

Métodos estatísticos empregados para comparação de métodos analíticos

Statistical methods used for comparison of analytical methods

Maria S. Aurora Prado¹; Martin Steppe²; Erika R. M. Kedor-Hackmann^{*2} & Maria Inês R. M. Santoro²

RESUMO – A estatística pode ser considerada como a ciência que estuda a organização, a descrição, a análise e a interpretação dos dados experimentais. A estatística é aplicável a qualquer ramo do conhecimento onde se manipulem dados experimentais. Assim, a Física, a Química, a Engenharia, a Economia, a Medicina, a Biologia, as Ciências Sociais, as Ciências Administrativas, etc., tendem cada vez mais a servir-se dos métodos estatísticos como ferramentas de trabalho. Profissionais com diferentes tipos de formação são preparados para efetuar determinações físicas, químicas e físico-químicas em diferentes materiais e amostras. Esta tarefa envolve, entre outros aspectos, a determinação quantitativa de elementos e/ou substâncias presentes em matrizes das mais diversas. Por outro lado, os laboratórios de medições são responsáveis pela garantia da qualidade de seus resultados analíticos. Desta maneira, os integrantes dos laboratórios necessitam estabelecer critérios para a avaliação dos resultados laboratoriais obtidos. O objetivo deste trabalho foi o de reunir, em língua portuguesa, conhecimentos básicos sobre o uso da estatística na comparação de métodos analíticos utilizados rotineiramente em laboratórios de pesquisa e análise de medicamentos e cosméticos.

PALAVRAS-CHAVE – Estatística, testes de significância, precisão, exatidão, medicamentos, cosméticos.

SUMMARY – Statistics can be considered as the science that studies the organization, description, analysis and the interpretation of experimental data. Statistics is widely used in many areas of scientific investigation. Thus, Physics, Chemistry, Engineering, Economy, Medicine, Biology, Social Sciences, Administrative Sciences etc., tend to use more and more statistical methods as working tools. Professionals of different areas are able to make physical and chemical determinations, in different materials and samples leading quantitative determination of elements and/or substances. On the other hand, the analytical laboratories are responsible for the quality of the results, so the analysts need to establish criteria for the evaluation of the data obtained in the analysis. The aim of this work was to present, in Portuguese, the basic statistical elements used to compare the results obtained in quantitative determination of the same samples using different methods in the analysis of pharmaceutical preparations and cosmetics.

KEYWORDS – Statistics, significance tests, precision, accuracy, pharmaceutical preparation, cosmetics.

INTRODUÇÃO

A estatística é uma ferramenta de grande utilidade para resolver grande variedade de problemas de Química Analítica. Seu emprego permite a avaliação objetiva de dados e promove o uso eficiente de fontes experimentais para extrair a quantidade máxima de informações a partir de dados obtidos por experimentos planejados (Anderson, 1987).

A estatística é aplicável a qualquer ramo do conhecimento onde se manipulem dados experimentais. Assim, a Física, a Química, a Engenharia, a Economia, a Medicina, a Biologia, as Ciências Sociais, as Ciências Administrativas, etc. tendem cada vez mais a servirem-se dos métodos estatísticos como ferramentas de trabalho, daí sua grande e crescente importância (Costa Neto, 1987).

Os métodos estatísticos de modo similar a outros procedimentos matemáticos, não podem revelar nada que já não esteja implícito nos dados.

Exatidão e precisão são os critérios mais importantes na avaliação de um método analítico. Exatidão e precisão, juntos determinam o erro de uma medida analítica e são os critérios usados quando

se julga a qualidade de um método analítico (Karnes & March, 1993; Lee & McAllister, 1996).

A exatidão representa o grau de concordância entre os resultados individuais encontrados e um valor aceito como referência. A precisão representa o grau de repetibilidade entre os resultados de análise individuais quando o procedimento é aplicado diversas vezes a uma mesma amostra homogênea, em idênticas condições de teste. Normalmente é expressa através do desvio padrão relativo para um número de amostras estatisticamente significativo (Kateman & Pijpers, 1981; Caulcutt & Boddy, 1994; Karnes & March, 1991; Swartz & Krull, 1998; Snyder et al., 1997; Bolton, 1990).

Conceitos gerais

Testes de significância

Os testes de significância estatística informam se a qualidade de um lote de material, a produção de certo tipo de produto ou a batelada de itens recebidos do fornecedor diferem "significativamente" de um valor padrão ou da qualidade de um segundo ou mais lotes ou fontes. Estes testes podem ser uti-

Recebido em 22/5/2001

¹Alunos de Programa de Pós-Graduação em Fármaco e Medicamentos, área Produção e Controle Farmacêuticos. Departamento de Farmácia, Faculdade de Ciências Farmacêuticas da Universidade de São Paulo;

²Professoras Titulares do Departamento de Farmácia, Faculdade de Ciências Farmacêuticas da Universidade de São Paulo.

lizados para comparar por exemplo, dois medicamentos diferentes, ou dois métodos analíticos, a precisão da medida de dois calibres diferentes, o rendimento de dois tornos diferentes, etc. Os testes de significância são amplamente utilizados na avaliação dos resultados experimentais. (Feigenbaum, 1994; Christmann, 1978; Miller & Miller, 1993).

Os testes de significância podem detectar diferenças em porcentagem de defeituosos, médias, dispersões e outras medidas. Esses testes foram desenvolvidos com a finalidade de solucionar problemas industriais freqüentemente encontrados, pois a diferença entre níveis satisfatórios e insatisfatórios da qualidade é quase sempre tão baixa que pode ser mascarada por variações aleatórias em amostras pequenas. A utilização de testes de significância reduz consideravelmente a probabilidade de se chegar a conclusões incorretas em tais casos (Feigenbaum, 1994).

