

PROBABILIDADES E ESTATÍSTICA

(LEQ, LEB E LQ)

NOTAS

SOBRE

ERROS E PROPAGAÇÃO DE ERROS

DE

CARLOS DANIEL PAULINO

E

PAULO SOARES

SETEMBRO DE 2004

SECÇÃO DE ESTATÍSTICA E APLICAÇÕES
DEPARTAMENTO DE MATEMÁTICA
INSTITUTO SUPERIOR TÉCNICO

ERROS ACIDENTAIS versus ERROS SISTEMÁTICOS. PRECISÃO versus EXACTIDÃO. PROPAGAÇÃO DE ERROS.

Qualquer processo experimental está sujeito a erros com múltiplas origens cuja eliminação é de todo impossível. Para que o experimentador possa ter uma atitude crítica sobre os resultados da sua análise interessa ter uma noção dos vários tipos de erros e suas causas e do modo como se propagam ao longo das suas operações de cálculo.

1. Definições

Em Estatística, a recolha de dados é encarada como um processo de observação de uma grandeza considerada como uma variável aleatória. A variabilidade presente nos dados recolhidos é assim um reflexo da própria natureza estocástica da grandeza observada. Em muitas situações experimentais pretende-se apenas avaliar o valor desconhecido de uma determinada constante, μ_0 , como, por exemplo, o valor de uma constante física ou a concentração de uma substância numa solução. Neste contexto, as medições experimentais podem também ser encaradas como resultantes da observação de uma variável aleatória, X , agora associada aos erros introduzidos por um qualquer dispositivo experimental. Desta forma, a variabilidade dos resultados experimentais é justificada pela intervenção desse dispositivo experimental. Os erros experimentais classificam-se habitualmente em **erros sistemáticos** e **erros acidentais** que, em geral, ocorrem em simultâneo.

Os **erros sistemáticos** (por vezes também apelidados de erros determinados) são erros que nas mesmas circunstâncias distorcem todas as medições sempre num dado sentido (ou para mais ou para menos) em relação ao seu verdadeiro valor, μ_0 . As suas causas residem em deficiências:

- de instrumentos de medição (aparelhos desregulados, material de vidro mal calibrado, etc) e reagentes (contaminação com impurezas) – erros instrumentais;
- do método usado – erros de método;
- da actuação do próprio operador (falta de prática, inabilidade, etc) – erros pessoais.

Os erros sistemáticos são, em geral, mais graves pois são frequentemente difíceis de detectar e a sua ocorrência pode facilmente passar despercebida. Por isso, um grande

cuidado deve ser colocado na sua detecção e, quando presentes, devem ser corrigidos (ou compensados) ou pelo menos minimizados, o que pode ser feito mediante a realização de ensaios em branco ou ensaios de calibração.

Os **erros acidentais** são erros devidos a causas que não se conhecem exactamente e que são responsáveis por pequenas e irregulares variações nas medições realizadas. Estes erros de carácter fortuito são alternativamente denominados de **erros aleatórios**, justificando que várias medições difiram algo umas das outras, em qualquer dos sentidos. A sua eliminação é naturalmente de todo impossível embora possam ser atenuados à custa de uma maior morosidade e/ou encarecimento dos ensaios.

O efeito dos dois tipos de erros referidos encontra-se ilustrado na Figura 1. A presença de um erro sistemático faz com que o conjunto de valores experimentais se afaste de μ_0 e a acção de outras causas desconhecidas introduz um erro aleatório que provoca a dispersão desses valores em torno de um valor médio $\mu \neq \mu_0$, também desconhecido. Ao erro aleatório atribui-se uma distribuição de probabilidade, sendo usual, na prática, adoptar-se distribuições Normais cuja justificação não é alheia a um importante resultado da Teoria da Probabilidade que dá pelo nome de Teorema Limite Central (aflorado no Cap. 5 do programa).

