

# CAPÍTULO IV

## Análise de variância

O objectivo principal da análise de variância (*analysis of variance* - ANOVA) é a comparação de mais do que dois grupos no que diz respeito à localização. Para exemplificar, considere-se a situação em que se pretende avaliar a eficácia de um novo medicamento no tratamento de determinada doença através da administração de quatro tratamentos diferentes: o novo medicamento, outros dois já existentes no mercado para tratar a doença em causa e um placebo. Os diferentes tratamentos são usados em indivíduos que sofrem daquela doença distribuídos aleatoriamente por quatro grupos. Será que se pode considerar que os quatro tratamentos têm resultados diferentes? Será que o novo medicamento é mais eficaz do que os já existentes no mercado? A análise de variância procura dar resposta a questões deste tipo através da comparação das localizações dos diferentes grupos. Esta comparação é feita a partir da análise da dispersão presente no conjunto de dados; daí a designação de **análise de variância**.

No exemplo acima descrito, as observações provêm de grupos classificados através de um só factor (a doença em causa); neste caso, fala-se em **análise de variância com um factor** - *one-way ANOVA*. Claro que só é legítimo considerar tal factor como sendo a causa das diferenças entre as médias se se puder garantir a homogeneidade das populações em relação a todos os outros factores que poderiam ser relevantes para a explicação do fenómeno.

Em muitas situações práticas há mais do que um factor a influenciar os resultados das observações. Considere-se o seguinte exemplo<sup>1</sup>: para tentar aumentar a duração de determinado componente para sapatos foram experimentadas cinco matérias-primas em três máquinas de tipos diferentes. Em cada máquina foi produzido um componente utilizando cada uma das matérias-primas e ensaiou-se a duração desses componentes. Coloca-se a questão de saber se a duração dos componentes é influenciada pelo tipo de máquina e pelas matérias-primas. Neste caso estamos perante uma **análise de variância com dois factores** - *two-way ANOVA*.

Por outro lado, diz-se que a análise de variância tem tantos **níveis** ou **efeitos** quantos grupos distintos se considerem. Na maior parte das situações, os grupos são determinados à partida; diz-se então que temos uma análise de variância com **efeitos fixos**. Em alternativa, os grupos podem ser retirados aleatoriamente de entre um conjunto alargado de possibilidades. Nesse caso teremos uma análise de variância com **efeitos aleatórios**.

### 1 Análise de variância com um factor e efeitos fixos

Designamos por  $g$  o número de grupos (efeitos) e por  $n_1, n_2, \dots, n_g$  as dimensões correspondentes. As observações relativas ao  $i$ -ésimo grupo são denotadas por  $x_{i1}, x_{i2}, \dots, x_{in_i}$ ,  $i = 1, 2, \dots, g$ . Admite-se que cada  $x_{ij}$  é uma observação de uma variável aleatória (v.a.)  $X_{ij}$  que verifica:

$$X_{ij} = \underbrace{m + \alpha_i}_{m_i} + \varepsilon_{ij} = \underbrace{m_i}_{m_i} + \varepsilon_{ij}, \quad \text{com } \varepsilon_{ij} \sim N(0, \sigma). \quad (1)$$

Admite-se ainda que as v.a.'s  $\varepsilon_{ij}$  são independentes.

Nestas condições, cada v.a.  $X_{ij}$  segue uma lei normal de média  $m_i = m + \alpha_i$ ,  $i = 1, 2, \dots, g$ , e desvio padrão  $\sigma$ ,

$$X_{ij} \sim N(m_i, \sigma),$$

---

<sup>1</sup>Guimarães, R.C. e Sarsfield Cabral, J.A. (2007) Estatística (2ª edição) McGraw-Hill.

e as v.a.'s  $X_{ij}$  são independentes.

Note-se que todas as v.a.'s envolvidas têm a mesma variância ( $\sigma^2$ ).

Resumindo, os pressupostos exigidos são os seguintes:

- Temos  $g$  grupos de observações independentes, sendo os grupos independentes entre si.
- Cada grupo de observações deve provir de uma população com distribuição normal.
- A variância das  $g$  populações deve ser a mesma (*homogeneidade das variâncias*).