Ao realizar-se o teste de significância comprova-se a veracidade de uma hipótese, a qual será considerada válida até prova em contrário. Essa hipótese será testada com base em resultados amostrais, podendo ser aceita ou rejeitada.

Designando-se como H_0 a hipótese existente a ser testada, denominada de **hipótese nula**, e como H_1 a **hipótese alternativa**, ter-se-ão elementos para decidir sobre a afirmação dada por H_0 através de testes sobre a hipótese alternativa H_1 . O teste irá levar à aceitação ou rejeição da hipótese H_0 , o que corresponde, portanto, à negação ou afirmação de H_1 . Para manter uniformidade, enuncia-se o resultado final sempre em termos da hipótese H_0 , ou seja, aceitar ou rejeitar H_0 . O valor α é denominado **nível de significância** do teste e **C**, é denominado **valor crítico**. A região da curva da distribuição que contém os valores para os quais rejeita-se a hipótese H_0 é chamada **região crítica** ou **região de rejeição**, e a região dos valores para os quais aceitamos a hipótese H_0 é chamada **região de aceitação**. Existem dois tipos de erros que se está sujeito a cometer; isto é, pode-se rejeitar H_0 (devido ao valor observado ser maior do que **C**) quando na realidade H_0 é verdadeiro. Neste caso o erro é denominado **erro do tipo I** e a probabilidade de sua ocorrência é representada por α . Alternativamente, pode-se aceitar a hipótese H_0 (devido ao valor observado ser menor ou igual a **C**) quando na realidade a hipótese é falsa. Neste caso o erro é chamado **erro do tipo II** e a probabilidade de sua ocorrência é representada por β (Christmann, 1978; Miller & Miller, 1993; Costa Neto, 1987).

Os testes de hipótese podem ser unilaterais ou monocaudais e bilaterais ou bicaudais, ou seja, são empregados para verificar se os desvios do valor real do parâmetro estão unicamente para mais ou unicamente para menos, ou para mais ou para menos, em relação ao valor testado (Christmann, 1978; Miller & Miller, 1993; Anderson, 1987).

Os testes de significância mais importantes são o teste **t**, e o teste **F**, (Miller & Miller, 1993; Feigenbaum, 1994).

2. Comparação da exatidão e precisão de métodos

Os testes de significância mais importantes são o teste **t**, usado para verificar a significância de uma média ou da diferença entre duas médias e o teste **F**,

utilizado para comparar duas variâncias, para um determinado nível de significância (Miller & Miller, 1993; Skoog et al., 1996; Ohlweiler, 1976).

2.1 Comparação da exatidão de dois ou mais métodos

2.1.1 Comparação das médias de dois métodos

A comparação dos dados obtidos com dois métodos ou por dois analistas diferentes ou pelo mesmo analista, mas sob diferentes condições experimentais é freqüentemente indispensável na Química Analítica para testar-se a validade de um método analítico. Para este propósito, as duas médias experimentais \bar{x}_1 e \bar{x}_2 são comparadas usando o teste **t** (Hadjiioannou et al., 1988; Christian, 1994; Bruce et al., 1998).

O teste **t** de Student é usado para verificar se a diferença entre as médias dos valores experimentais obtidos por dois métodos diferentes é significativa, para um determinado nível de significância. Por exemplo, para comparar dois métodos analíticos, aplicados à mesma amostra, o teste estatístico poderá indicar se a diferença entre as duas médias é devida à presença de um erro determinado em um dos métodos ou decorrente das flutuações associadas aos erros indeterminados. O teste **t** também pode servir para estabelecer a identidade ou falta de identidade entre duas amostras que são analisadas pelo mesmo método. Nestas condições, as duas médias são, comparadas (Skoog et al., 1996; Ohlweiler, 1976).

Considera-se à \bar{x}_1 e \bar{x}_2 como variáveis aleatórias independentes e normalmente distribuídas com média μ_1 e μ_2 , respectivamente, desconhecidas, e desvios padrão σ_1 e σ_2 também desconhecidos, mas iguais a um valor σ . A variância, baseada em ambas amostras, é dada pela Equação 1:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{Eq. 1}$$

Onde:

S_1^2 = variância da amostra pelo método 1

S_2^2 = variância da amostra pelo método 2

n_1 = número de determinações pelo método 1

n_2 = número de determinações pelo método 2

O valor de **t** é calculado pela Equação 2:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{Eq. 2}$$

Onde:

\bar{x}_1 = média da amostra pelo método 1

\bar{x}_2 = média da amostra pelo método 2

S = desvio-padrão, baseada em ambos métodos, obtido a partir da raiz quadrada da variância calculada para os dois métodos.

t apresenta $n_1 + n_2 - 2$ graus de liberdade

Se o valor encontrado para **t** excede o valor tabelado ou crítico, considera-se que a diferença entre os resultados dos dois métodos é significativa, para um determinado nível de significância (Miller & Miller, 1993; Skoog et al., 1996; Ohlweiler, 1976; Da Silva Guedes et al., 1988).

A decisão do teste será dada pelo critério apresentado na Tab. I onde $t_{n_1+n_2-2, \alpha}$ e $t_{n_1+n_2-2, \alpha/2}$ são

obtidos na Tab. da distribuição t de Student para (n_1+n_2-2) graus de liberdade (Chui et al., 1997).