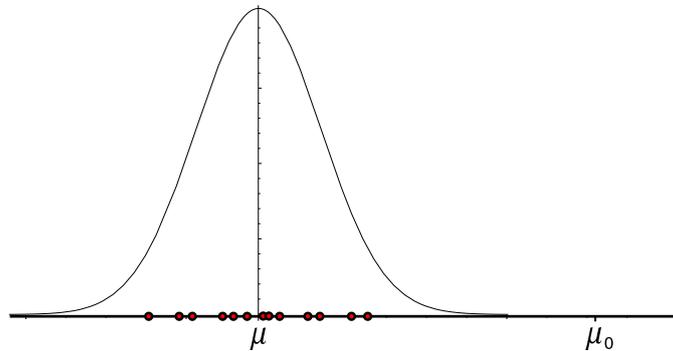


Figura 1: Representação esquemática dos erros sistemáticos e acidentais.

A análise do efeito dos dois tipos de erros supracitados materializa-se na avaliação da **exactidão** e da **precisão** do método experimental. Note-se que, muitas vezes, os termos *exactidão* e *precisão* são erradamente usados como sinónimos na linguagem quotidiana (e em alguma da linguagem técnica).

A chamada **exactidão** (*accuracy*) do método de determinação de μ_0 tem precisamente que ver com a magnitude dos erros sistemáticos no sentido em que reflecte a propriedade de fiabilidade dos resultados da aplicação daquele método. Este será tanto menos inexacto quanto menor for a diferença entre os seus resultados (que reflectem de algum modo o valor esperado de X) e o verdadeiro valor μ_0 .

Como medidas de exactidão têm-se os erros simples, absolutos e relativos. O **erro simples** de uma determinação x é definido por $d = x - \mu_0$, sendo o **erro absoluto** dado por $|x - \mu_0|$. Ambos se exprimem nas mesmas unidades de X (e.g., em % ou g/l). A razoabilidade da sua magnitude depende naturalmente da ordem de grandeza de μ_0 . Por exemplo, na determinação da percentagem de um dado elemento numa dada substância, $|d| = 0.05\%$ será um valor substancial quando $\mu_0 = 0.01\%$ mas já não o é quando $\mu_0 = 60\%$. Daí o recurso ao **erro relativo** $e = |d/\mu_0|$, para $\mu_0 \neq 0$, que é uma medida adimensional habitualmente expressa em % ou em partes por mil. Quando há várias determinações os erros associados ao método podem definir-se pelas expressões correspondentes substituindo x pela média \bar{x} dos resultados dessas determinações.

Nota 1: As expressões acima para os erros pouca utilidade têm em geral pelo facto de μ_0 ser desconhecido. Uma excepção ocorre quando na avaliação de um método analítico se usa uma amostra padrão de características conhecidas. Na prática usam-se valores aproximados obtidos substituindo μ_0 em d pelo valor referente a uma amostra padrão, se disponível, ou pelo valor obtido na amostra corrente por um método comprovadamente bem estabelecido, se existente. Por vezes, μ_0 é substituído no denominador de e pelo próprio x , ou pela média \bar{x} dos resultados de um preferencialmente grande número de determinações, prática tanto menos recomendável quanto mais inexacto se esperar que seja o método.

Note-se que se se substituir em d e e μ_0 por \bar{x} as medidas resultantes já não reflectem a exactidão mas sim a precisão, pois os erros sistemáticos serão eliminados em $x - \bar{x}$ que, por isso, envolve apenas os erros acidentais. ■

Nota 2: Em face do exposto a avaliação da exactidão de um método exige o conhecimento de μ_0 . Uma forma de avaliar em termos aproximados se um método é exacto será vista no Cap. 8 do programa através da formulação da hipótese incisiva de que o valor médio da quantidade X determinada por esse método coincide com o valor μ_0 , e da adopção de um critério de comparação desse valor com a média dos valores experimentais obtidos com esse método (e.g., o teste sobre o valor esperado de uma distribuição Normal). Se esta média empírica estiver significativamente afastada daquele valor hipotético, há indícios de que o método não deverá ser exacto (atente-se, no entanto, que a discrepância entre \bar{x} e o valor postulado para μ poderá também ser devida a erros acidentais, e não só a erros sistemáticos). ■

A chamada **precisão** (*precision*) do método de determinação da quantidade μ_0 (ou precisão de X) está associada com a maior ou menor concordância entre os

resultados, x_i , de várias determinações paralelas de X nas mesmas circunstâncias, traduzindo assim a propriedade de reprodutibilidade dos resultados desse método. Quanto menos (mais) dispersos estiverem estes valores, tanto maior (menor) será a precisão do método.

