Fundamentos de Análise de Dados

para cursos de Ciências Exactas e Engeharia

José Mariano
Departamento de Física - FCT
Universidade do Algarve
jmariano@ualg.pt

Última revisão: Janeiro de 2024

Conteúdo

1 Introdução

É um facto de observação corrente que, se repetirmos a medição de uma mesma grandeza física em condições supostas idênticas, não obtemos sempre o mesmo resultado, mas sim um conjunto de valores diferentes. Cada um destes valores representa um *valor medido* da referida grandeza, e torna-se evidente que não se pode esperar que o valor medido represente o seu valor verdadeiro (exacto). Nenhuma medição é exacta. As medidas de massa, comprimento, tempo, e todas as propriedades derivadas como o volume, densidade, força, energia, são inevitavelmente de precisão limitada. Nestas condições, a crítica dos resultados obtidos numa experiência é parte fundamental da própria experiência.

Essa "falta de perfeição" é designada, actualmente, por *incerteza*. A palavra "erro", que durante largos anos foi utilizada com esse mesmo significado, está hoje em dia reservada para designar o afastamento entre o valor obtido numa medição e o correspondente valor verdadeiro, o qual é, em geral, desconhecido.

Ao realizar uma medição, não basta indicar o número que se obteve como resultado: é necessário fazê-lo acompanhar de um outro que indique em que medida o experimentador está certo do valor que apresenta. Pode-se assim afirmar que o objectivo de uma experiência de laboratório é determinar o valor de uma ou mais quantidades físicas, *o(s) mesurando(s)* (X), através de um procedimento adequado. De maneira geral, o resultado da medida (x) é apenas uma aproximação, ou estimativa, do valor do mesurando, sendo portanto acompanhado de uma determinada *incerteza* (u) nessa medição.

De maneira geral, o resultado do processo de medição deve ser reportado na forma:

$$X = (x \pm u) \text{ unidades.}$$

Por exemplo, o resultado da medição de um comprimento deve ser reportado na forma

$$L = (256,07 \pm 0,02) \text{ mm.}$$

2 Incerteza

Todas as medições são afectadas por um erro experimental devido às inevitáveis imperfeições nos aparelhos de medida ou às limitações impostas pelos nossos sentidos (visão, audição, etc) que registam a informação, o que se traduz na existência de uma *incerteza* associada à medida. A incerteza numa medida reflecte a falta de conhecimento do valor exacto (ou real) do mesurando. O resultado de uma medida é sempre uma estimativa do valor do mesurando devido à incerteza introduzida pelos erros. A preocupação fundamental do experimentador que realiza uma medição é, naturalmente, é identificar as fontes de erro que podem afectar o processo de medição, e quantificar essas fontes de erro.

2.1 Conceitos clássicos: exactidão e precisão

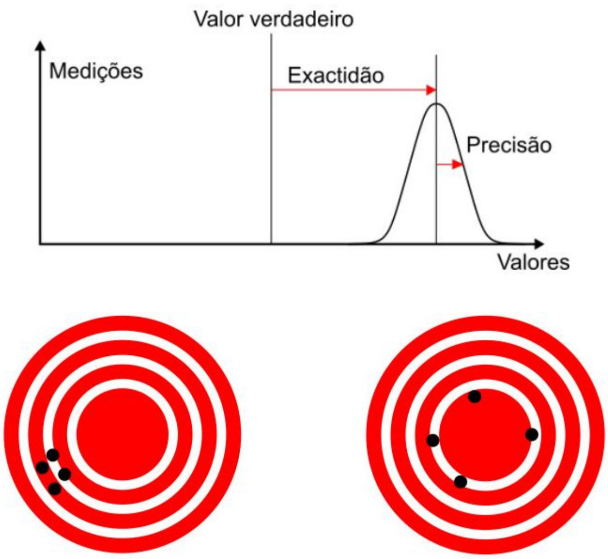

Figura 1: Exactidão vs precisão.

Tem sido prática corrente a utilização dos conceitos de *exactidão* e *precisão* para caracterizar o grau de rigor com que uma medição é efectuada. Entende-se por *exactidão* a maior ou menor aproximação entre o resultado obtido e o valor verdadeiro. A *precisão* está associada à dispersão dos valores resultantes da repetição das medições. Se se fizer uma analogia com o disparo de um projéctil contra um alvo, naquilo que se chama tiro de grupamento, pode-se dizer que a *exactidão* corresponde a acertar no (ou próximo do) centro do alvo, enquanto que se fala de *precisão* quando os vários disparos conduzirem a acertar em pontos muito próximos entre si. A figura 1 procura ilustrar estes conceitos.

2.2 Erros de medição

Consideram-se normalmente três tipos de erros: os *erros grosseiros*, os *erros sistemáticos* e os *erros aleatórios* ou acidentais.

Os erros grosseiros são devidos à falta de atenção, pouco treino ou falta de perícia do operador. Por exemplo, uma troca de algarismos ao registar um valor lido. São geralmente fáceis de detectar e eliminar.

Os erros sistemáticos são devidos a defeitos constantes do método escolhido, do instrumento, ou da pessoa que faz a medição. São erros que se reproduzem sempre nas mesas condições, afectando o resultado sempre no mesmo sentido.

