CAPÍTULO
13

Análise de Variância

O Capítulo 12 explica como comparar médias de duas populações, com base em amostras dessas populações. Mas às vezes é preciso comparar médias de *mais de duas populações*. Por exemplo, para verificar se pessoas com diferentes níveis de renda, isto é, alto, médio e baixo têm, em média, o mesmo peso corporal, é preciso comparar médias de três populações.

Outras vezes, é preciso comparar várias situações experimentais. Por exemplo, se um pesquisador separa, ao acaso, um conjunto de pacientes em 4 grupos e administra uma droga diferente a cada grupo, terá que comparar médias de quatro "populações".

Para comparar médias de mais de duas populações aplica-se o teste F, na forma descrita nesse Capítulo, desde que a variável em estudo tenha distribuição normal ou aproximadamente normal. Mas antes de mostrar como se faz esse teste, convém apresentar um exemplo.

Imagine que 4 amostras casuais simples, todas com cinco elementos mas cada uma proveniente de uma população, conduziram aos dados apresentados na Tabela 13.1. As médias dessas amostras estão na última linha dessa tabela. Será que as diferenças das médias das amostras são suficientemente grandes para que se possa afirmar que as médias das populações são diferentes? Para responder a esta pergunta, é preciso um teste estatístico.

Tabela 13.1

Dados de 4 amostras e respectivas médias

Amostras			
A	B	C	D
11	8	5	4
8	5	7	4
5	2	3	2
8	5	3	0
8	5	7	0
8	5	5	2

13.1 - Análise de Variância para Experimentos ao Acaso

Se a variável em estudo tem distribuição normal ou aproximadamente normal, para comparar mais de duas médias aplica-se o teste F. Primeiro, é preciso estudar as *causas de variação*. Por que os dados variam? Uma explicação é o fato de as amostras provirem de populações diferentes. Outra explicação é o acaso, porque mesmo dados provenientes da mesma população variam.

O teste F é feito através de uma *análise de variância*, que separa a variabilidade devido aos "tratamentos" (no exemplo, devido às amostras terem provindo de populações diferentes) da variabilidade residual, isto é, devido ao acaso. Para aplicar o teste F é preciso fazer uma série de cálculos, que exigem conhecimento da notação.

A Tabela 13.2 apresenta os dados de k tratamentos, cada um com r repetições (no exemplo, denominam-se repetições os elementos da mesma amostra). A soma das r repetições de um mesmo tratamento constitui o *total* desse tratamento. O *total geral* é dado pela soma dos k totais de tratamentos.

Para fazer a análise de variância é preciso calcular as seguintes quantidades:

a) os graus de liberdade:
de tratamentos: $k - 1$
do total: $n - 1$
de resíduo: $(n - 1) - (k - 1) = n - k$

b) o valor C, dado pelo total geral elevado ao quadrado e dividido pelo número de dados. O valor C é chamado *correção*.

$$C = \frac{(\sum x)^2}{n}$$

c) a soma de quadrados total:

$$\text{SQT} = \sum x^2 - C$$

d) a soma de quadrados de tratamentos:

$$\text{SQTr} = \frac{\sum T^2}{r} - C$$

e) a soma de quadrados de resíduo:

$$\text{SQR} = \text{SQT} - \text{SQTr}$$

f) o quadrado médio de tratamentos:

$$\text{QMTr} = \frac{\text{SQTr}}{k-1}$$

g) o quadrado médio de resíduo:

$$\text{QMR} = \frac{\text{SQR}}{n-k}$$

h) o valor de F

$$F = \frac{\text{QMTr}}{\text{QMR}}$$

Tabela 13.2

Notação para a análise de variância

	Tratamento					
	1	*2*	*3*	...	*k*	Total
	x_{11}	x_{21}	x_{31}		x_{k1}	
	x_{12}	x_{22}	x_{32}		x_{k2}	
	.	.	.		.	
	.	.	.		.	
	.	.	.		.	
	x_{1r}	x_{2r}	x_{3r}		x_{kr}	
Total	T_1	T_2	T_3		T_k	$\Sigma T = \Sigma x$
Nº de repetições	r	r	r		r	$n = kr$
Média	$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$		$\overline{x}_k$	

Em seguida, é preciso comparar o valor calculado de F com o valor dado na tabela de F, ao nível de significância estabelecido e com $(k - 1)$ graus de liberdade no numerador e $(n - k)$ graus de liberdade no denominador. Toda vez que o valor calculado de F for igual ou maior do que o dado na tabela de F conclui-se, ao nível de significância estabelecido, que as médias de tratamentos não são iguais.

