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij\bullet}-\bar{x}_{i\bullet\bullet}-\bar{x}_{\bullet j\bullet}+\bar{\bar{x}})^2$	(a-1)(b-1)	$MS_I = \frac{SS_I}{(a-1)(b-1)}$	$\frac{MS_I}{MS_E}$
Residual	$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij\bullet})^2$	ab(n-1)	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{\bar{x}})^2$	abn-1		

Para o **Exemplo 5** temos:

Soma de quadrados

Factor Loja

				Total
$(\bar{x}_{i\bullet\bullet} - \bar{\bar{x}})^2$	1	4	1	6

$SS_{Loja}=3\times 3\times 6=54$

Factor Empregado

				Total
$(\bar{x}_{\bullet j\bullet} - \bar{\bar{x}})^2$	9	16	1	26

$SS_{Emp}=3\times 3\times 26=234$

Interacção

$(\bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} + \bar{\bar{x}})^2$			
0	1	1	
4	1	1	
4	0	4	Total
			16

$SS_I=3\times 16=48$

Residual

$(x_{ijk} - \bar{x}_{ij\bullet})^2$			
0, 1, 1	4, 1, 1	4, 4, 0	
9, 4, 1	4, 1, 1	1, 1, 4	
4, 1, 1	0, 4, 4	1, 4, 1	Total
			62

$SS_E=62$

Total

$(x_{ijk} - \bar{\bar{x}})^2$			
4, 1, 9	4, 25, 25	1, 25, 9	
100, 25, 36	9, 0, 0	9, 9, 36	
0, 9, 9	9, 25, 1	9, 0, 9	Total
			398

$SS_T=398$

A Tabela ANOVA é então,

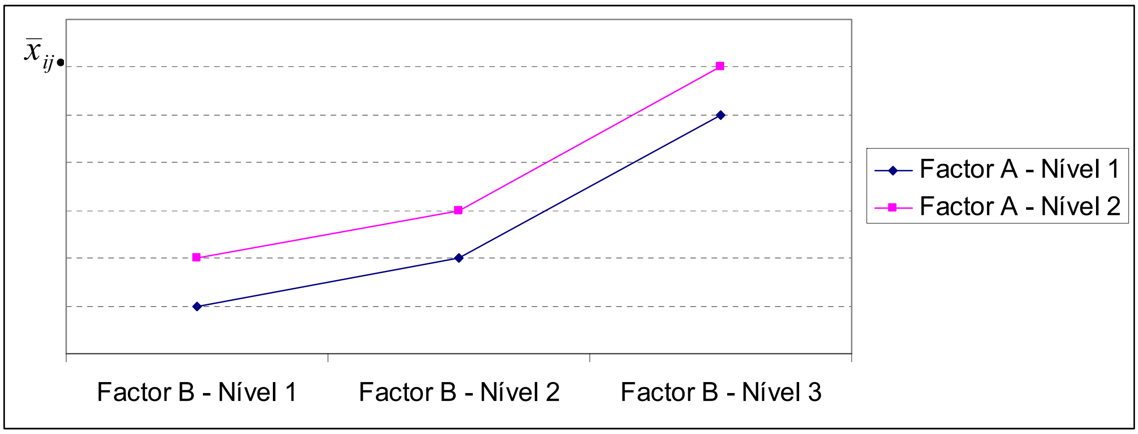
Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberdade	Variância (Soma Média dos Quadrados)	Razões F
Loja	$SS_{Loja}= 54$	2	$MS_{Loja}=27$	7.85
Empregado	$SS_{Emp}= 234$	2	$MS_{Emp}=117$	34.01
Interacção	$SS_{I}= 48$	4	$MS_{I}=12$	3.49
Residual	$SS_{E}= 62$	18	$MS_{E}=3.44$	
Total	$SS_{T}= 398$	26		

Salienta-se que **quando existe interacção entre os dois factores o efeito de um deles depende dos níveis do outro**.

Assim, na presença de uma interacção significativa o efeito de cada um dos factores isoladamente pode ser "mascarado" pela interacção e, consequentemente, os **testes à significância da influência de cada um dos factores podem ficar desprovidos de sentido**.

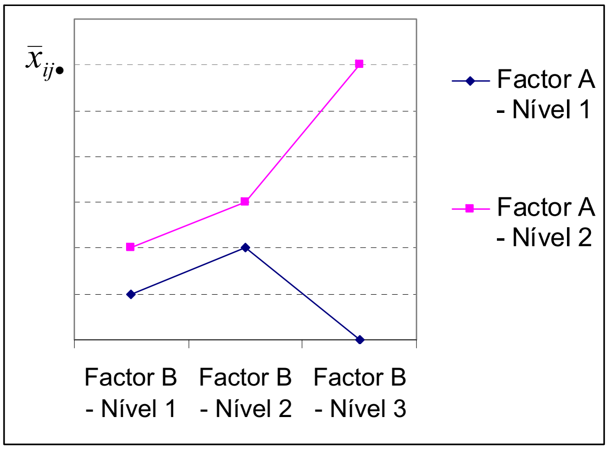
Por esta razão, em primeiro lugar deve-se fazer o teste relativo à interacção, isto é, deve-se testar a hipótese nula de que não existe interacção entre os dois factores.

