ISSN 0871-3413 • ©*ArquiMed*, 2006

Desvio Padrão ou Erro Padrão

Nuno Lunet, Milton Severo, Henrique Barros
Serviço de Higiene e Epidemiologia da Faculdade de Medicina da Universidade do Porto

A distinção fundamental entre estatística descritiva e inferência estatística pode ser ilustrada pelo diferente significado dos termos desvio padrão e erro padrão. Contudo, o uso do erro padrão da média para descrever a variabilidade das observações numa amostra é dos erros mais frequentemente observados na literatura médica (1), muitas vezes por desconhecimento de princípios básicos da estatística. Neste texto procuraremos definir estas duas medidas, colocando em evidência as suas diferenças e as situações em que cada uma delas pode ser, adequadamente, utilizada.

DESVIO PADRÃO

O desvio padrão é uma medida de dispersão e o seu valor reflecte a variabilidade das observações em relação à média.

A dispersão das observações que constituem uma amostra pode ser caracterizada pelos desvios de cada observação em relação à média $(\chi_i - \bar{\chi})$, podendo tomar valores positivos ou negativos, e o somatório dos desvios de cada observação em relação à média amostral é zero. Contudo, os desvios ao quadrado$(\chi_i - \bar{\chi})^2$, tomam sempre um valor positivo, e a respectiva média é a variância da amostra. Se existir uma grande dispersão das observações a variância é grande. Se os valores de cada uma das observações forem próximos da média a variância é pequena.

Uma vez que a variância é obtida a partir dos quadrados dos desvios, esta exprime-se na unidade da variável ao quadrado (*e.g.* se as observações tiverem "cm" como unidade, a variância exprime-se em "cm^2"). O desvio padrão é a raiz quadrada da variância (fórmula 1), pelo que as suas unidades são as mesmas da média da variável. O cálculo do desvio padrão é exemplificado no anexo 1.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i}(x_i - \bar{x})^2} \qquad (1)$$

A magnitude do desvio padrão depende da dispersão das observações relativamente à média, não variando com o aumento do tamanho das amostras.

Quando a variável segue uma distribuição normal, o desvio padrão fornece uma informação adicional acerca da forma como as observações se distribuem em torno da média, cerca de 68,2% das observações estão contidas no intervalo definido por média ±1 desvio padrão, 95,4% no intervalo média ± 2 desvios padrão e 99,7% no intervalo média ± 3 desvios padrão.

O desvio padrão, para além de sumariar a informação relativa à dispersão das observações relativamente à média amostral, é uma estimativa da dispersão na população de que a amostra é proveniente. Contudo, esta estimativa é sistematicamente inferior ao valor real do desvio padrão da população, principalmente nas amostras pequenas, pelo que é habitualmente calculado o desvio padrão corrigido (fórmula 2), que não apresenta o referido erro sistemático (2).

$$\sigma_{n-1} = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2} \qquad (2)$$

ERRO PADRÃO

Quando extraímos uma amostra aleatória da população e calculamos o valor médio de uma determinada variável, o objectivo último é inferir sobre a média da população de onde a amostra é originária, ou seja, a média na amostra avaliada é uma estimativa da média na população, cuja precisão depende da dispersão da população e do tamanho da amostra.

Se várias amostras aleatórias forem obtidas de uma dada população, elas vão diferir relativamente ao valor médio da população em cada uma e, à semelhança do que acontece com as observações de cada amostra individualmente, a distribuição das médias amostrais tem também um desvio padrão. O erro padrão da média de uma amostra é uma estimativa do desvio padrão da distribuição das médias de amostras com o mesmo tamanho obtidas da mesma população, e dessa forma uma medida da incerteza associada à estimativa da média na população (anexo 2).

No caso do erro padrão da média, este é obtido dividindo o desvio padrão da amostra pela raiz quadrada do número de observações na amostra.