Para entender como se procura o valor de F na tabela, observe a Figura 13.1, que reproduz parte da Tabela A.4 dada neste livro, em Apêndice. Está sombreado o valor de F ao nível de significância de 5%, com 3 graus de liberdade para tratamentos (numerador) e 8 graus de liberdade para resíduo (denominador).

Figura 13.1 *Valor de F ao nível de significância de 5%, com 3 e 8 graus de liberdade*

	1	2	3	4
1	161	200	216	225
2	18,5	19,0	19,2	19,2
3	10,1	9,55	9,28	9,12
4	7,71	6,94	6,59	6,39
5	6,61	5,79	5,41	5,19
6	5,99	5,14	4,76	4,53
7	5,59	4,74	4,35	4,12
8	5,32	4,46	4,07	3,84
9	5,12	4,26	3,86	3,63

Um exemplo ajuda a entender como se aplica o teste F para a comparação de médias. Veja os dados apresentados na Tabela 13.1. Para fazer a análise de variância é preciso calcular:

a) os graus de liberdade:
de tratamentos: $k - 1 = 4 - 1 = 3$
do total: $n - 1 = 20 - 1 = 19$
de resíduo: $n - k = 20 - 4 = 16$

b) o valor de C:

$$C = \frac{(11 + 8 + \ldots + 0)^2}{20} = \frac{100^2}{20} = 500$$

c) a soma de quadrados total:

$$SQT = 11^2 + 8^2 + ... + 0^2 - 500$$
$$= 658 - 500$$
$$= 158$$

d) a soma de quadrados de tratamentos:

$$SQTr = \frac{40^2 + 25^2 + 25^2 + 10^2}{5} - 500$$
$$= 590 - 500$$
$$= 90$$

e) a soma de quadrados de resíduo:

$$SQR = 158 - 90$$
$$= 68$$

f) o quadrado médio de tratamentos:

$$QMTr = \frac{90}{3} = 30$$

g) o quadrado médio do resíduo:

$$QMR = \frac{68}{16} = 4{,}25$$

h) o valor de F:

$$F = \frac{30}{4{,}25} = 7{,}06$$

As quantidades calculadas são apresentadas numa *tabela de análise de variância*. Veja a Tabela 13.3.

Tabela 13.3

Análise de variância dos dados da Tabela 13.1

Causas de variação	GL	SQ	QM	F
Tratamentos	3	90	30	7,06
Resíduo	16	68	4,25	
Total	19	158		

Ao nível de significância de 5%, o valor de F na Tabela A.4 do Apêndice, com 3 e 16 graus de liberdade, é 3,24. Como o valor obtido é maior do que 3,24, conclui-se que as médias não são iguais, ao nível de significância de 5%.

13.2 - Teste de Tukey para Comparação de Médias

Uma análise de variância permite estabelecer se as médias das populações em estudo são, ou não são, estatisticamente iguais. No entanto, esse tipo de análise não permite detectar quais são as médias estatisticamente diferentes das demais. Por exemplo, a análise de variância apresentada na Tabela 13.3 mostrou que as médias das populações não são iguais, mas não permite concluir qual é, ou quais são, as médias diferentes das demais.

O teste de Tukey permite estabelecer a *diferença mínima significante*, ou seja, a menor diferença de médias de amostras que deve ser tomada como estatisticamente significante, em determinado nível. Essa diferença (d.m.s.) é dada por:

$$\text{d. m. s.} = q\sqrt{\frac{\text{QMR}}{r}}$$

onde q é um valor dado em tabela, QMR é o quadrado médio do resíduo da análise de variância e r é o número de repetições de cada tratamento.

A tabela de q é dada neste livro, em Apêndice. Para entender como se usa essa tabela, primeiro observe a Figura 13.2. O valor sombreado seria utilizado para comparar as médias de 3 tratamentos, quando o número de graus de liberdade do resíduo da análise de variância é igual a 6, ao nível de significância de 5%.