As medidas mais usadas da precisão de X são o **desvio padrão**, $\sigma = \sqrt{\text{Var}(X)}$ (ou o seu inverso, por vezes denominado **precisão**), como parâmetro de dispersão, e o **coeficiente de variação**, $\sigma/|\mu|$, como parâmetro de dispersão relativa (na literatura química é frequentemente rotulado de **desvio padrão relativo**).

Nota 3: Como estas quantidades são desconhecidas é necessário obter valores aproximados para elas com base em várias determinações de X , sejam eles valores pontuais (e.g., \bar{x} para μ e $s = \sqrt{\sum_i (x_i - \bar{x})^2 / (n - 1)}$ para σ) ou intervalos de valores que as contenham com elevada probabilidade. Uma questão relacionada é calcular o número de determinações de X a efectuar de modo a obter uma dada precisão. Também interessa com frequência comparar dois métodos para determinação de μ_0 do ponto de vista quer das suas precisões quer das suas médias. Tudo isto será (parcialmente) abordado nos Caps. 6 (Estimação pontual), 7 (Estimação por intervalos) e 8 (Testes de hipóteses) do programa. ■

Na Figura 2 ilustra-se a aplicação de 4 métodos experimentais que se distinguem em termos de precisão e exactidão dos seus resultados.

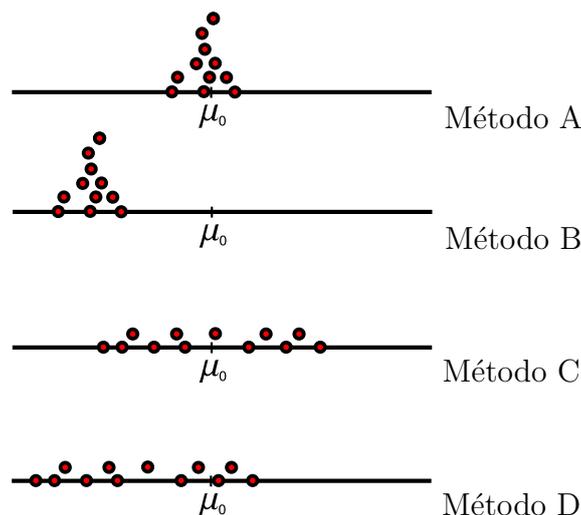


Figura 2: Representação esquemática dos conceitos de precisão e exactidão (A: preciso e exacto; B: preciso e inexacto; C: impreciso e exacto; D: impreciso e inexacto).

Exercício 1¹: Uma amostra padrão de soro sanguíneo contém 42.0g de albumina por litro. Cinco laboratórios fazem cada um 6 determinações da concentração de albumina (em g/l) com os seguintes resultados.

A	42.5	41.6	42.1	41.9	41.1	42.2
B	39.8	43.6	42.1	40.1	43.9	41.9
C	43.5	42.8	43.8	43.1	42.7	43.3
D	35.0	43.0	37.1	40.5	36.8	42.2
E	42.2	41.6	42.0	41.8	42.6	39.0

Comente sobre a precisão e exactidão de cada um deste conjuntos de resultados.

(R: A – preciso e exacto; B – impreciso e exacto; C – preciso e inexacto; D – impreciso e inexacto; E – preciso e exacto se o valor aberrante (*outlier*) for considerado espúrio e, como tal, removido.) ■

2. Propagação de erros acidentais

Num trabalho experimental a quantidade escalar de interesse, agora denotada por Y , é uma variável aleatória medida indirectamente à custa de outra(s) variável(is) aleatória(s). Tem assim importância avaliar (pelo menos, aproximadamente) a precisão na sua medição, ou mais geralmente, a sua distribuição de probabilidade, à custa da correspondente característica das variáveis de que é função.