O erro padrão da estimativa diminui com o aumento do tamanho da amostra, reflectindo o aumento de precisão da estimativa com o tamanho da amostra (anexo 2).

QUANDO UTILIZAR O DESVIO PADRÃO E O ERRO PADRÃO DA MÉDIA?

Se o objectivo é descrever a variabilidade observada numa amostra deve-se utilizar o desvio padrão.

O desvio padrão, como medida de dispersão, não deve ser usado quando a população não segue uma distribuição normal ou aproximadamente normal. Nestes casos, o desvio padrão pode não ser uma boa estimativa de dispersão, pelo facto da média, que é utilizada no seu cálculo, ser pouco resistente a observações extremas. Também quando a distribuição da população é normal podem ocorrer observações extremas se o tamanho das amostras for pequeno.

Nestas situações, poderá ser mais adequada a descrição da dispersão com outras medidas (*e.g.* distância inter-quartis) ou indicando percentis próximos dos dois extremos da distribuição (*e.g.* percentis 25 e 75 ou os percentis 10 e 90).

Se o objectivo for indicar a imprecisão associada à estimativa de um determinado parâmetro (*e.g.* média), pode utilizar-se o erro padrão. Contudo, de uma forma geral, os intervalos de confiança podem ser interpretados de forma mais directa que os erros padrão, sendo preferível a apresentação dos primeiros. O erro padrão é um passo intermédio no cálculo de intervalos de confiança.

O facto do erro padrão ser quantitativamente menor do que o desvio padrão pode contribuir para que alguns autores optem por apresentar o erro padrão quando pretendem quantificar a dispersão das observações da amostra, transmitindo uma falsa ideia de precisão aos leitores menos atentos e com poucos conhecimentos de estatística.

APRESENTAÇÃO DOS RESULTADOS

O desvio padrão e o erro padrão têm as mesmas unidades das medidas das quais resultam.

É frequente (3) a utilização do símbolo ± entre o valor numérico da média e o do respectivo desvio padrão ou erro padrão (*e.g.* 0,350 ± 0,062 g/cm^2) sem que seja indicada qual a quantidade a que o número colocado após o sinal ± se refere. É necessário indicar de forma clara no texto se é apresentada a média e a respectiva medida da incerteza (*e.g.* os resultados apresentados são média (erro padrão)) ou a média e uma medida da dispersão (*e.g.* os resultados apresentados são média (desvio padrão)) (4).

Leitura recomendada
Altman DG, editor. Practical statistics for medical research. Chapman &Hall; 1991.
Bland M, editor. An inroduction to medical statistics, 3ª ed. Oxford University Press; 2000.
Coggon D. Statistics in clinical practice, 2nd ed. BMJ books, 2003.

REFERÊNCIAS

1 - Lunet N, Barros H. Sobre a necessidade da metodologia. Arq Med 2003;17:132-5.
2 - Murteira BJF. Análise exploratória de dados: estatística descritiva. McGraw-Hill, 1993.
3 - Olsen CH. Review of the use of statistics in Infection and Immunity. Infect Immun 2003;71:6689-92.
4 - Altman DG, Bland JM. Standard deviations and standard errors. BMJ 2005;331:903.

Correspondência:
Prof. Nuno Lunet
Serviço de Higiene e Epidemiologia
Faculdade de Medicina da Universidade do Porto
Alameda Prof.. Hernâni Monteiro
4200-319 Porto

e-mail: nlunet@med.up.pt

Anexo 1 - **Cálculo do desvio padrão para os valores do perímetro de cintura (PC) numa amostra de 10 indivíduos.**