A forma do modelo (1) põe em evidência a seguinte relação:

$$\text{Valor observado} = \text{Média da população subjacente ao } i\text{-ésimo grupo} + \text{Valor residual.}$$

Nas situações práticas descritas por este tipo de modelo interessa fundamentalmente testar se as médias  $m_1, m_2, \dots, m_g$  das  $g$  populações associadas às  $g$  amostras são significativamente diferentes umas das outras.

As hipóteses do teste fundamental da ANOVA podem então ser especificadas do seguinte modo:

$$\begin{aligned} H_0: m_1 = m_2 = \dots = m_g = m & \quad (\text{ou: } \alpha_1 = \alpha_2 = \dots = \alpha_g = 0) \\ H_1: \text{os valores } m_i \text{ não são todos iguais} & \quad (\text{ou: algum } \alpha_i \text{ é diferente de zero}). \end{aligned}$$

A ideia de base para testar estas hipóteses é a seguinte: estima-se a variância  $\sigma^2$  por dois métodos diferentes, um que não depende da veracidade de  $H_0$  e outro que depende da veracidade de  $H_0$ . Depois comparam-se as duas estimativas. Se  $H_0$  é verdadeira, então as duas estimativas devem ser próximas; caso contrário, devem diferir significativamente.

## NOTAÇÕES:

$$N = \sum_{i=1}^g n_i \quad (\text{número total de observações})$$

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad (\text{média das observações do grupo } i)$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{n_i} X_{ij} \quad (\text{média de todas as observações})$$

A variabilidade total das observações é dada pela soma dos quadrados das distâncias de cada observação à média global:

$$SST = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2.$$

Verifica-se a seguinte igualdade:

$$\underbrace{\sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2}_{SST} = \underbrace{\sum_{i=1}^g n_i (\bar{X}_i - \bar{X})^2}_{SSE} + \underbrace{\sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}_{SSD}. \quad (2)$$

*SST*: variabilidade total das observações  $X_{ij}$  em relação à média global  $\bar{X}$ .

*SSE*: variabilidade das observações **entre** grupos - corresponde à soma ponderada das variações das médias de cada grupo,  $\bar{X}_i$ , em torno da média global,  $\bar{X}$  (a ponderação é feita pelo número de observações de cada grupo,  $n_i$ .)

*SSD*: variabilidade das observações **dentro** dos grupos - corresponde à soma das variações das observações  $X_{ij}$  dentro de cada um dos diferentes grupos (para cada grupo  $i$ , a variação das observações é calculada relativamente à média desse grupo,  $\bar{X}_i$ ).

Definimos ainda:

$$MSE = \frac{SSE}{g-1} : \text{média da soma dos quadrados entre grupos;}$$

$$MSD = \frac{SSD}{N-g} : \text{média da soma dos quadrados dentro dos grupos.}$$

Os valores de *MSD* e *MSE* são as duas estimativas de  $\sigma^2$  anteriormente referidas (sendo *MSE* aquela que depende da veracidade de  $H_0$ ). Assim, quando a hipótese  $H_0$  é verdadeira, estes valores devem ser próximos e, conseqüentemente, a razão  $\frac{MSE}{MSD}$  terá um valor próximo de 1. Se  $H_0$  não for verdadeira, então o valor de *MSE* será significativamente superior ao de *MSD*. Assim, a hipótese  $H_0$  é rejeitada para valores elevados de  $\frac{MSE}{MSD}$ .

$F = \frac{MSE}{MSD}$  é precisamente a **estatística de teste** usada para efectuar o teste de

$$H_0 : m_1 = m_2 = \dots = m_g = m \quad \text{contra} \quad H_1 : \text{nem todos os } m_i \text{ são iguais.}$$

Sob a validade de  $H_0$ , tem-se

$$F \sim F(g-1, N-g),$$

onde  $F(g-1, N-g)$  representa a *distribuição de Fisher com  $g-1$  e  $N-g$  graus de liberdade*.

Como foi referido acima, a hipótese  $H_0$  é rejeitada para valores elevados de  $F$ , pelo que

$$p\text{-valor} = P(F \geq f_0),$$

onde  $f_0$  representa o valor observado de  $\frac{MSE}{MSD}$ .