TABELA I
Comparação de médias de dados não emparelhados com desvios-padrão desconhecidos mas supostamente iguais

Hipótese nula	Hipótese alternativa	Rejeita-se H_0 se
$H_0: \mu_1 = \mu_2$	$H_1: \mu_1 \neq \mu_2$ (caso bilateral)	$ t_{\text{calc}} > (t_{n_1+n_2-2, \alpha/2})$
$H_0: \mu_1 = \mu_2$	$H_1: \mu_1 < \mu_2$ (caso unilateral)	$t_{\text{calc}} < (-t_{n_1+n_2-2, \alpha})$
$H_0: \mu_1 = \mu_2$	$H_1: \mu_1 > \mu_2$ (caso unilateral)	$t_{\text{calc}} > (t_{n_1+n_2-2, \alpha})$

Exemplos

1. Caso unilateral

Dois fornecedores de reagentes químicos para a determinação de fósforo pelo método colorimétrico foram testados pelo laboratório de química do Instituto de Pesquisas Tecnológicas (IPT-USP). Foram analisadas 7 amostras de dois lotes diferentes. Ao nível de 5% de significância, pode-se afirmar que os resultados obtidos com os reagentes químicos do fornecedor A são mais baixos dos que os obtidos com reagentes do fornecedor B?

Temos:

$$\begin{aligned} \bar{x}_A &= 0,016886 & S_A &= 0,003664 & n_1 &= 7 \\ \bar{x}_B &= 0,017057 & S_B &= 0,003109 & n_2 &= 7 \\ \text{Hipótese nula} & & H_0: \mu_A &= \mu_B \\ \text{Hipótese alternativa} & & H_1: \mu_A &< \mu_B \end{aligned}$$

Substituindo os dados nas Equações 1 e 2 temos:

$$S^2 = \frac{(7-1)(0,003664)^2 + (7-1)(0,003109)^2}{7+7-2}$$

$$S^2 = 1,154572 \times 10^{-5}$$

$$S = 0,0033979$$

$$t_{\text{calc}} = \frac{0,016886 - 0,017057}{0,0033979 \sqrt{\frac{1}{7} + \frac{1}{7}}} = -0,09415$$

$$t_{\text{tabulado}} = t_{n_1+n_2-2, \alpha} = t_{12; 5\%} = 1,782$$

$$t_{\text{calculado}} > -t_{\text{tabulado}} \quad (-0,09415 > -1,782)$$

Aceita-se a hipótese nula, ou seja, pode-se aceitar que os resultados obtidos com os reagentes de ambos os fornecedores são iguais e, portanto, não apresentam diferença significativa (Chui et al., 1997).

2. Caso bilateral

Na comparação de dois métodos: eletroforese capilar (EC) e cromatografia líquida de alta eficiência (CLAE) para a determinação de diclofenaco de sódio em preparações farmacêuticas (comprimidos, amostra comercial) obtiveram-se os seguintes resultados (Aurora Prado et al., 1999):

$$\begin{aligned} 1: \text{ método da eletroforese capilar} \\ 2: \text{ método da cromatografia líquida de alta eficiência} \\ \bar{x}_1: 98,8 & S_1: 1,65 & n_1: 10 \\ \bar{x}_2: 99,12 & S_2: 0,37 & n_2: 10 \end{aligned}$$

Ao nível de 5% de significância ($P=95\%$), pode-se afirmar que as médias obtidas pelos dois métodos são iguais?

$$\begin{aligned} \text{Hipótese nula} & H_0: \mu_1 = \mu_2 \\ \text{Hipótese alternativa} & H_1: \mu_1 \neq \mu_2 \end{aligned}$$

Substituindo os dados nas Equações 1 e 2 temos:

$$S^2 = \frac{(10-1)(1,65)^2 + (10-1)(0,37)^2}{10+10-2}$$

$$S^2 = 1,4297$$

$$S = 1,1957$$

$$|t_{\text{calc}}| = \frac{98,08 - 99,12}{1,1957 \sqrt{\frac{1}{10} + \frac{1}{10}}} = |-1,9448| = 1,9448$$

$$t_{\text{tabulado}} = t_{n_1+n_2-2, \alpha/2} = t_{18; 2,5\%} = 2,101$$

$$|t_{\text{calculado}}| < t_{\text{tabulado}} \quad (1,9448 < 2,101)$$

Aceita-se a hipótese nula, ou seja, pode-se aceitar que as médias dos dois métodos são iguais e portanto as médias dos dois métodos (EC e CLAE) não apresentam diferença significativa.

2.1.2 Comparação das médias de mais de dois métodos

Em todo trabalho analítico são necessárias comparações nas quais podem intervir mais de duas médias. Existem casos em que, por exemplo, comparam-se resultados como a concentração média de proteínas em amostras armazenadas em diferentes condições, os resultados médios obtidos da concentração de um analito utilizando diferentes métodos, a média dos resultados obtidos numa titulação efetuada por diferentes operadores com os mesmos equipamentos. Para solucionar estes problemas utiliza-se a *Análise de Variância* (ANOVA, *Analysis of Variance*).

A análise de variância é um método suficientemente poderoso para poder identificar diferenças entre as médias populacionais devidas a várias causas atuando simultaneamente sobre os elementos da população (Miller & Miller, 1993; Costa Neto, 1987). Assim, quando se quer testar uma hipótese envolvendo mais de duas populações (partindo da suposição que as populações são normalmente distribuídas e com variâncias iguais) emprega-se a análise de variância (Chui et al., 1997).