	PC (cm) χ_i	observação — média $\chi_i - \bar{\chi}$	(observação — média)^2 $(\chi_i - \bar{\chi})^2$
Observaçã0 1	87,1	87,1-88,1= -1,0	1,0
Observação 2	87,8	87,8-88,1= -0,3	0,1
Observação 3	90,9	90,9-88,1= 2,8	7,8
Observação 4	78,8	78,8-88,1= -9,3	86,5
Observação 5	97,7	97,7-88,1= 9,6	92,2
Observação 6	84,9	84,9-88,1= -3,2	10,2
Observação 7	82,5	82,5-88,1= -5,6	31,4
Observação 8	74,2	74,2+88,1= -13,9	193,2
Observação 9	104,8	104,8-88,1= 16,7	278,9
Observação 10	92,3	92,3-88,1= 4,2	17,6

média

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{1}{10} \times 881$$

$$\bar{x} = 88,1$$

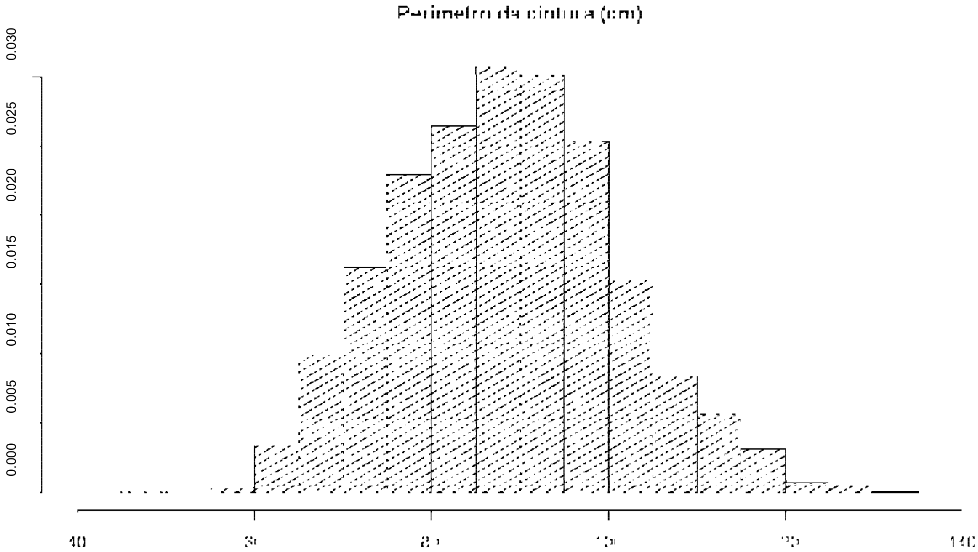
$$\sum_{i}^{n} (x_i - \bar{x})^2 = 718,9$$

A variância é obtida dividindo o somatório dos quadrados das diferenças entre cada observação e o valor médio das observações pelo número de observações (718,9/10=71,9). O desvio padrão é a raiz quadrada da variância √71,9=8,5). Uma vez que o tamanho da amostra é pequeno deve ser calculado o desvio padrão corrigido:

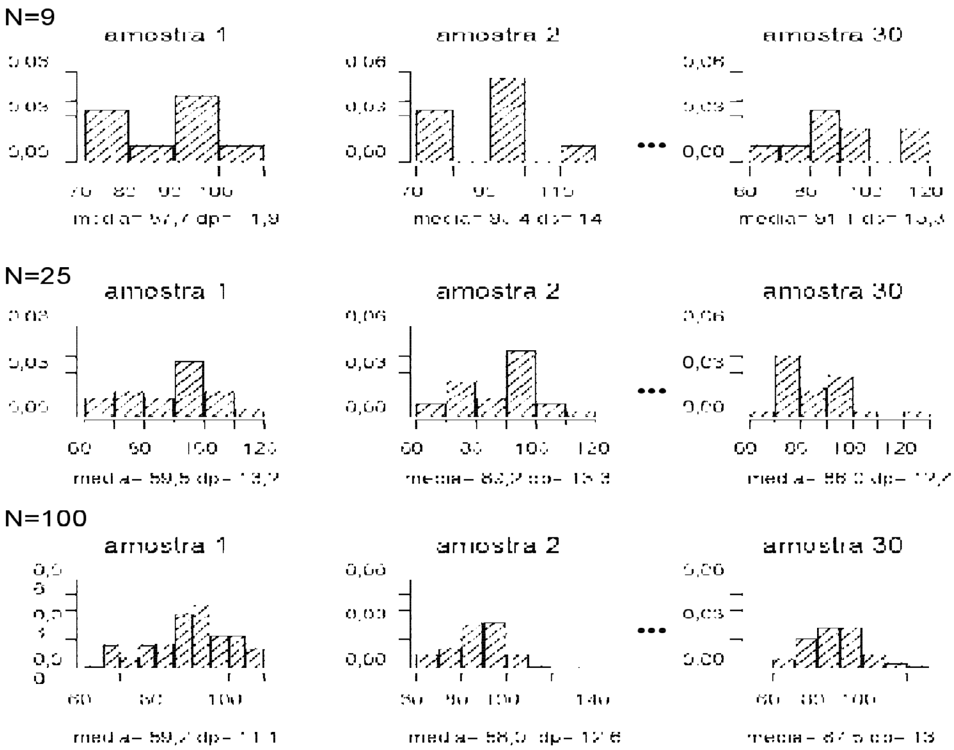
$$\sigma_c = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{10-1}718,9} = \sqrt{\frac{1}{9}718,9} = \sqrt{79,9} = 8,9$$

Anexo 2 - **Cálculo do erro padrão da média em amostras aleatórias com diferente tamanho.**

Considere que a distribuição da variável perímetro da cintura da seguinte população segue uma distribuição normal com média=88,8 cm e desvio padrão (dp)=12,5 cm.



Extraímos 30 amostras com tamanho fixo (9, 25 e 100)



Média para cada amostra de tamanho fixo 9, 25 e 100

Amostras	Média n=9	Média n=25	Média n=100
1	87,7	89,8	89,2
2	90,4	89,2	88
3	92,5	92,8	87,9
4	91,8	87,7	88,2
5	94,1	91,9	89,7
6	94,1	87,7	88,1
7	91,6	86,4	86,4
8	81,5	91,1	88,6
9	88,8	90,7	88,9
10	91,8	87,6	87,2
11	87,9	86,4	88,6
12	95,7	85,2	89,6
13	89,6	81,5	90,2
14	92,0	86,6	88,7
15	90,9	87,8	89,8
16	87,1	88,7	90,3
17	87,2	87,9	88,7
18	86,2	89,1	90,3
19	90,6	88,1	90,8
20	91,0	91,4	89,3
21	88,8	85,3	92,4
22	84,9	86,1	88,8
23	81,0	92,7	88,4
24	98,3	89,4	87,3
25	85,8	84,6	89,2
26	92,9	90,6	88,5
27	88,5	87,9	90,1
28	86,3	86,5	87,9
29	85,8	88,8	89,1
30	91,1	86,0	87,5
Média das medias das amostras	89,5	88,1	88,9
Desvio padrão das medias das amostras	3,9	2,6	1,2

O erro padão é igual ao desvio padrão a dividir pela raiz quadrada de n.

O desvio padrão da população é neste caso conhecido, pelo que podemos calcular o erro padrão para cada tamanho amostral. O erro padrão da média é para n=9

$\frac{12.5}{\sqrt{9}} = 4,2$,

para n=25 $\frac{12,5}{\sqrt{25}} = 2,5$

e para n=100 $\frac{12,5}{\sqrt{100}} = 1,2$

Pode verificar-se que estes valores são muito semelhantes aos dos desvios padrão das médias, apresentando na tabela, ilustrando que o erro padrão da média de uma amostra é uma estimativa do desvio padrão da distribuição das médias de amostras com o mesmo tamanho obtidas da mesma população.
Também se verifica que o erro padrão da estimativa diminui com o aumento da amostra 3,9>2,6>1,2 para respectivamente n=9, 25 e 100, reflectindo o aumento de precisão da estimativa com o tamanho da amostra.