Figura 13.2 *Valor de q para α = 5%, 3 tratamentos e 6 graus de liberdade no resíduo*

Nº de graus de lib. do resíduo	2	3	4	5
1	18,0	27,0	32,8	37,1
2	6,08	8,33	9,80	10,9
3	4,50	5,91	6,82	7,50
4	3,93	5,04	5,76	6,29
5	3,64	4,60	5,22	5,67
6	3,46	4,34	4,90	5,30
7	3,34	4,16	4,68	5,06

Considere agora os dados da Tabela 13.1. A análise de variância apresentada na Tabela 13.3 mostra um valor de F significante ao nível de 5%. Então as médias de A, B, C e D não são estatisticamente iguais. Mas qual é, ou quais são, as médias diferentes entre si?

A pergunta pode ser respondida com a aplicação do teste de Tukey. Ao nível de significância de 5%, o valor de q para comparar 4 tratamentos (A, B, C e D), com 16 graus de liberdade no resíduo (veja a Tabela 13.3), é 4,05. Como QMR = 4,25 e r = 5, segue-se que:

$$\text{d.m.s.} = 4{,}05\sqrt{\frac{4{,}25}{5}}$$

$$= 3{,}73$$

De acordo com o teste de Tukey, duas médias são estatisticamente diferentes toda vez que o valor absoluto da diferença entre elas for igual ou superior ao valor da d.m.s. No caso do exemplo, o valor da d.m.s. é 3,73 e os valores absolutos das diferenças entre as médias estão apresentados na Tabela 13.4. É fácil ver que a diferença entre as médias A e D é maior do que a d.m.s. Então, ao nível de 5%, a média de A é significativamente maior do que a média de D.

Tabela 13.4
Valores absolutos das diferenças entre as médias dos tratamentos A, B, C e D

Pares de médias	Valor absoluto da diferença
A e B	8 – 5 = 3
A e C	8 – 5 = 3
A e D	8 – 2 = 6
B e C	5 – 5 = 0
B e D	5 – 2 = 3
C e D	5 – 2 = 3

13.3 - Análise de Variância com Número Diferente de Repetições

Muitas vezes o pesquisador dispõe de diversas amostras, cada uma proveniente de uma população, mas essas amostras não têm todas o mesmo tamanho. Mesmo assim, é possível conduzir a análise de variância. Aliás, todos os cálculos, com exceção da soma de quadrados de tratamentos, são feitos na forma já apresentada na Seção 13.2.

Para entender como se calcula a soma de quadrados de tratamentos — quando os tratamentos não têm o mesmo número de repetições — primeiro observe a Tabela 13.5.

Tabela 13.5

Notação para a análise de variância

	Tratamento					
	1	2	3	...	k	Total
	x_{11}	x_{21}	x_{31}		x_{k1}	
	x_{12}	x_{22}	x_{32}		x_{k2}	
	.	.	.		.	
	.	.	.		.	
	.	.	.		.	
	x_{1r_1}	x_{2r_2}	x_{3r_3}		x_{kr_k}	
Total	T_1	T_2	T_3		T_k	$\Sigma T = \Sigma x$
Nº de repetições	r_1	r_2	r_3		r_k	$n = \Sigma r$
Média	$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$		$\overline{x}_k$	

A soma de quadrados de tratamentos é dada pela fórmula:

$$\text{SQTr} = \frac{T_1^2}{r_1} + \frac{T_2^2}{r_2} + \ldots + \frac{T_k^2}{r_k} - C,$$

onde C é a correção, definida na Seção 13.1.

É mais fácil entender a aplicação de fórmulas através de um exemplo. Veja então os dados apresentados na Tabela 13.6. Note que o tratamento A tem 6 repetições, o tratamento B tem 4 repetições e o tratamento C tem 5 repetições.