2.1.2.1 Critérios de classificação

A. Amostras do mesmo tamanho

Considerando-se h como número de amostras retiradas de h populações, cada uma com n elementos e média μ_i ($i = 1, 2, \dots, h$). Em primeiro lugar deve-se testar a hipótese nula, ou seja, $H_0: \mu_1 = \mu_2 = \dots = \mu_h$ contra a hipótese alternativa (H_1) supondo que pelo menos uma das médias populacionais seja diferente.

Será empregada a notação segundo a qual

$$x_{ij} (i = 1, 2, \dots, h; j = 1, 2, \dots, n)$$

é o j -ésimo valor da i -ésima amostra de n elementos. As médias das amostras serão $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_h$ e a média de todos os valores agrupados juntos será \bar{x} (Miller & Miller, 1993; Costa Neto, 1987).

A Tabela II apresenta, como exemplo, os resultados obtidos na pesquisa sobre a estabilidade de um reagente fluorescente, em diferentes condições de armazenamento. Os dados correspondem aos sinais de fluorescência obtidos a partir de soluções diluídas com mesma concentração (Miller & Miller, 1993).

O teste de hipóteses pode ser formulado:

H_0 : Hipótese nula (não existem diferenças nas médias)

H_1 : Hipótese alternativa (aceitam-se diferenças nas médias)

O estudo do caso exemplificado será efetuado nas seguintes etapas:

- Estimativa da variância populacional a partir da variância entre médias amostrais (VEA)
- Estimativa da variância populacional da variância dentro de amostras (VDA)
- Calculando a relação $F = VEA/VDA$
- Comparação do F (calculado) com o F obtido da Tabela F -Snedecor (o nível de significância e os graus de liberdade a serem adotados neste teste de hipótese devem ser definidos previamente).
- Decisão sobre a hipótese nula (H_0 : as médias são iguais)
 - Se $F_{calculado} > F_{tabelado}$ Rejeitar H_0
 - Se $F_{calculado} < F_{tabelado}$ Aceitar H_0

Sabendo-se ainda que,

1) Número total de medidas (dado pela Equação 3):

$$N = hn \quad \text{Eq. 3}$$

$$N = 4 \times 3 = 12$$

2) Soma de quadrados totais (SQT), dada pela Equação 4:

$$SQT = \sum_i \sum_j (x_{ij} - \bar{x})^2 \quad \text{Eq. 4}$$

$$SQT = (102-98)^2 + (100-98)^2 + \dots + (94-98)^2 = 210$$

3) Soma de quadrados entre amostras (SQE), dada pela Equação 5:

$$SQE = \sum_i \sum_j (\bar{x}_i - \bar{x})^2 \quad \text{Eq. 5}$$

$$SQE = 3 [(101-98)^2 + (102-98)^2 + (97-98)^2 + (92-98)^2] = 186$$

4) Soma dos quadrados residuais (SQR), dada pela Equação 6:

$$SQR = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 \quad \text{Eq. 6}$$

$$SQR = (102-101)^2 + (100-101)^2 + \dots + (94-92)^2 = 24$$

TABELA II
Sinais de fluorescência obtidos para soluções do reagente em diferentes condições de armazenamento

Soluções (condições)	Leituras (\bar{x}_i)	Média das 3 leituras (\bar{x}_i)
A (recentemente preparada)	102, 100, 101	101
B (1 hora ao abrigo da luz)	101, 101, 104	102
C (1 hora sob luz fraca)	97, 95, 99	97
D (1 hora sob luz brilhante)	90, 92, 94	92
		$\bar{X} = 98$

5) Estimativa entre amostras (S_E^2) ou quadrado médio entre amostras, dada pela Equação 7:

$$S_E^2 = \frac{n \sum_i (\bar{x}_i - \bar{x})^2}{h-1} = \frac{SQE}{h-1} \quad \text{Eq. 7}$$

Onde:

$$h-1 = \text{graus de liberdade}$$

$$h = 4$$

$$S_E^2 = \frac{186}{4-1} = 62$$

6) Estimativa residual (S_R^2) ou quadrado médio residual, dada pela Equação 8:

$$S_R^2 = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{h(n-1)} = \frac{SQR}{h(n-1)} \quad \text{Eq. 8}$$

Onde:

$$h(n-1) = \text{graus de liberdade}$$

$$h = 4; n = 3$$

$$S_R^2 = \frac{24}{4(3-1)} = 3$$

7) Número dos graus de liberdade:

Graus de liberdade entre amostras:

$$h-1 = 4-1 = 3$$

Graus de liberdade residuais:

$$h(n-1) = 4(3-1) = 8$$

Graus de liberdade totais:

$$hn-1 = 4 \times 3 - 1 = 11$$

8) Cálculo de F , Equação 9:

$$F = \frac{S_E^2}{S_R^2} \quad \text{Eq. 9}$$

$$F = \frac{63}{3} = 20,7 (F_{calculado})$$

O valor calculado de F é comparado com o valor obtido da Tabela F -Snedecor, ao nível de significância " α " e com $(h-1)$ e $h(n-1)$ graus de liberdade.

$$F_{tabelado} (F_{h-1, h(n-1); \alpha}); \alpha = 0,05$$

$$F_{3,8,0,05} = 4,066$$

Como $F_{calculado} > F_{tabelado}$, rejeita-se a hipótese nula, ou seja, as médias das amostras não são iguais ou diferem em forma significativa.

Na Tab. III está apresentada a análise de variância para o exemplo escolhido.