**Exemplo 1<sup>2</sup>**: Um determinado departamento governamental está preocupado com os aumentos dos custos verificados no decurso de projectos de investigação e desenvolvimento encomendados aos institutos  $A$ ,  $B$ ,  $C$  e  $D$ . Assim, decidiu analisar os custos associados a diferentes projectos, calculando para cada um deles a razão entre o custo final incorrido e o custo inicialmente previsto. Os resultados apresentam-se na tabela seguinte:

<sup>2</sup>Guimarães, R.C. e Sarsfield Cabral, J.A. (2007) Estatística (2ª edição) McGraw-Hill.

Instituto						
A	1.0	0.8	1.9	1.1	2.7	
B	1.7	2.5	3.0	2.2	3.7	1.9
C	1.0	1.3	3.2	1.4	1.3	2.0
D	3.8	2.8	1.9	3.0	2.5	

A questão que se coloca é a de saber se os quatro institutos têm um comportamento distinto em relação ao agravamento de custos.

A resposta a esta questão passa por efectuar uma análise de variância com um factor (instituto) e efeitos fixos.

Para efectuar a análise de variância no SPSS, as 22 observações disponíveis têm que estar todas numa mesma coluna no ficheiro de dados. Assim, é necessário criar outra variável (numérica) que identifique o grupo (instituto) a que pertence cada uma das observações.

Podemos começar por observar a localização relativa dos quatro grupos através da construção de boxplots paralelos, como se mostra na figura 1.

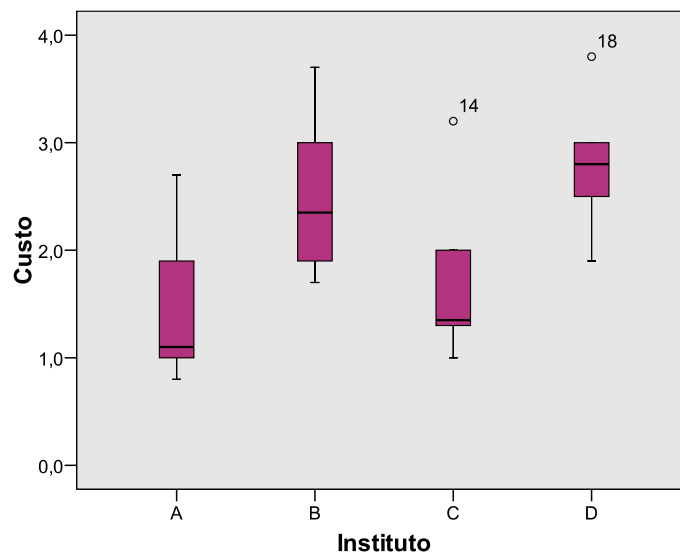


Figura 1: Boxplots das amostras do exemplo 1.

Vamos agora analisar se cada uma das quatro amostras pode ser considerada como proveniente de uma população normal. O teste adequado é o de Shapiro-Wilk (dimensões das amostras  $< 20$ ). A análise dos p-valores presentes no quadro<sup>3</sup> da figura 2, permite-nos aceitar as hipóteses correspondentes ao nível de significância 0.05.

<sup>3</sup>Um quadro deste tipo e boxplots paralelos podem ser obtidos simultaneamente através de *Analyze* → *Descriptive Statistics* → *Explore*, colocando a variável “Custo” em *Dependent List* e a variável “Instituto” em *Factor List* e seleccionando, em *Plots*, *Normality plots with tests*.

### Tests of Normality

Instituto		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
A	Custo	,294	5	,184	,876	5	,292
B	Custo	,167	6	,200*	,945	6	,703
C	Custo	,312	6	,069	,810	6	,072
D	Custo	,187	5	,200*	,987	5	,967

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figura 2: Testes de normalidade para as amostras do exemplo 1.

Os resultados da análise de variância conseguem-se através de

*Analyze* → *Compare Means* → *One way ANOVA*.

Em *Options* devemos seleccionar *Homogeneity of Variances* que fornece o p-valor de um teste da hipótese  $H_0$ : “as variâncias das  $g$  populações são iguais” contra a hipótese  $H_1$ : “há pelo menos duas variâncias diferentes”.