Para fazer a análise de variância dos dados apresentados na Tabela 13.6, é preciso calcular:

a) os graus de liberdade:
do total: $15 - 1 = 14$
de tratamentos: $3 - 1 = 2$
do resíduo: $14 - 2 = 12$

b) o valor de C:

$$C = \frac{245^2}{15} = 4001{,}67$$

Tabela 13.6

Dados, segundo o tratamento e os respectivos totais

Tratamento		
A	B	C
15	23	19
10	16	15
13	19	21
18	18	14
15		16
13		
84	76	85

c) a soma de quadrados total:

$$SQT = 15^2 + 10^2 + ... + 16^2 - 4\,001{,}67 = 159{,}33$$

d) a soma de quadrados de tratamentos:

$$SQTrat = \frac{84^2}{6} + \frac{76^2}{4} + \frac{85^2}{5} - 4001{,}67 = 63{,}33$$

e) a soma de quadrados de resíduo:

$$SQR = SQT - SQTr$$
$$= 159{,}33 - 63{,}33 = 96{,}00$$

f) o quadrado médio de tratamentos:

$$QMTr = \frac{63{,}33}{2} = 31{,}67$$

g) o quadrado médio de resíduo:

$$QMR = \frac{96{,}00}{12} = 8{,}00$$

h) o valor de F:

$$F = \frac{31{,}67}{8{,}00} = 3{,}96$$

Os valores calculados estão apresentados na Tabela 13.7

Tabela 13.7

Análise de variância dos dados apresentados na Tabela 13.6

Causas de variação	GL	SQ	QM	F
Tratamentos	2	63,33	31,67	3,96
Resíduo	12	96,00	8,00	
Total	14	159,33		

Ao nível de significância de 5%, com dois e 12 graus de liberdade, o valor de F dado na Tabela A.4 do Apêndice é 3,89. Como o valor calculado de F é 3,96, maior do que 3,89, conclui-se que as médias não são iguais.

Para comparar as médias de tratamentos duas a duas, aplica-se o teste de Tukey que, neste caso, é aproximado, porque os tratamentos têm número diferente de repetições. A diferença mínima significativa (d.m.s.) é dada pela fórmula:

$$\text{d. m. s.} = q\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right)\frac{QMR}{2}}$$

onde r_i é o número de repetições do i-ésimo tratamento e r_j é o número de repetições do j-ésimo tratamento.

No caso do exemplo, para comparar a média de A com a média de B, tem-se:

$$\text{d. m. s.} = 3{,}77\sqrt{\left(\frac{1}{6} + \frac{1}{4}\right)\frac{8{,}00}{2}}$$
$$= 4{,}87$$

Para comparar A com C, tem-se:

$$\text{d. m. s.} = 3{,}77\sqrt{\left(\frac{1}{6} + \frac{1}{5}\right)\frac{8{,}00}{2}}$$
$$= 4{,}57$$

Para comparar B com C, tem-se:

$$\text{d. m. s.} = 3{,}77\sqrt{\left(\frac{1}{4} + \frac{1}{5}\right)\frac{8{,}00}{2}}$$
$$= 5{,}06$$

Os valores absolutos das diferenças entre as médias estão na Tabela 13.8. Como o valor absoluto da diferença entre A e B é maior do que a respectiva d.m.s., conclui-se que, em média, A é diferente de B, ao nível de significância de 5%.

Tabela 13.8
Valores das diferenças entre as médias

Pares de médias	Valor absoluto da diferença
A e B	$\|14-19\| = 5$
A e C	$\|14-17\| = 3$
B e C	$\|19-17\| = 2$

13.4 - Exercícios Resolvidos

13.4.1 - *Com base nos dados apresentados na Tabela 13.9, verifique se existe diferença estatística entre os grupos. Note que são três grupos em comparação. No grupo operado foi feita a remoção das glândulas salivares maiores, e no grupo pseudo-operado foram executados todos os tempos cirúrgicos, mas nenhuma glândula foi removida.*

Tabela 13.9
Taxa de glicose, em miligramas por 100ml de sangue, em ratos Wistar machos de 60 dias, segundo o grupo

Grupo		
Operado	Pseudo-operado	Normal
96,0	90,0	86,0
95,0	93,0	85,0
100,0	89,0	105,0
108,0	88,0	105,0
120,0	87,0	90,0
110,5	92,5	100,0
97,0	87,5	95,0
92,5	85,0	95,0

Fonte: GUIMARÃES et alii (1979)