A determinação exacta destas características de Y apresenta, em geral, grandes dificuldades de ordem analítica. Por exemplo, se $Y = g(X)$ é uma função contínua complicada da variável aleatória contínua X , a avaliação de

$$E(Y) = \int_{-\infty}^{+\infty} g(x)f_X(x)dx$$

está longe, em geral, de ser tranquila. *A fortiori*, o mesmo acontece se Y for uma função complicada de duas ou mais variáveis. Contudo, se Y for uma função matematicamente bem comportada é possível obter expressões analíticas aproximadas para a sua média e variância em particular, sem grande esforço. É isto que se indica nos resultados seguintes.

Proposição 1. Seja X uma variável aleatória com valor médio μ e variância σ^2 e $Y = g(X)$ uma função continuamente diferenciável (pelo menos até à 3^a ordem) em

¹Fonte: Miller and Miller (1993). *Statistics for Analytical Chemistry*. 3rd ed.. Ellis Horwood. Chichester.

$x = \mu$. Então

$$E(Y) \simeq g(\mu) + g''(\mu)\sigma^2/2$$

$$Var(Y) \simeq [g'(\mu)]^2 \sigma^2.$$

Dem. (*esboço*): Expandindo $g(x)$ para qualquer ponto do seu domínio numa série de Taylor de 2^a ordem em torno de μ tem-se

$$Y = g(\mu) + g'(\mu)(X - \mu) + g''(\mu)(X - \mu)^2/2 + R_2$$

onde o resto de ordem 2, R_2 , é uma quantidade desprezável quando comparada com $(x - \mu)$ à medida que $x \rightarrow \mu$. Desprezando esta quantidade e tomando o valor esperado de ambos os membros, obtém-se a 1^a relação.

Considerando agora a expansão de Taylor de 1^a ordem de $g(\cdot)$ tem-se

$$Y = g(\mu) + g'(\mu)(X - \mu) + R_1$$

De novo desprezando o resto, R_1 , de 1^a ordem e aplicando o operador variância a ambos os membros, obtém-se a 2^a relação. ■

Observação: Se se usar a expansão de 1^a ordem para a avaliação de $E(Y)$ obtém-se a aproximação menos precisa $E(Y) \simeq g(\mu)$ que, em certos casos, pode coincidir praticamente com a do resultado anterior.

Exemplo: Seja $\mu = 0.501$ e $\sigma = 0.001$ o valor médio e o desvio padrão da transmitância X de uma solução (razão entre a intensidade da luz transmitida e a intensidade da luz incidente). Então o valor médio e o desvio padrão da absorvância dessa solução, $Y = -\log X$ (log denota logaritmo decimal) são aproximadamente dados por

$$E(Y) \simeq -\log 0.501 + \frac{0.4343}{2 \times 0.501^2} \times 10^{-6} = 0.300 + 0.865 \times 10^{-6} \simeq 0.300$$

$$\sigma(Y) \simeq (0.4343/0.501) \times 10^{-3} = 8.6710^{-3},$$

onde $0.4343 = 1/\ln 10 = \log e$ (note-se que $(\log u)' = \frac{u'}{u \ln 10}$ e $(\log u)'' = \frac{uu'' - (u')^2}{u^2 \ln 10}$).

Como a concentração da substância em questão na referida solução é uma função linear da absorvância, segundo a lei de Lambert-Beer, facilmente se obtém neste pressuposto as expressões da média e da variância da concentração com base em $E(Y)$ e $Var(Y)$. ■

Proposição 2. Seja (X_1, X_2) um par aleatório com valores médios $\mu_i = E(X_i)$, desvios padrões $\sigma_i = \sqrt{Var(X_i)}$, $i = 1, 2$ e covariância $\sigma_{12} = E[(X_1 - \mu_1)(X_2 - \mu_2)]$, e