A aceitação da hipótese  $H_0$  permite validar o pressuposto da homogeneidade de variâncias.

Na figura seguinte apresenta-se o output relativo ao exemplo 1.

### Test of Homogeneity of Variances

Custo

Levene Statistic	df1	df2	Sig.
,133	3	18	,939

### ANOVA

Custo

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6,159	3	2,053	3,533	,036
Within Groups	10,460	18	,581		
Total	16,619	21			

Figura 3: Teste da homogeneidade de variâncias e tabela de análise de variância.

O p-valor da primeira tabela (0.939) permite-nos considerar que as variâncias das populações subjacentes aos quatro grupos em análise são iguais.

O p-valor da segunda tabela (0.036) leva-nos à rejeição da hipótese da igualdade das médias (ao nível de significância 0.05), concluindo-se que os institutos têm comportamentos distintos no que diz respeito ao agravamento dos custos.

## 1.1 Comparações múltiplas

Quando se rejeita a hipótese nula da igualdade das médias, não se tem informação sobre qual ou quais dos grupos são responsáveis pela diferença. Uma das possibilidades para efectuar comparações múltiplas consiste em comparar todos os pares de médias,  $m_k$  e  $m_l$ , para  $k \neq l$ .

**Problema:** Fixado o nível de significância do teste correspondente a cada comparação, o nível de significância global de um conjunto de comparações tende a ser tanto mais elevado quanto maior é o número de comparações, o que não é desejável. Efectivamente, o facto do nível de significância global de um conjunto de comparações ser elevado significa que é elevada a probabilidade de se rejeitar indevidamente **pelo menos uma** das hipóteses nulas do conjunto<sup>4</sup>.

Existem métodos que procuram tomar em consideração este problema. Vamos abordar alguns deles.

### 1.1.1 Definição conjunta dos intervalos de confiança pelo método de Tukey

O método de Tukey consiste na construção de intervalos de confiança para todos os pares de médias de tal forma que o conjunto de todos os intervalos tenha um determinado grau de confiança  $\gamma$  (por exemplo, se  $\gamma = 0.95$ , temos 95% de confiança de que os intervalos obtidos contêm, **todos**, as respectivas diferenças  $m_k - m_l$ ).

O método de construção destes intervalos depende do facto dos grupos terem ou não a mesma dimensão.

**Amostras equilibradas:** Todos os grupos têm a mesma dimensão  $n$ .

Os intervalos de confiança para  $m_k - m_l$ ,  $k \neq l$ , com grau de confiança  $\gamma$ , são dados por

$$(\bar{x}_k - \bar{x}_l) \pm q_{1-\gamma}(g, N - g) \sqrt{\frac{MSD}{n}}, \quad (3)$$

onde  $q_{1-\gamma}(g, N - g)$  é um valor tabelado especialmente para os intervalos de confiança obtidos pelo método de Tukey<sup>5</sup> (note-se que  $N = ng$ , no caso particular de que aqui se trata).

**Amostras pouco desequilibradas:** Os grupos têm diferentes dimensões, mas exige-se que *dimensão máxima*  $\leq 2$  *dimensão mínima*<sup>6</sup>.

Os intervalos de confiança têm uma expressão do tipo (3), substituindo-se  $n$  por  $\frac{g}{\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_g}}$ : média harmónica (*harmonic mean*) dos valores  $n_1, n_2, \dots, n_g$ .

### 1.1.2 Método de Bonferroni

O método de Bonferroni consiste em efectuar cada um dos testes individuais ( $H_0: m_k = m_l$ ,  $H_1: m_k \neq m_l$ ) com um nível de significância muito reduzido de modo que o nível global seja o desejado. Assim, considera-se para cada uma das  $r$  comparações individuais um nível de significância  $\alpha_r = \frac{\alpha}{r}$  por forma a garantir que o nível total seja, no máximo,  $\alpha$ .

Este método funciona bem desde que o número de comparações a efectuar não seja demasiado elevado.