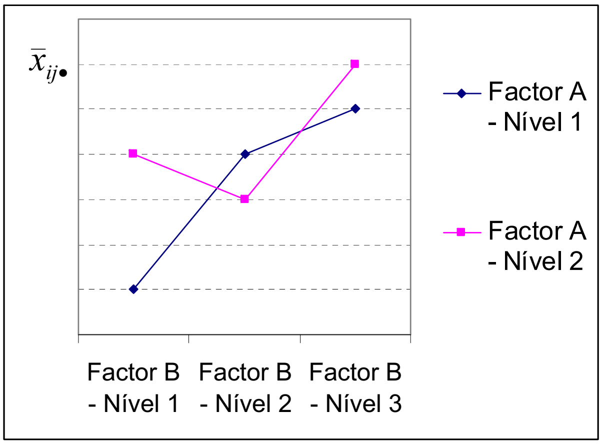
Representando as médias amostrais $\bar{x}_{ij\bullet}$ graficamente, como se ilustra nas figuras seguintes, é possível averiguar se existe ou não uma interacção significativa.



Ausência de interacção significativa: *Segmentos de recta paralelos* – A diferença entre os valores médios para quaisquer dois níveis do Factor A é igual para todos os níveis do factor B e vice-versa.

Neste caso, é possível comparar os níveis do Factor A sem ter de especificar o nível do Factor B envolvido e vice-versa.





Existência de interacção significativa: A diferença entre os valores médios para dois níveis do Factor A pode depender do nível do factor B envolvido e vice-versa.

Neste caso, nem sempre é possível comparar os níveis do Factor A sem ter de especificar o nível do Factor B envolvido e vice-versa.

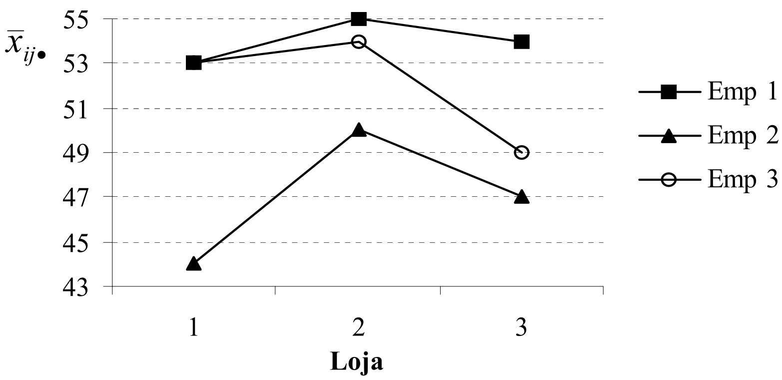
Para o **Exemplo 5**, tem-se, sob H_0^3

$$F = \frac{MS_I}{MS_E} \sim F_{ab(n-1)}^{(a-1)(b-1)}, \qquad \text{com } (a-1)(b-1) = 4 \qquad \text{e} \qquad ab(n-1)=18.$$

Mais:

- Quantil de probabilidade (1-0.05) da distribuição F_{18}^{4}: 2.93;
- R.C.: [2.93, +∞[
- $F_{obs}=3.49 \in$ R.C., logo rejeitamos H_0^3 - a interacção entre o factor Loja e o factor Empregado é significativa, o que conduz à conclusão de que **o desempenho de um vendedor depende da loja onde está a trabalhar**.

Coloca-se então a questão de saber se podemos atribuir algum significado aos testes relativos a cada um dos factores.



A análise do gráfico revela que:

A o empregado 1 tem mais êxito nas vendas do que o empregado 3 e este do que o empregado 2, independentemente da loja;

B a loja 2 apresenta maior volume de vendas do que as outras duas lojas, independentemente do empregado.

Parece então fazer sentido testar a hipótese H_0^1 e a hipótese H_0^2 para avaliar se estas diferenças são ou não significativas.

Sob H_0^1, $F = \dfrac{MS_{Loja}}{MS_E} \sim F_{ab(n-1)}^{a-1}$, com a-1=2 e ab(n-1)=18.

Mais:

- Quantil de probabilidade (1-0.05) da distribuição F_{18}^2: 3.55;
- R.C.: [3.55, +∞[
- $F_{obs}=7.85 \in$ R.C., logo rejeitamos H_0^1 - há evidência para concluir que as três lojas diferem no que diz respeito ao volume médio de vendas semanais.