Para fazer a análise de variância, é preciso calcular:

$$C = \frac{2292,0^2}{24} = 218\ 886,00$$

$$SQT = 220\ 722,00 - 218\ 886,00 = 1\ 836,00$$

$$SQTr = \frac{1756826,00}{8} - 218\ 886,00 = 717,25$$

$$SQR = 1\ 836,00 - 717,25 = 1\ 118,75$$

$$QMT = \frac{717,25}{2} = 358,625$$

$$QMR = \frac{1118,75}{21} = 53,274$$

$$F = \frac{358,625}{53,274} = 6,73$$

Estes valores estão apresentados na Tabela 13.10. O valor de F é significante ao nível de 5%.

Tabela 13.10

Análise de variância dos dados da Tabela 13.9

Causas de variação	GL	SQ	QM	F
Grupos	2	717,25	358,625	6,73
Resíduo	21	1 118,75	53,274	
Total	23	1 836,00		

Para aplicar o teste de Tukey ao nível de significância de 5%, é preciso obter o valor de q, para 3 tratamentos e 21 graus de liberdade no resíduo. Como na tabela não existe esse valor, faz-se uma interpolação. Com 3 tratamentos e 20 graus de liberdade no resíduo, $q = 3,58$; com 3 tratamentos e 24 graus de liberdade do resíduo, $q = 3,53$. Então, aumentando $24 - 20 = 4$ graus de liberdade, o valor de q diminuiu $3,58 - 3,53 = 0,05$. Logo, aumentando de $21 - 20 = 1$ grau de liberdade, o valor q diminuirá x, tal que:

$$4 \rightarrow 0,05$$
$$1 \rightarrow x$$

$$x = \frac{0,05}{4} = 0,0125$$

Portanto, com 3 tratamentos e 21 graus de liberdade no resíduo,

$$q = 3,58 - 0,01 = 3,57$$

Para proceder ao teste de Tukey, é preciso calcular:

$$\text{d.m.s.} = 3,57\sqrt{\frac{53,247}{8}} = 9,21.$$

Os valores absolutos das diferenças de médias estão apresentados na Tabela 13.11. A taxa de glicose é, em média, maior nos operados do que nos pseudo-operados, ao nível de significância de 5%.

Tabela 13.11
Valores absolutos das diferenças de médias

Pares de médias	Diferença (Valor absoluto)
Operado vs pseudo	$\lvert 102{,}375 - 89{,}000 \rvert = 13{,}375$
Operado vs normal	$\lvert 102{,}375 - 95{,}125 \rvert = 7{,}250$
Pseudo vs normal	$\lvert 89{,}000 - 95{,}125 \rvert = 6{,}125$

13.4.2 - *Faça a análise de variância e aplique o teste de Tukey aos dados apresentados na Tabela 13.12.*

Tabela 13.12
Dados segundo a amostra

Amostra		
A	B	C
24	9	20
19	16	25
26	14	18
25	9	19
22	12	18
26		
23		
27		

Para fazer a análise de variância é preciso calcular:

$$C = \frac{352^2}{18} = 6883{,}56$$

$$SQT = 7\,448 - 6883{,}56 = 564{,}44$$

$$SQTr = 7\,328 - 6883{,}56 = 444{,}44$$

$$SQR = 564{,}44 - 444{,}44 = 120{,}00$$

$$QMTr = \frac{444{,}44}{2} = 222{,}22$$

$$QMR = \frac{120{,}00}{15} = 8{,}00$$

$$F = \frac{222{,}22}{8{,}00} = 27{,}78$$

Os valores calculados estão apresentados na Tabela 13.13. O valor de F é significante ao nível de 5%.

Tabela 13.13

Análise de variância dos dados da Tabela 13.12

Causas de variação	GL	SQ	QM	F
Amostras	2	444,44	222,22	27,78
Resíduo	15	120,00	8,00	
Total	17	564,44		

A comparação de médias, 2 a 2, é feita pelo teste de Tukey. Para comparar as médias de A e B, ou as médias de A e C, calcula-se:

$$\text{d.m.s.} = 3{,}67\sqrt{\left(\frac{1}{8}+\frac{1}{5}\right)\frac{8{,}00}{2}} = 4{,}18$$

e, para comparar as médias de B e C, calcula-se:

$$\text{d.m.s.} = 3{,}67\sqrt{\frac{8{,}00}{2}} = 4{,}64$$

Os valores absolutos das diferenças entre as médias das amostras estão na Tabela 13.14. Ao nível de 5%, a média de B é significativamente menor do que as médias de A e C.