<sup>4</sup>O nível de significância de um teste corresponde à probabilidade de se rejeitar indevidamente a hipótese  $H_0$ .

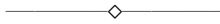
<sup>5</sup>Em rigor,  $z = q_{1-\gamma}(g, N - g)$  é tal que  $P(R \leq z) = \frac{1+\gamma}{2}$ , sendo  $R$  uma v.a. cuja distribuição é designada por *Studentized Range distribution*.

<sup>6</sup>Caso contrário, usar o **método de Scheffé**.

### 1.1.3 Teste de Dunnett

Por vezes, o objectivo de uma análise de comparações múltiplas é o de comparar um grupo particular (grupo de controlo) com cada um dos restantes grupos. Neste caso, temos  $g - 1$  comparações a efectuar. Dunnett propôs um procedimento para este efeito, semelhante ao de Tukey.

Quando se pretende realizar o teste de Dunnett no SPSS, o grupo de controlo deve ser o primeiro ou o último na coluna do ficheiro de dados que contém as observações de todos os grupos.



Para todos os métodos apresentados, o SPSS fornece os intervalos de confiança e também os testes correspondentes. Se usarmos um nível de significância  $\alpha$  para os testes, então os intervalos de confiança terão grau de confiança  $\gamma = 1 - \alpha$ . O nível  $\alpha$  é especificado pelo utilizador.

O trajecto para obter os resultados no SPSS é

*Analyze → Compare Means → One way ANOVA → Post Hoc.*

**Continuação do exemplo 1** Concluiu-se que as médias dos quatro institutos não são todas iguais. Vamos usar o método de Tukey para tentar descobrir quais as médias que são de facto diferentes. Usamos o nível 0.05.

#### Multiple Comparisons

Dependent Variable: Custo  
Tukey HSD

(I) Instituto	(J) Instituto	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	-1,0000	,4616	,171	-2,305	,305
	C	-,2000	,4616	,972	-1,505	1,105
	D	<b>-1,3000</b>	,4821	<b>,065</b>	-2,663	,063
B	A	1,0000	,4616	,171	-,305	2,305
	C	,8000	,4401	,298	-,444	2,044
	D	-,3000	,4616	,914	-1,605	1,005
C	A	,2000	,4616	,972	-1,105	1,505
	B	-,8000	,4401	,298	-2,044	,444
	D	-1,1000	,4616	,116	-2,405	,205
D	A	<b>1,3000</b>	,4821	<b>,065</b>	-,063	2,663
	B	,3000	,4616	,914	-1,005	1,605
	C	1,1000	,4616	,116	-,205	2,405

Figura 4: Comparações múltiplas (método de Tukey).

Neste quadro, nenhum dos p-valores é inferior a 0.05. *Grosso modo*, não há diferenças entre os diferentes pares de médias. No entanto, começámos por rejeitar a igualdade das quatro médias!...

Note-se que o p-valor 0.065, correspondente à diferença  $m_A - m_D$  (ou  $m_D - m_A$ ), é muito próximo de 0.05: é melhor não aceitarmos facilmente a hipótese  $m_A = m_D$ !

Vejamos que conclusões nos são sugeridas pelo método de Bonferroni.

### Multiple Comparisons

Dependent Variable: Custo

Bonferroni

(I) Instituto	(J) Instituto	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	-1,0000	,4616	,264	-2,368	,368
	C	-,2000	,4616	1,000	-1,568	1,168
	D	<b>-1,3000</b>	,4821	<b>,089</b>	-2,728	,128
B	A	1,0000	,4616	,264	-,368	2,368
	C	,8000	,4401	,515	-,504	2,104
	D	-,3000	,4616	1,000	-1,668	1,068
C	A	,2000	,4616	1,000	-1,168	1,568
	B	-,8000	,4401	,515	-2,104	,504
	D	-1,1000	,4616	,170	-2,468	,268
D	A	<b>1,3000</b>	,4821	<b>,089</b>	-,128	2,728
	B	,3000	,4616	1,000	-1,068	1,668
	C	1,1000	,4616	,170	-,268	2,468

Figura 5: Comparações múltiplas (método de Bonferroni).

As mesmas!!...