Sob H_0^2,

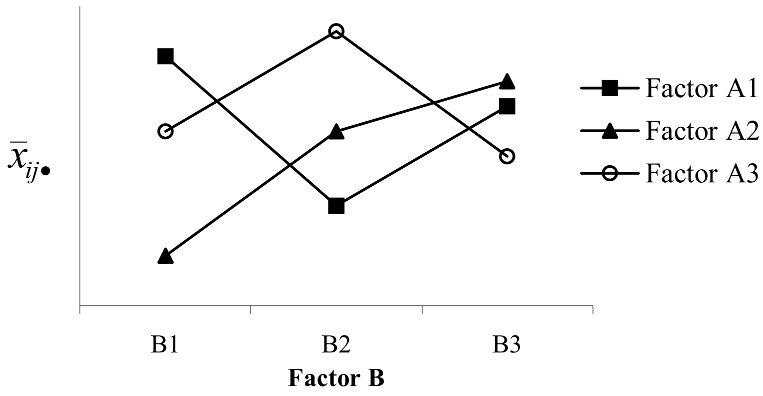
$$F = \frac{MS_{Emp}}{MS_E} \sim F_{ab(n-1)}^{b-1}, \quad \text{com b-1=2 e ab(n-1)=18.}$$

Mais:

- Quantil de probabilidade (1-0.05) da distribuição F_{18}^2: 3.55;
- R.C.: [3.55, +∞[
- $F_{obs}=34.01 \in$ R.C., logo rejeitamos H_0^2 - há evidência de que existem diferenças entre os empregados no que diz respeito ao seu volume médio de vendas.

Podemos concluir que tanto o factor Loja como o factor Empregado exercem uma influência significativa sobre o volume de vendas.

Como já dissemos, a existência de interacção entre os factores pode levar a que os testes relativos aos factores A e B não tenham significado. Na figura seguinte representa-se uma situação deste tipo (compare-a com a Figura anterior).



Comparações Múltiplas

O teste que vamos considerar é, uma vez mais, o **teste de Tuckey**.

Para o Factor A

A hipótese nula ***H*₀: *μ_r* = *μ_s*** (os níveis *r* e *s* do factor A têm médias iguais) é rejeitada se

$$\left|\bar{x}_{r\bullet} - \bar{x}_{s\bullet}\right| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{nb}}$$

onde,

$S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da ***"Studentized Range"*** com (a, ab(n-1)) graus de liberdade

Para o Factor B

A hipótese nula H_0: $\mu_r = \mu_s$ (os níveis r e s do factor B têm médias iguais) é rejeitada se

$$|\bar{x}_{\bullet r} - \bar{x}_{\bullet s}| \geq S_{T(1-\alpha)}\sqrt{\frac{MS_E}{na}}$$

onde,

$S_{T(1-\alpha)}$ é o quantil de probabilidade (1-α) da distribuição da ***"Studentized Range"*** com (b, ab(n-1)) graus de liberdade

Exemplo 5

Vamos apenas avaliar se são significativas as diferenças registadas em **A** e **B** do *slide 85*.

Factor Empregado

Para α=0.05, tem-se $S_{T(1-\alpha)}\sqrt{\frac{MS_E}{na}}=3.64\sqrt{\frac{3.44}{9}}=2.25$

e

$$|\bar{x}_{\bullet 1\bullet} - \bar{x}_{\bullet 2\bullet}|=|54\text{-}47|=7>2.25$$

$$|\bar{x}_{\bullet 1\bullet} - \bar{x}_{\bullet 3\bullet}|=|54\text{-}52|=2<2.25$$

$$|\bar{x}_{\bullet 2\bullet} - \bar{x}_{\bullet 3\bullet}|=|47\text{-}52|=5>2.25$$

Há, portanto, evidência de que o empregado 2 tem um volume médio de vendas diferente dos empregados 1 e 3. A análise do gráfico da Figura deste exemplo revela que essa diferença é favorável aos empregados 1 e 3.

Para o **factor Loja**

tem-se $S_{T(1-\alpha)}\sqrt{\dfrac{MS_E}{nb}}=3.64\sqrt{\dfrac{3.44}{9}}=2.25$

e

$$\left|\bar{x}_{1\bullet\bullet}-\bar{x}_{2\bullet\bullet}\right|=|50-53|=3>2.25$$

$$\left|\bar{x}_{2\bullet\bullet}-\bar{x}_{3\bullet\bullet}\right|=|53-50|=3>2.25.$$

Concluímos portanto que a loja 2 difere significativamente das lojas 1 e 3, no que diz respeito ao volume médio de vendas. A análise do gráfico da figura deste exemplo revela que essa diferença é favorável à loja 2.

Note que, não faz sentido comparar as lojas 1 e 3, pois, devido à interacção, o desempenho destas lojas depende do empregado envolvido (confirme na figura do *slide 88*).

ANÁLISE DE VARIÂNCIA

É importante notar que, além de dois factores, podem ainda ser acrescentados mais factores ao estudo da variação de uma característica (ANOVA com *k* factores). Uma consulta deste assunto pode ser feita em “Applied Statistics and Probability for Engineers”, Montegomery e Runger.