Tabela 13.14

Valores absolutos das diferenças entre as médias das amostras

Pares de médias	Valor absoluto da diferença
A e B	$\lvert 24-12 \rvert = 12$
A e C	$\lvert 24-20 \rvert = 4$
B e C	$\lvert 12-20 \rvert = 8$

13.5 — Exercícios Propostos

13.5.1 - *Faça a análise de variância e aplique o teste de Tukey aos dados apresentados na Tabela 13.15.*

13.5.2 - *Faça a análise de variância e aplique o teste de Tukey aos dados apresentados na Tabela 13.16.*

Tabela 13.15
Dados segundo o tratamento

Tratamento				
A	B	C	D	E
17	15	15	15	20
19	8	23	9	13
13	9	17	10	14
19	12	21	14	17

Tabela 13.16
Dados segundo o tratamento

Tratamento			
A	B	C	D
11	1	21	14
20	9	13	5
18	8	11	7
15	2		10

13.5.3 - *Faça a análise de variância dos dados apresentados na Tabela 12.8 do Capítulo 12. Verifique que o valor calculado de F é igual ao quadrado do valor de t, calculado para os mesmos dados. Isto não é coincidência; com 1 grau de liberdade no numerador,* $F = t^2$.

13.5.4 - *Verifique que os valores de F com um grau de liberdade no numerador, apresentados nas tabelas do Apêndice, são iguais aos quadrados dos valores de t, apresentados na tabela do Apêndice, para o mesmo* α *e com os mesmos graus de liberdade. Então, para comparar dois tratamentos, é indiferente usar teste t ou o teste F. Ambos levam à mesma conclusão.*

CAPÍTULO

14

Intervalo de Confiança

Imagine uma amostra casual simples de n elementos. A média dos dados dessa amostra constitui uma *estimativa* da média da população de onde essa amostra proveio. Para indicar a *precisão* dessa estimativa, calcula-se o intervalo de confiança para a média na forma dada neste Capítulo. Mas, antes de conceituar intervalo de confiança, é preciso entender o que é erro padrão da média.

14.1 - Erro Padrão da Média

Uma população é constituída pelos valores 4,10 e 16. A média dessa população, que se indica por μ (lê-se mi), é:

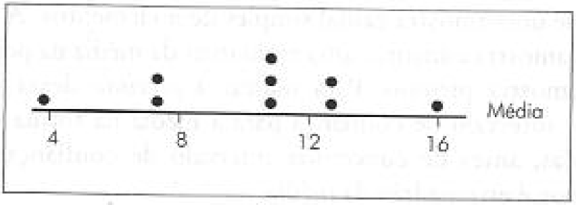
$$\mu = \frac{4+10+16}{3} = 10$$

Considere todas as amostras possíveis de dois elementos que podem ser retirados dessa população, admitindo que todo elemento retirado para compor a amostra é reposto, antes da retirada do segundo. Essas amostras, e as respectivas médias, estão na Tabela 14.1. É fácil ver, observando a Figura 14.1, que as médias das amostras distribuem-se em torno da média $\mu = 10$ da população.

Tabela 14.1
Médias das amostras de dois elementos obtidos da população constituída pelos números 4, 10 e 16

Amostra	Média
4 e 4	4
4 e 10	7
4 e 16	10
10 e 4	7
10 e 10	10
10 e 16	13
16 e 4	10
16 e 10	13
16 e 16	16

Figura 14.1 *Distribuição das médias das amostras*



Para medir o grau de dispersão das médias das amostras em torno da média da população, calcula-se a *variância da média*. Essa medida, que se indica por $\sigma^2_{\bar{x}}$, é dada pela fórmula:

$$\sigma^2_{\bar{x}} = \frac{\sum_{i=1}^{r}(\bar{x}_i - \mu)^2}{r}$$

onde $\bar{x}_i$ é a média da *i*-ésima amostra e *r* o número de amostras que podem ser obtidas da população.