Na figura 4, podemos observar que o p-valor correspondente a  $m_C - m_D$  também é relativamente baixo (0.116). Vamos então tentar obter mais alguns resultados com o teste de Dunnett, tomando o grupo  $D$  como controlo.

Observemos que  $m_A$  e  $m_C$  tendem a ser inferiores a  $m_D$ . Vamos então optar por fazer testes unilaterais que tenham em conta esta tendência. Assim, depois de seleccionar *Dunnett*, seleccionamos  $< Control$  em *Test*. Na figura seguinte podemos observar o output correspondente.

### Multiple Comparisons

Dependent Variable: Custo

Dunnett t (<control)<sup>a</sup>

(I) Instituto	(J) Instituto	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval
					Upper Bound
A	D	-1,3000*	,4821	<b>,019</b>	-,240
B	D	-,3000	,4616	,470	,715
C	D	-1,1000*	,4616	<b>,035</b>	-,085

\*. The mean difference is significant at the 0.05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Figura 6: Comparações múltiplas (teste de Dunnett).

Podemos então considerar que tanto  $m_A$  como  $m_C$  são inferiores a  $m_D$  e que  $m_B$  e  $m_D$  não diferem significativamente. Além disso, nas tabelas das figuras 4 e 5 podemos observar p-valores muito elevados associados às comparações entre  $m_A$  e  $m_C$ , o que nos leva a aceitar a igualdade destas duas médias.

Em resumo, conclui-se que  $m_A = m_C < m_D = m_B$ .



**NOTA:** Quando usamos o método de Tukey no SPSS, além da tabela das comparações múltiplas, é produzida uma tabela de grupos homogêneos. Trata-se de uma tabela que subdivide o conjunto dos  $g$  grupos de observações em subconjuntos dentro dos quais podemos considerar que as médias não apresentam diferenças significativas (ao nível de significância adoptado para as comparações múltiplas).

No caso do exemplo aqui tratado, o método de Tukey não evidenciou diferenças de médias ao nível 0.05. Assim, a tabela de grupos homogêneos correspondente apresenta apenas o conjunto:  $\{A, B, C, D\}$ .

No entanto, se repetirmos o procedimento com o nível 0.07, a tabela de grupos homogêneos apresenta dois subconjuntos:  $\{A, B, C\}$  e  $\{B, C, D\}$ .

## 2 Análise de variância com um factor e efeitos aleatórios

No exemplo 1, que serviu de introdução ao modelo de efeitos fixos, o objectivo da análise foi o de verificar se existiam diferenças entre os quatro institutos  $A$ ,  $B$ ,  $C$  e  $D$  (e só esses!). Considere-se agora a situação descrita no exemplo seguinte:

**Exemplo 2<sup>7</sup>:** Um determinado departamento governamental tem projectos de investigação e desenvolvimento encomendados a um número elevado de institutos. Por razões ligadas ao controlo de custos, está interessado em saber se, na generalidade dos casos, os institutos a quem adjudica projectos têm ou não comportamentos idênticos no tocante ao agravamento dos custos inicialmente orçamentados. Com este objectivo, o departamento decide escolher uma amostra aleatória de quatro institutos e, em relação a cada um deles, analisar o agravamento de custos relativos a um conjunto de projectos seleccionados ao acaso.

Admita-se que os institutos e os resultados associados às amostras de projectos são os que foram considerados no exemplo 1.

A diferença essencial entre as situações descritas nos dois exemplos é a de que, no segundo, os quatro institutos constituem uma *amostra aleatória* dos institutos aos quais são encomendados projectos de investigação e desenvolvimento. Assim, o valor esperado das observações, por exemplo, do instituto  $A$ , é um valor seleccionado aleatoriamente entre vários valores possíveis do mesmo tipo (valores esperados das observações dos institutos  $E$ ,  $F$ , etc.). Nestas circunstâncias, os parâmetros  $m_i$  e  $\alpha_i$  que integravam o modelo (1) passam a ser *variáveis aleatórias*. Passamos então a usar as notações  $M_i$  e  $A_i$ , respectivamente. O modelo de análise de variância com um factor e efeitos aleatórios é então descrito da seguinte forma:

$$X_{ij} = m + A_i + \varepsilon_{ij} = M_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, g, \quad j = 1, 2, \dots, n_i, \quad (4)$$

admitindo-se que as v.a.'s  $A_i$  são independentes e têm todas a mesma distribuição  $N(0, \sigma_A)$  e que as v.a.'s  $\varepsilon_{ij}$  são independentes e têm todas a mesma distribuição  $N(0, \sigma)$ . Admite-se ainda que qualquer das variáveis do conjunto  $\{A_i, \quad i = 1, 2, \dots, g\}$  é independente de qualquer das variáveis do conjunto  $\{\varepsilon_{ij}, \quad i = 1, 2, \dots, g, \quad j = 1, 2, \dots, n_i\}$ .

Neste modelo, o parâmetro  $m$  representa o valor esperado do conjunto das observações (sem especificação do grupo a que pertencem).

---

<sup>7</sup>Guimarães, R.C. e Sarsfield Cabral, J.A. (2007) Estatística (2ª edição) McGraw-Hill.

Num modelo de efeitos aleatórios, a forma mais apropriada de testar a igualdade das médias dos vários grupos é através das hipóteses

$$H_0 : \sigma_A^2 = 0 \quad \text{contra} \quad H_1 : \sigma_A^2 > 0. \quad (5)$$

Claro que a hipótese alternativa só pode ser  $\sigma_A^2 > 0$ , uma vez que a variância não pode ser negativa.

Note-se que, se os efeitos tiverem variância nula, então a média dos grupos não poderá variar: é esta a base da análise de variância com efeitos aleatórios.

Tal como no modelo de efeitos fixos, se  $H_0$  é verdadeira, os valores de  $MSD$  e  $MSE$  são estimativas de  $\sigma^2$ . Assim, continuamos a usar a razão  $\frac{MSE}{MSD}$  para testar as hipóteses (5).

Quando  $H_0$  é rejeitada, faz sentido estimar a variância dos efeitos,  $\sigma_A^2$ . A estimativa correspondente é dada por

$$\hat{\sigma}_A^2 = \frac{MSE - MSD}{h},$$

com

$$h = \frac{1}{N(g-1)} \left( N^2 - \sum_{i=1}^g n_i^2 \right).$$

Num modelo de efeitos aleatórios não analisamos as comparações múltiplas, devido à natureza aleatória dos grupos.

No SPSS, o procedimento para efectuar uma análise de variância com efeitos aleatórios é semelhante ao procedimento usado para o caso de efeitos fixos, mas devemos seleccionar (em *Options*) a opção

*Fixed and random effects.*

Esta opção fornece, além da tabela ANOVA, uma tabela (*Descriptives*) onde aparece uma linha relativa a *Random Effects* que contém a estimativa de  $\sigma_A^2$  acima apresentada (*Between-Component Variance*). Claro que esta estimativa só deve ser tida em conta quando rejeitarmos a hipótese  $H_0 : \sigma_A^2 = 0$ .

A tabela *Descriptives* relativa ao exemplo apresentado nesta secção figura a seguir.

<b>Descriptives</b>						
Custo						
		Std. Deviation	Std. Error	95% Confidence Interval for Mean		<b>Between-Component Variance</b>
				Lower Bound	Upper Bound	
Model	Fixed Effects	,7623	,1625	1,781	2,464	
	Random Effects		,3067	1,147	3,099	<b>,2683</b>

Figura 7: Tabela com informação relativa a um modelo ANOVA de efeitos aleatórios.

A correspondente tabela ANOVA é a mesma do exemplo 1 (figura 3), a partir da qual decidimos rejeitar  $H_0$ . Assim, devemos ter em conta a estimativa para a variância dos efeitos:  $\hat{\sigma}_A^2 = 0.2683$ .

### 3 ANOVA não paramétrica: Teste de Kruskal-Wallis

Em muitas situações, não podemos usar os modelos de análise de variância descritos nas secções anteriores porque os dados contrariam fortemente a hipótese de normalidade ou da igualdade de variâncias<sup>8</sup>.