B. Amostras de tamanhos diferentes

Com pequenas modificações, o método apresentado no item anterior pode ser adaptado para o caso de amostras de tamanhos diferentes. Neste caso o índice j referente à caracterização do elemento dentro da amostra variará de 1 a n_i , sendo n_i o tamanho da i -ésima amostra. Como exemplo para ilustrar este caso, na Tab. IV, estão apresentados os resultados experimentais obtidos na determinação quantitativa da cefalexina, em suspensão (amostra simulada), através de três métodos analíticos (Steppe, 2000).

Ao nível de 5% de significância ($P=95\%$), pode-se afirmar que as médias obtidas pelos três métodos são iguais?

Sabendo-se que,

Hipótese nula $H_0: \mu_1 = \mu_2 = \mu_3$

Hipótese alternativa H_1 : pelo menos uma das médias é diferente.

Tem-se,

1) Número total de medidas:

$$N = \sum_i n_i$$

$$N = 6 + 9 + 9 = 24$$

2) Soma de quadrados totais (SQT):

$$SQT = \sum_i \sum_j (x_{ij} - \bar{x})^2$$

$$SQT = (100,35-100,95)^2 + (102,32-100,95)^2 + \dots + (100,36-100,95)^2 = 10,5799$$

3) Soma de quadrados entre amostras (SQE)

$$SQE = n_i \sum_i (\bar{x}_i - \bar{x})^2 \quad \text{Eq. 4}$$

$$SQE = 6(100,81-100,95)^2 + 9(101,13-100,95)^2 + 9(100,91-100,95)^2 = 0,4344$$

4) Soma dos quadrados residuais (SQR)

$$SQR = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

$$SQR = (100,35-100,81)^2 + (102,32-100,81)^2 + \dots + (99,98-100,91)^2 + (100,36-100,91)^2 = 10,1455$$

5) Estimativa entre amostras (S_E^2) ou quadrado médio entre amostras

$$S_E^2 = \frac{n_i \sum_i (\bar{x}_i - \bar{x})^2}{h-1} = \frac{SQE}{h-1}$$

Onde:

$h-1$ = graus de liberdade

$h = 3; 3 - 1 = 2$

$$S_E^2 = \frac{0,4344}{3-1} = 0,2172$$

TABELA III
Análise de variância para os dados do exemplo considerado

Fonte de variação	Soma dos quadrados	Graus de liberdade	Quadrado médio	F	Fa
Entre amostras	186	3	62	20,7	$F_{5\%}=4,066$
Residual	24	8	3		
Total	210*	11*	-		

* Soma dos valores obtidos.

TABELA IV
Resultados experimentais obtidos na determinação quantitativa da cefalexina em suspensão (amostra simulada), empregando ensaio microbiológico (EM), cromatografia líquida de alta eficiência (CLAE) e cromatografia capilar eletrocínica micelar (MEKC)

Amostra	EM	CLAE	MEKC
Simulada	100,35	101,94	101,27
	102,32	101,24	100,71
	101,27	100,99	101,78
	99,81	100,41	100,16
	100,19	101,68	101,34
	100,94	101,80	101,02
	-	100,72	101,54
	-	101,07	99,98
	100,36	100,36	
Média (\bar{x}_i)	100,81	101,13	100,91
n	6	9	9

Média global (\bar{x}) = 100,95

6) Estimativa residual (S_R^2) ou quadrado médio residual

$$S_R^2 = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{\sum_i n_i - h} = \frac{SQR}{\sum_i n_i - h}$$

Onde:

$$\sum_i n_i - h = \text{graus de liberdade}; h = 3$$

$$\sum_i n_i = 6 + 9 + 9 = 24; 24 - 3 = 21$$

$$S_R^2 = \frac{10,1455}{21} = 0,4831$$

7) Número de graus de liberdade:

Graus de liberdade entre amostras:

$$h - 1 = 3 - 1 = 2$$

Graus de liberdade residual:

$$\sum_i n_i - h = 24 - 3 = 21$$

Graus de liberdade total:

$$\sum_i n_i - 1 = 6 + 9 + 9 - 1 = 23$$

8) Cálculo de F :

$$F = \frac{S_U^2}{S_R^2}$$

$$F = \frac{0,2172}{0,4831} = 0,4495 \quad (F_{\text{calculado}})$$

O valor calculado de F é comparado com o valor obtido da Tabela F -Snedecor, ao nível de significância " α " e com $(h-1)$ e $\sum n_i - h$ graus de liberdade.

Assim,

$$F_{\text{tabelado}} (F_{n-1, i}; \alpha); \alpha = 0,05$$

$$F_{2,21;0,05} = 3,47$$

Como $F_{\text{calculado}} < F_{\text{tabelado}}$, aceita-se a hipótese nula, ou seja, pode-se dizer que as médias dos métodos são iguais.

Na Tab. V apresentam-se os dados da análise de variância para o exemplo selecionado.

TABELA V

Análise de variância dos resultados experimentais obtidos na determinação quantitativa da cefalexina, em suspensão (amostra simulada), empregando o ensaio microbiológico (EM), cromatografia líquida de alta eficiência (CLAE) e cromatografia capilar eletrocínética micelar (MEKC)

Fonte de variação	Soma dos quadrados	Graus de liberdade	Quadrado médio	F	F α
Entre amostras	0,4344	2	0,2172	0,4495	F $_{9\%}$ = 3,47
Residual	10,1455	21	0,4831		
Total	10,5799*	23*	-		

* Soma dos valores obtidos.

$$F_{2,21;0,05} = 3,47 \quad (F_{\text{tabelado}})$$

$$F_{\text{calculado}} = 0,4495$$

Portanto, $F_{\text{calculado}} < F_{\text{tabelado}}$, aceita-se a hipótese nula, ou seja, pode-se dizer que as médias dos três métodos são iguais.