Para as médias apresentadas na Tabela 14.1, a variância da média é:

$$\sigma^2_{\bar{x}} = \frac{(4-10)^2 + (7-10)^2 + \ldots + (16-10)^2}{9} = \frac{108}{9} = 12$$

Na prática, é impossível calcular a variância da média pela fórmula apresentada, porque o pesquisador dispõe de uma única amostra para estimar a média μ da população e obter uma medida de precisão dessa estimativa — e não de todas as amostras possíveis.

Mas existe uma saída. Já se demonstrou que uma estimativa da variância da média é dada pela fórmula:

$$s_{\bar{x}}^2 = \frac{s^2}{n}$$

onde s^2 é a variância da amostra.

As médias, as variâncias e as variâncias das médias das amostras apresentadas na Tabela 14.1 estão na Tabela 14.2. Note que a média das médias coincide com a média $\mu = 10$ da população e que a média das variâncias das médias das amostras é igual a $\sigma_{\bar{x}}^2 = 12$, calculada anteriormente.

Tabela 14.2

Médias, variâncias e variâncias das médias das amostras apresentadas na Tabela 14.1

Amostra	Média	Variância	Variância da média
4 e 4	4	0	0
4 e 10	7	18	9
4 e 16	10	72	36
10 e 4	7	18	9
10 e 10	10	0	0
10 e 16	13	18	9
16 e 4	10	72	36
16 e 10	13	18	9
16 e 16	16	0	0
Média	10	24	12

Por definição, *erro padrão da média* é a raiz quadrada com sinal positivo da variância da média. Indica-se a estimativa do erro padrão da média por $s_{\bar{x}}$. Logo

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

14.2 - Intervalo de Confiança

Seja X uma variável aleatória com distribuição normal de média μ e variância σ^2. Com base em uma amostra casual simples de n elementos dessa população, obtêm-se as estimativas $\bar{x}$ e s^2, de μ e σ^2, respectivamente.

A expressão

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

é o *intervalo de confiança* para a média μ da população. Nessa expressão, t é um valor encontrado na tabela de t (dada neste livro em Apêndice), com $n - 1$ graus de liberdade e ao nível de significância α.

Para melhor entender como se calcula um intervalo de confiança, convém discutir um exemplo. Seja X a variável aleatória que representa a taxa de colesterol no plasma sangüíneo humano. Imagine que, com base em uma amostra casual simples de $n = 25$ indivíduos, foram obtidos a média $\bar{x} = 198$mg/100ml e o desvio padrão $s = 30$mg/100ml.

É fácil calcular o intervalo de confiança para μ. Seja $\alpha = 10\%$. O valor de t na Tabela A.6 do Apêndice, com $n - 1 = 25 - 1 = 24$ graus de liberdade, é 1,71. A expressão do intervalo de confiança fica, então, como segue:

$$198 - 1{,}71\frac{30}{\sqrt{25}} < \mu < 198 + 1{,}71\frac{30}{\sqrt{25}}$$

$$187{,}74 < \mu < 208{,}26$$

A interpretação do intervalo de confiança exige cuidado. Quando são obtidas muitas amostras de n elementos de uma mesma população e se determina, para cada amostra, um intervalo de confiança $(100 - \alpha)\%$ desses intervalos contém a média μ da população.

Mas, na prática, o pesquisador dispõe de uma única amostra, que fornece uma estimativa da média e uma estimativa do desvio padrão. Então o pesquisador não sabe se a média da população está, ou não está contida no intervalo que calculou. Sabe, porém, que $(100 - \alpha)\%$ dos intervalos de confiança calculados dessa forma contêm a média da população.

14.3 - Alguns Pontos Básicos

O valor $(100 - \alpha)\%$ é denominado *nível de confiança*. Diz-se, por isso, que os intervalos são de confiança $(100 - \alpha)\%$.

Na área biológica, é comum apresentar os valores $\bar{x}$ e $s_{\bar{x}}$ escritos na forma $\bar{x} \pm s_{x}$. Esta expressão pode ser vista como um intervalo de confiança para μ, mas com nível de confiança indeterminado. Isto porque o valor de t, nessa expressão, é igual a 1. Mas o valor de t, na tabela, depende do número de graus de liberdade associados a $s_{\bar{x}}$. Então, se n for pequeno, o intervalo $\bar{x} \pm s_{x}$ pode ter nível de confiança relativamente baixo. Veja um exemplo.