$Y = g(X_1, X_2)$ uma função continuamente diferenciável (pelo menos até à 3ª ordem) em $(x_1, x_2) = (\mu_1, \mu_2)$. Então, fazendo $\mu = (\mu_1, \mu_2)$

$$E(Y) \simeq g(\mu_1, \mu_2) + \left[\frac{\partial^2 g}{\partial x_1^2} \Big|_{\mu} \right] \frac{\sigma_1^2}{2} + \left[\frac{\partial^2 g}{\partial x_2^2} \Big|_{\mu} \right] \frac{\sigma_2^2}{2} + \left[\frac{\partial^2 g}{\partial x_1 \partial x_2} \Big|_{\mu} \right] \sigma_{12}.$$

$$Var(Y) \simeq \left[\frac{\partial g}{\partial x_1} \Big|_{\mu} \right]^2 \sigma_1^2 + \left[\frac{\partial g}{\partial x_2} \Big|_{\mu} \right]^2 \sigma_2^2 + 2 \left[\frac{\partial g}{\partial x_1} \Big|_{\mu} \right] \left[\frac{\partial g}{\partial x_2} \Big|_{\mu} \right] \sigma_{12},$$

com desaparecimento dos termos envolvendo σ_{12} se X_1 e X_2 forem não correlacionadas.

Dem. (*esboço*): A prova deste resultado segue as linhas da demonstração do resultado anterior, usando os desenvolvimentos de Taylor de 2ª e 1ª ordem para a expressão de $E(Y)$ e $Var(Y)$, respectivamente. Para o 1º caso tem-se então

$$Y \simeq g(\mu_1, \mu_2) + \sum_{i=1}^2 \left[\frac{\partial g}{\partial x_i} \Big|_{\mu} \right] (X_i - \mu_i) + \frac{1}{2} \sum_{i=1}^2 \left[\frac{\partial^2 g}{\partial x_i^2} \Big|_{\mu} \right] (X_i - \mu_i)^2 +$$

$$+ \left[\frac{\partial^2 g}{\partial x_1 \partial x_2} \Big|_{\mu} \right] (X_1 - \mu_1)(X_2 - \mu_2) + R_2,$$

donde se obtém a expressão para $E(Y)$. Adoptando o desenvolvimento

$$Y \simeq g(\mu_1, \mu_2) + \sum_{i=1}^2 \left[\frac{\partial g}{\partial x_i} \Big|_{\mu} \right] (X_i - \mu_i) + R_1,$$

e atendendo à expressão da variância de uma combinação linear de duas variáveis aleatórias, chega-se à expressão para $Var(Y)$. ■

A generalização do resultado anterior traduz-se em:

Proposição 3. Se $Y = g(X_1, X_2, \dots, X_n)$ for uma função continuamente diferenciável (pelo menos até à 3ª ordem) em $(x_1, x_2, \dots, x_n) = (\mu_1, \mu_2, \dots, \mu_n) \equiv \mu$, onde $\mu_i = E(X_i)$, e sendo $\sigma_i^2 = Var(X_i)$, $Cov(X_i, X_j) = \sigma_{ij}$, $i \neq j$, $\forall i, j = 1, \dots, n$, então

$$E(Y) \simeq g(\mu_1, \mu_2, \dots, \mu_n) + \frac{1}{2} \sum_{i=1}^n \left[\frac{\partial^2 g}{\partial x_i^2} \Big|_{\mu} \right] \sigma_i^2 + \sum_{i=1}^n \sum_{j>i}^n \left[\frac{\partial^2 g}{\partial x_i \partial x_j} \Big|_{\mu} \right] \sigma_{ij},$$

$$Var(Y) \simeq \sum_{i=1}^n \left[\frac{\partial g}{\partial x_i} \Big|_{\mu} \right]^2 \sigma_i^2 + 2 \sum_{i=1}^n \sum_{j>i}^n \left[\frac{\partial g}{\partial x_i} \Big|_{\mu} \right] \left[\frac{\partial g}{\partial x_j} \Big|_{\mu} \right] \sigma_{ij},$$

onde as parcelas envolvendo as covariâncias serão nulas se as variáveis X_i forem mutuamente não correlacionadas.