2.2 Comparação da precisão de métodos

2.2.1 Comparação de duas variâncias

A comparação da precisão de dois métodos é realizada pelo teste F de Snedecor (Costa Neto, 1987; Anderson, 1987). Este teste permite comprovar se um método é mais preciso do que um outro método ou se os dois métodos diferem significativamente na precisão (Miller & Miller, 1993; Skoog et al., 1996; Ohlweiler, 1976).

No teste F considera-se a razão das duas variâncias amostrais, ou seja, a razão dos quadrados dos desvios padrão de uma mesma amostra obtidos por métodos diferentes. O cálculo de F é dado pela Equação 10:

$$F = \frac{S_1^2}{S_2^2}, \text{ sendo que } S_1^2 > S_2^2 \quad \text{Eq. 10}$$

Onde:

S_1^2 = variância da amostra obtida pelo método 1.

S_2^2 = variância da amostra obtida pelo método 2.

O teste F está associado aos graus de liberdade de S_1^2 e S_2^2 , respectivamente. Se o valor encontrado excede ao valor tabelado, em certo nível de probabilidade, a diferença entre a precisão dos dois métodos é significativa.

Os critérios para aceitação ou rejeição da hipó-

tese nula estão apresentados na Tab. VI, utilizando-se a Tabela de F -Snedecor, com n_1-1 e n_2-1 graus de liberdade.

A comparação das variâncias pode ser feita de duas maneiras:

a. Caso unilateral: para provar se o método 1 é mais ou menos preciso do que o método 2;

b. Caso bilateral: para verificar se existe diferença na precisão dos dois métodos.

TABELA VI
Comparação de variâncias

Hipótese nula	Hipótese alternativa	Rejeita-se H_0 , se
$H_0: \sigma_1^2 = \sigma_2^2$	$H_1: \sigma_1^2 > \sigma_2^2$ (caso unilateral)	$F_{\text{calc}} > F_{n_1-1, n_2-1, \alpha}$
		$F_{\text{calc}} = S_1^2/S_2^2$
$H_0: \sigma_1^2 = \sigma_2^2$	$H_1: \sigma_1^2 < \sigma_2^2$ (caso unilateral)	$F_{\text{calc}} < F_{n_1-1, n_2-1, \alpha}$
		$F_{\text{calc}} = S_1^2/S_2^2$
$H_0: \sigma_1^2 = \sigma_2^2$	$H_1: \sigma_1^2 \neq \sigma_2^2$ (caso bilateral)	$F_{\text{calc}} > F_{n_1-1, n_2-1, \alpha/2}$
		$F_{\text{calc}} = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)}$

Os dados apresentados a seguir exemplificam a comparação da precisão entre dois métodos pelo caso bilateral, empregando-se a electroforese capilar (EC) e a cromatografia líquida de alta eficiência (CLAE) para a determinação de diclofenaco de sódio em preparações farmacêuticas (comprimidos, amostra comercial) (Aurora Prado et al, 1999):

$$S_1: 1,65 \quad n_1: 10 \quad (\text{EC})$$

$$S_2: 0,37 \quad n_2: 10 \quad (\text{CLAE})$$

Ao nível de 5% de significância ($P=95\%$), pode-se afirmar que as variâncias obtidas pelos dois métodos são iguais?

Sabendo-se que,

$$\text{Hipótese nula} \quad H_0: \sigma_1^2 = \sigma_2^2$$

$$\text{Hipótese alternativa} \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

E, substituindo-se os dados na Equação 10 tem-se:

$$F_{\text{calculado}} = \frac{(1,65)^2}{(0,37)^2} = 19,88$$

$$F_{\text{tabelado}} = F_{n_1-1, n_2-1, \alpha/2} = F_{9;9, 2,5\%} = 4,026$$

Como $F_{\text{calculado}} > F_{\text{tabelado}}$, rejeita-se a hipótese nula, ou seja, as variâncias dos dois métodos (EC e CLAE) não são iguais, e portanto os métodos apresentam diferença significativa na precisão.

2.2.2 Comparação de mais de duas variâncias

Existem dois testes para comparar mais de duas variâncias: o teste de Cochran, empregado quando todas as variâncias têm os mesmos graus de liberdade, e o teste de Bartlett, utilizado quando todas as variâncias têm ou não igual número de graus de liberdade (Anderson, 1987).

2.2.2.1 Teste de Cochran

O teste de Cochran usa a razão das variâncias assim como no teste F , mas de maneira diferente. A razão das variâncias é chamada g estatístico, calculado utilizando-se a Equação 11.

$$g = \frac{\max(S^2)}{S_1^2 + S_2^2 + S_3^2 + \dots + S_k^2} \quad \text{Eq. 11}$$

Onde:

max S^2 = máxima variância.

$S_1^2 + S_2^2 + S_3^2 + \dots + S_k^2$ = variâncias 1,2,3,...k

Comparar o valor g calculado com o valor tabelado ou crítico para um nível de significância de 0,05 ou 0,01 (95,0 ou 99,0% de nível de confiança). Para tanto, entrar com n , (número de observações associadas a cada variância) e k , (número de variâncias a serem comparadas) na tabela dos valores críticos de g (Anderson, 1987).