Considere uma amostra de seis elementos. Para calcular um intervalo de 90% de confiança para μ, usando esta amostra, o valor de t dado na Tabela A.6 do Apêndice é 2,02. Então o intervalo

$$\bar{x} \pm s_{\bar{x}}$$

é menor do que o intervalo de 90% de confiança, que é

$$\bar{x} \pm 2{,}02\, s_{\bar{x}}.$$

14.4.1 - *A taxa de glicose no sangue humano é uma variável aleatória com distribuição aproximadamente normal. Suponha que, com base em uma amostra de 30 pessoas, foi obtida a média $\overline{x}$ = 102mg de glicose por 100ml de sangue e o desvio padrão s = 6mg. Calcule o intervalo de 90% de confiança para μ.*

O intervalo de confiança é dado pela expressão:

$$\overline{x}-t\frac{s}{\sqrt{n}}<\mu<\overline{x}+t\frac{s}{\sqrt{n}}$$

Na tabela de t, com $30 - 1 = 29$ graus de liberdade e ao nível de significância de $\alpha = 10\%$, encontra-se $t = 1{,}70$. Então o intervalo de confiança fica como segue:

$$102-1{,}70\frac{6}{\sqrt{30}}<\mu<102+1{,}70\frac{6}{\sqrt{30}}$$

$$100{,}14 < \mu < 103{,}86$$

14.4.2 - *Com base nos dados apresentados na Tabela 14.3, determine o intervalo de 95% de confiança para μ.*

Tabela 14.3

Taxa de glicose, em miligramas por 100ml de sangue, de ratos Wistar machos de 40 dias

100,0	92,5
87,5	94,0
110,0	100,0
99,5	100,0

Fonte: GUIMARÃES et alii (1979)

A média e o desvio padrão dos dados apresentados na Tabela 14.3 são, respectivamente, $\overline{x} = 97{,}9$ e $s = 6{,}7$. O valor de t, com $8 - 1 = 7$ graus de liberdade e para $\alpha = 5\%$ é 2,36. Então o intervalo de 95% de confiança para μ é:

$$97{,}9-2{,}36\frac{6{,}7}{\sqrt{8}}<\mu<97{,}9+2{,}36\frac{6{,}7}{\sqrt{8}}$$

$$92{,}3 < \mu < 103{,}5$$

14.5 - Exercícios Propostos

14.5.1 - *Seja X a variável aleatória que representa a pressão sangüínea sistólica em indivíduos com idade entre 20 e 25 anos. Essa variável tem distribuição aproximadamente normal. Suponha que, com base em uma amostra de 100 indivíduos, foi obtida a média* $\bar{x} = 123mm$ *de mercúrio e o desvio padrão* $s = 8$ *milímetros de mercúrio. Determine o intervalo de 90% de confiança para* μ.

14.5.2 - *Seja X a variável aleatória que representa a taxa de hemoglobina em mulheres. Imagine que, com base em uma amostra aleatória de 20 mulheres, obteve-se a média* $\bar{x} = 16{,}2g$ *de hemoglobina por 100ml de sangue e o desvio padrão* $s = 1{,}1g$. *Determine o intervalo de 99% de confiança para* μ, *supondo que X é uma variável com distribuição normal.*

14.5.3 - *Seja X a variável aleatória que representa a estatura ao nascer para o sexo masculino. Com base em 28 recém-nascidos masculinos, obtiveram-se* $\bar{x} = 50cm$ *e* $s = 2{,}5$ *cm. Supondo distribuição normal, determine o intervalo de 90% de confiança para* μ.

14.5.4 - *Seja X a variável aleatória que representa a taxa de glicose no sangue humano. Determine o intervalo de 95% de confiança para* μ, *supondo que uma amostra de 25 pessoas forneceu a média* $\bar{x} = 95mg$ *de glicose por 100 ml de sangue e o desvio padrão* $s = 6mg$. *Suponha que X tem distribuição normal.*