Recorremos então a técnicas não paramétricas para fazer a comparação das várias localizações, sendo o teste de Kruskal-Wallis uma das mais utilizadas. Este teste destina-se a verificar se há diferenças na localização das populações (com distribuições contínuas) subjacentes aos  $g$  grupos.

As hipóteses em teste são:

$H_0$ : as  $g$  populações têm a mesma localização,

$H_1$ : pelo menos duas das  $g$  populações não têm a mesma localização.

O procedimento a aplicar para efectuar o teste de Kruskal-Wallis é semelhante ao do teste de Mann-Whitney: ordenam-se as  $N$  observações em conjunto e atribuem-se-lhes *ranks* (posições:  $1, 2, \dots, N$ ).

Quando há empates (observações repetidas) atribui-se o *rank* médio às observações empatadas.

A ideia base do teste é a de que, se  $H_0$  for verdadeira, os *ranks* correspondentes aos vários grupos estarão misturados de forma aleatória; caso contrário, deverão existir grupos com predominância de *ranks* reduzidos e outros grupos com predominância de *ranks* elevados.

A estatística de teste de Kruskal-Wallis, baseia-se nas posições médias das observações de cada grupo (*mean ranks*), avaliando o grau de disparidade entre tais posições médias.

Para efectuar o teste de Kruskal-Wallis no SPSS, seguimos o trajecto

*Analyse* → *Nonparametric Tests* → *K Independent Samples*.

---

<sup>8</sup>Na prática, a homogeneidade das variâncias só se torna importante quando as dimensões dos grupos são muito diferentes, mais precisamente, quando  $(n_i)_{max} \geq 2(n_i)_{min}$ . Quando as amostras não são fortemente desequilibradas, o efeito da heterogeneidade das variâncias, mesmo se acentuada, é pouco significativo.

**Exemplo 3** Determinada empresa farmacêutica lançou um concurso para recrutar novos trabalhadores para três dos seus laboratórios: controlo da qualidade (lab. 1), microbiologia (lab. 2) e investigação e desenvolvimento (lab. 3). Concorreram 54 pessoas: 20 para o lab. 1, 18 para o lab. 2 e 16 para o lab. 3.

Todas estas pessoas realizaram o mesmo teste sobre temas de Química, tendo-lhes sido atribuída uma classificação entre 0 e 20 valores. Na figura 8 apresenta-se um resumo dos resultados obtidos em cada um dos grupos.

Nota		
1	N	20
	Mean	10,225
	Median	9,800
	Variance	1,668
2	N	18
	Mean	13,022
	Median	12,500
	Variance	3,878
3	N	16
	Mean	12,806
	Median	12,550
	Variance	,946

Figura 8: Resumo da amostra das classificações do exemplo 3.

Será que os três grupos diferem significativamente no nível de conhecimentos relativamente aos quais foram testados?

A resposta a esta pergunta requer uma análise de variância.

Em primeiro lugar, vejamos se as três amostras podem ser consideradas como provenientes de leis normais.

Tests of Normality							
Nota	Grupo	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Nota	1	,192	20	,052	,805	20	,001
	2	,252	18	,004	,764	18	,000
	3	,162	16	,200*	,907	16	,105

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figura 9: Testes de normalidade relativos ao exemplo 3.

Como se pode observar no quadro da figura 9, os p-valores sugerem a rejeição das hipóteses de normalidade, pelo menos no que respeita aos grupos 1 e 2. Assim, somos conduzidos a uma ANOVA não paramétrica.

Os resultados do teste de Kruskal-Wallis apresentam-se na figura 10.

**Ranks**

	Grupo	N	Mean Rank
Nota	1	20	12,43
	2	18	35,64
	3	16	37,19
	Total	54	

**Test Statistics<sup>a,b</sup>**

	Nota
Chi-Square	29,284
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: Grupo

Figura 10: Teste Kruskal-Wallis relativo ao exemplo 3.

Perante o p-valor apresentado no segundo quadro (0.000), rejeitamos a hipótese de que os concorrentes têm o mesmo nível de conhecimentos. Por outro lado, a observação dos valores *Mean Rank* do primeiro quadro sugere que o grupo 1 tem um nível de conhecimentos mais baixo e que o nível de conhecimentos dos outros dois grupos parece não diferir muito.