Dem.: Omitida (basta seguir o raciocínio da demonstração da proposição anterior, com base nas expansões de Taylor de uma função de n variáveis reais). ■

Nota 4: Como os parâmetros indicados de X_i são geralmente desconhecidos as fórmulas anteriores são usadas na prática substituindo $\{\mu_i\}, \{\sigma_i^2\}$ e $\{\sigma_{ij}\}$ pelas suas contrapartidas amostrais obtidas das m , digamos, determinações x_{ik} das variáveis X_i , respectivamente dadas por $\bar{x}_i = \sum_k x_{ik}/m$, $s_i^2 = \sum_k (x_{ik} - \bar{x}_i)^2/(m-1)$ e $s_{ij} = \sum_k (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)/(m-1)$, e encarando os correspondentes segundos membros como aproximações da média e da variância das decorrentes determinações $y_k = g(x_{1k}, \dots, x_{nk})$ de Y . As respectivas expressões do desvio padrão (ou da variância ou do coeficiente de variação) exprimem a denominada **lei de propagação (ou da acumulação) dos erros (acidentais)**.

Exercício 2: Demonstre os resultados aproximados abaixo discriminados relativos a funções que surgem frequentemente em análises químicas (a , b e n são constantes):

Y	$E(Y)$	$\sigma(Y)$
$Y = a + bX^n$	$a + b\mu^n + \frac{bn(n-1)}{2}\mu^{n-2}\sigma^2$	$bn\mu^{n-1}\sigma$
$Y = e^{a(X-b)}$	$e^{a(\mu-b)} + \frac{a^2}{2}e^{a(\mu-b)}\sigma^2$	$ae^{a(\mu-b)}\sigma$
$Y = aX_1X_2$	$a(\mu_1\mu_2 + \sigma_{12})$	$a\sqrt{\mu_2^2\sigma_1^2 + \mu_1^2\sigma_2^2 + 2\mu_1\mu_2\sigma_{12}}$
$Y = aX_1/X_2$ *	$a\left(\frac{\mu_1}{\mu_2} + \frac{\sigma_2^2\mu_1}{\mu_2^3} - \frac{\sigma_{12}}{\mu_2^2}\right)$	$\frac{a}{\mu_2}\sqrt{\sigma_1^2 + \frac{\sigma_2\mu_1^2}{\mu_2^2} - 2\frac{\sigma_{12}\mu_1}{\mu_2}}$
$Y = aX_1X_2/(X_3X_4)$ **	$a\frac{\mu_1\mu_2}{\mu_3\mu_4}$	$a\frac{\mu_1\mu_2}{\mu_3\mu_4}\sqrt{\sum_{i=1}^4(\sigma_i/\mu_i)^2}$

* ($\mu_2 > 0$); ** ($\{X_i\}$ indep., $\{\mu_i > 0\}$) ■

Quando se reconhece a existência de erros sistemáticos e, não sendo possível eliminá-los, se consegue majorar alguma das medidas de exactidão mencionadas, então também se pode analisar a propagação desses erros ao longo dos vários cálculos mas a sua acumulação segue regras diferentes pelo facto de estes erros ocorrerem num sentido definido (e conhecido), enquanto os erros aleatórios podem neutralizar-se numa certa medida. Com efeito, se as determinações de X_1 e X_2 forem afectadas de um erro sistemático de $+1$, o erro sistemático acumulado em $Y = X_1 + X_2$ será de $+2$. Porém, se X_1 e X_2 forem afectadas de um erro aleatório de ± 1 , o erro aleatório acumulado em Y não é ± 2 porque poderá haver determinações de X_1 e X_2 com erros aleatórios de $+1$ e -1 , respectivamente (ou vice-versa), resultando no cancelamento do erro na correspondente avaliação de Y .

Devido à não aleatoriedade dos erros sistemáticos, as regras da sua propagação não serão aqui abordadas, pelo que se remetem os interessados neste tópico para livros de Análise Química Quantitativa, por exemplo.