O teste de hipótese pode ser formulado:

H_0 : Hipótese nula (as variâncias são iguais, não apresentam diferença significativa)

H_1 : Hipótese alternativa (as variâncias não são iguais, apresentam diferença significativa)

Se o valor calculado para g excede o valor tabelado ou crítico, considera-se que as variâncias não são iguais; e a hipótese nula é rejeitada.

Exemplo: Para determinar a pureza de três tipos diferentes de cetonas, amostras de cada uma foram analisadas 11 vezes por um método geral. Os desvios-padrão calculados para cada amostra, com 10 graus de liberdade foram: $S_1 = 0,16$, $S_2 = 0,20$ e $S_3 = 0,14$. Para o nível de 5% de significância, pode-se afirmar que o método apresenta a mesma precisão para as três amostras de cetonas?

Substituindo-se os dados na Equação 11 tem-se:

$$g_{calc} = \frac{0,20^2}{0,16^2 + 0,20^2 + 0,14^2} = 0,4695$$

$$g_{tabelado} = g_{\alpha, n, k}; \alpha = 5\%; n = 11; k = 3$$

$$g_{tabelado} = 0,6025$$

$$g_{calculado} < g_{tabelado} (0,4695 < 0,6025)$$

Aceita-se a hipótese nula, ou seja, as três variâncias são iguais, portanto, o método apresenta a mesma precisão para as três amostras de cetonas.

2.2.2.2 Teste de Bartlett

O teste Bartlett é similar ao teste F onde a razão de F é obtida e comparada com o valor crítico de F . Os testes se diferenciam na maneira de como as variâncias são manipuladas e como os graus de liberdade são calculados. Para o teste de Bartlett, são utilizadas as seguintes equações:

$$S_p^2 = \sum (n_i - 1)S_i^2 / (N - k) \quad \text{Equação 12}$$

$$M = (N - k) \ln S_p^2 - \sum (n_i - 1) \ln S_i^2 \quad \text{Equação 13}$$

$$df_1 = k - 1 \quad \text{Equação 14}$$

$$A = \frac{1}{3(k-1)} \left(\sum \left(\frac{1}{n_i - 1} \right) - \frac{1}{N - k} \right) \quad \text{Equação 15}$$

$$df_2 = (k + 1) / A^2 \quad \text{Equação 16}$$

$$b = \frac{df_2}{1 - A + \frac{2}{df_2}} \quad \text{Equação 17}$$

$$F = \frac{df_2 M}{df_1 (b - M)} \quad \text{Equação 18}$$

Onde:

n_i = número de observações associados com cada variância S_i^2

N = soma de todos os n

k = número das variâncias a serem comparadas

df_1 = graus de liberdade do numerador

df_2 = graus de liberdade do denominador.

Advertência: Quando aplicado a dados cujas distribuições não são normais, este teste poderia dar um resultado errôneo de não homogeneidade. Devido a esta sensibilidade com relação à distribuição, alguns estatísticos não recomendam sua aplicação. Este teste é empregado, algumas vezes, quando as distribuições são, pelo menos, aproximadamente normais, pois não existe outro tipo de teste para comprovar-se a homogeneidade de mais de duas variâncias, quando os graus de liberdade são diferentes.

Exemplo: Utilizando-se os mesmos dados do exemplo anterior, pode-se empregar o teste de Bartlett para comparar os três desvios-padrão. O teste de Bartlett pode ser utilizado para comparar variâncias que tenham graus de liberdade iguais ou diferentes.

O teste de hipótese pode ser formulado:

H_0 : Hipótese nula (as variâncias são iguais)

H_1 : Hipótese alternativa (as variâncias não são iguais)

Substituindo-se os dados nas Equações 12 a 18 obtém-se:

$$S_1 = 0,16, 10 \text{ df}; S_2 = 0,20, 10 \text{df}; S_3 = 0,10, 10 \text{df}$$

$$n_1 = n_2 = n_3 = 11$$

$$N = 33$$

$$k = 3$$

$$S_p^2 = \frac{((11-1)0,16^2) + ((11-1)0,20^2) + ((11-1)0,14^2)}{33-3} = \frac{0,8520}{30} = 0,0284$$

$$M = (33-3) \ln 0,02840 - ((11-1) \ln 0,16^2 + (11-1) \ln 0,20^2 + (11-1) \ln 0,14^2)$$

$$= -106,84098 - (-36,65163) - 32,18876 - 39,32226 = 1,3217$$

$$A = \frac{1}{3(3-1)} \left(\frac{1}{11-1} + \frac{1}{11-1} + \frac{1}{11-1} - \frac{1}{33-3} \right) = 0,0444$$

$$df_1 = 3-1 = 2$$

$$df_2 = \frac{3+1}{0,0444^2} = 2029$$

$$b = \frac{2029}{1 - 0,0444 + \frac{2}{2029}} = 2121,0854$$

$$F = \frac{2029(1,3217)}{2(2121,0854 - 1,3217)} = 0,6326$$

$$F_{calc.} = 0,6326$$

$$F_{\text{tabulado}} = F_{\alpha, df_1, df_2}; \alpha = 5\%; df_1 = 2; df_2 = 2029$$

$$F_{\text{tabulado}} = F_{0,05,2,2029} = 3,00$$

$$F_{\text{calculado}} < F_{\text{tabulado}} (0,6326 < 3,00)$$

Aceita-se a hipótese nula, ou seja, as variâncias são iguais, e portanto não apresentam diferença significativa. Esta conclusão também é obtida com o teste de Cochran. (Nota: Da tabela dos valores críticos de g , foi utilizada a linha dos graus de liberdade infinito para o denominador, pois o valor 2029 encontra-se entre as linhas 1000 e o infinito.)

CONCLUSÃO

Neste estudo foi ilustrada a aplicação da estatística como ferramenta básica na comparação de métodos analíticos na análise química.

Foram utilizados os testes de significância estatística para informar se a qualidade de um lote de material, a produção de certo tipo de produto ou a batelada de itens recebidos do fornecedor diferem "significativamente" de um valor padrão ou da qualidade de um segundo ou mais lotes ou fontes. Ao realizar-se o teste de significância comprova-se a veracidade de uma hipótese, a qual será considerada válida até prova em contrário.

Os testes de significância mais importantes e utilizados neste trabalho foram o teste t , e o teste F . O teste t foi utilizado para comparar a exatidão de dois métodos e o teste F , foi utilizado para comparar a precisão de duas metodologias. Foi também utilizada a análise da variância (ANOVA) para comparar a exatidão de mais de duas médias ou metodologias, e os testes de Cochran e de Bartlett para comparar a precisão de mais de duas variâncias ou metodologias.

A exatidão e a precisão são os critérios mais importantes para verificar se um método é apropriado para a avaliação de uma metodologia em particular e se os dados obtidos são aceitáveis. Estes dois critérios juntos permitem determinar o erro de uma medida analítica e são utilizados quando se está avaliando a validade de um método analítico.

AGRADECIMENTOS

À FAPESP e à CNPq pelo auxílio financeiro.

REFERÊNCIAS

1. Anderson, R. L. *Practical statistics for analytical chemists*. New York: Van Nostrand Reinhold. p.17, 51-66, 1987.
2. Aurora Prado, M.S.; Steppe, M.; Tavares, M. F. M.; Kedor-Hackmann, E.R.M.; Santoro, M. I. R. M. Comparison of capillary electrophoresis and liquid chromatography for determination of diclofenac sodium in pharmaceuticals. *J. Cap. Elec. and Microchip Tech.* 6:125-129, 1999.
3. Bolton, S. *Pharmaceutical statistics. Practical and clinical applications*. 2nd ed. New York: Marcel Dekker Inc., pp.27-30, 118-181, 1990.
4. Bruce, P.; Minkinen, P.; Riekkola, M. L. Practical method validation: validation sufficient for an analytical method. *Mikrochim. Acta*, Wien, 128:93-106, 1998.
5. Caulcutt, R.; Boddy, R. *Statistics for analytical chemists*. London: Chapman & Hall, p.65-78, 1994.
6. Chui, Q.S.H.; Nagamine, R.; Olivieri, J. C. *Curso de estatística básica aplicada à análise química*. Instituto de Pesquisas Tecnológicas. Divisão de Química, São Paulo. p. 32-52, 1997.
7. Christian, G. D. *Analytical chemistry*. New York: John Wiley & Sons, Inc. p.38-43, 1994.
8. Christmann, R. U. *Estatística aplicada*. São Paulo: Edgard Blücher. p.58-61, 113-118, 1978.
9. Costa Neto, P. L. *Estatística*. São Paulo: Edgard Blücher Ltda. p.84-86; 115-120, 152-167, 1987.
10. Da Silva Guedes, M. L.; Da Silva Guedes, J. *Bioestatística para profissionais da Saúde*. 1^a ed. Rio de Janeiro: Ao Livro Técnico S.A. p.126-131, 1988.
11. Feigenbaum, A. V. *Controle de qualidade total. Métodos estatísticos aplicados à qualidade*. Brasil: Makron. p.288-289, 1994.
12. Hadjiannou, T. P.; Christian, G. D.; Efsthathiou, C. E.; Nikolettis, D. P. *Problem solving in analytical chemistry*. Oxford: Pergamon Press. p.13-18, 1988.
13. Karnes, H. T.; Shiu, G.; Shah, V. P. Validation of bioanalytical methods. *Pharm. Res.*, New York, 8(4):421-426, 1993.
14. Karnes, T.; March, C. Precision, accuracy, and data acceptance criteria in biopharmaceutical analysis. *Pharm. Res.*, New York, 10(10):1421-1426, 1993.
15. Kateman, G.; Pijpers, F. W. *Quality control in analytical chemistry*. New York: John Wiley & Sons. p.90-98, 1981.
16. Lee, K. R.; McAllister, P. R. Helping analytical scientists apply statistics. *Drug Dev. Ind. Pharm.*, 22(9/10):891-908, 1996.
17. Miller, J. C.; Miller, J. N. *Estatística para química analítica*. 2^a ed. Estados Unidos: Addison-Wesley Iberoamericana, S.A. p.40-57, 1993.
18. Ohlweiler, O. A. *Química analítica quantitativa*. 2^a ed. Rio de Janeiro: Livros Técnicos e Científicos. v.1, p.346-358, 1976.
19. Skoog, D. A.; West, D. M.; Holler, F. J. *Fundamentals of analytical chemistry*. 7th ed. New York: Saunders College Publishing. p.53-57, 1996.
20. Snyder, L. L. R.; Kirkland, J. J.; Glajch, J. L. *Practical HPLC method development*. 2nd ed. John Wiley, New York. pp.687-689, 1997.
21. Steppe, M. *Estudo da estabilidade térmica da cefalexina em suspensão oral pediátrica*. Tese para obtenção de grau de Doutor, 2000.
22. Swartz, M. L.; Krull, I. S. Validação de métodos cromatográficos. *Pharm. Tech.* 2(3):12-20, 1998.

Endereço para correspondência

Erika R. M. Kedor-Hackmann

E-mail: ermkedor@usp.br