Os erros aleatórios, são devidos a causa fortuitas de que não se tem perfeito conhecimento, tanto podem ser por defeito como por excesso, e têm tendência a anular-se quando se faz uma média. Pode-se ser aplicado um cálculo estatístico da teoria das probabilidades.

Os erros sistemáticos podem eliminar-se, uma vez conhecida a sua causa. Para isso é necessário fazer variar as condições da experiência, o método, o observador. Os erros acidentais não podem eliminar-se: mas atenuam-se os seu efeitos, aumentando o número de medições e pode-se avaliar a sua magnitude.

Normalmente consideram-se dois tipos de incertezas, consoante a forma como estas são avaliadas: *incertezas do tipo A* e *incertezas do tipo B*, consoante o número de medidas que se realiza.

2.3 Incerteza do tipo A

Incertezas do tipo A são as que são obtidas a partir de várias medidas do mesurando, que se supõe ser uma variável aleatória de variância σ^2. Neste caso, o valor mais provável do mesurando é estimado através da *média amostral*:

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

em que x_i é o valor individual de cada uma das n medidas efectuadas. Pode-se portanto afirmar que:

$$X = \bar{x}$$

Para se estimar a incerteza u, faz-se uso da *variância amostral*, que estima a variância téorica σ^2. A variância amostral s^2 é dada por :

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2}$$

vindo expressa nas mesmas unidades que x_i^2.

O *desvio padrão amostral* ou *desvio padrão experimental* s, estimativa do desvio padrão teórico, define-se como:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3}$$

vindo expresso nas mesmas unidades que x_i. Se s tiver um valor pequeno em relação ao valor médio $\bar{x}$, isso significa que os dados estão concentrados em torno de $\bar{x}$ e a precisão da medida é elevada, isto é, a incerteza é baixa.

A variância do valor médio de x, $\sigma^2(\bar{x})$ é melhor estimada pela *variância experimental da média* s_m^2:

$$s_m^2 = \frac{s^2}{n} = \frac{1}{n(n-1)}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4}$$

Esta expressão é tanto mais certa quanto maior for a dimensão da amostra. No entanto, adopta-se por convenção que pode ser utilizada quando $N \geq 10$.

Pode-se então dizer que, para incertezas do tipo A, se tem:

$$u = s_m = \sqrt{\frac{s^2}{n}}$$

em que s_m é chamado o *desvio padrão da média*[1].

[1] s_m pode ser calculado no Office Excel e LibreOffice Calc usando a função **STDEV.S** (**DESVPAD.S**, se o Excel estiver em Português) dividida por $\sqrt{n}$. Consulte as respectivas páginas de ajuda.

2.4 Incerteza do tipo B

Quando uma grandeza X não é obtida através de várias observações, mas de apenas uma, a melhor estimativa para o valor do mesurando é o resultado da medida, isto é, $X = x$, sendo a sua incerteza padrão avaliada com base no conhecimento prévio da possível variabilidade de X. Por exemplo, no caso de uma única leitura de uma régua, a incerteza padrão é normalmente tomada como igual a *metade da menor divisão da escala.* No caso de uma régua graduada em milímetros, isto significa que $u = 0,5$ mm. A esta incerteza, relacionada com o intrumento de medida, chama-se usualmente *incerteza de leitura.* No caso de um aparelho digital, a incerteza de leitura é normelmente considerada igual à menor divisão avaliável da escala.

2.5 Incerteza padrão combinada

Considere-se um mesurando Y que não é medido directamente, mas calculado a partir de n outras quantidades $X_1, X_2, \ldots, X_n$ através de uma relação funcional f, muitas vezes designada *equação de medida*[2].

$$Y = f(X_1, X_2, \ldots, X_n) \tag{5}$$

A estimativa do mesurando Y, designada por y, é obtida a partir da equação 5, recorrendo a n estimativas $x_1, x_2, \ldots, x_n$ das n quantidades $X_1, X_2, \ldots, X_n$. Portanto, a estimativa de y, que é o resultado da medição, é dada por:

$$y = f(x_1, x_2, \ldots, x_n) \tag{6}$$

A incerteza padrão de y, a que se chama *incerteza padrão combinada* $u_c(y)$, é dada pela raiz quadrada positiva da variância estimada de y, $u_c^2(y)$, obtida a partir da expressão

$$u_c^2(y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) \tag{7}$$

isto é,

$$u_c(y) = \sqrt{\left(\frac{\partial f(x_1, x_2 \ldots x_n)}{\partial x_1}\right)^2 u(x_1)^2 + \left(\frac{\partial f(x_1, x_2 \ldots x_n)}{\partial x_2}\right)^2 u(x_2)^2 + \ldots + \left(\frac{\partial f(x_1, x_2 \ldots x_n)}{\partial x_n}\right)^2 u(x_n)^2} \tag{8}$$

A esta última expressão chama-se *lei de propagação das incertezas.*

2.5.1 Formas simplificadas

A eq. 7 reduz-se a formas mais simples em casos de interesse prático. Por exemplo:

Soma ou subtracção Soma ou subtracção de quantidades X_i multiplicadas por constantes a_i:

$$y = a_1 x_1 \pm a_2 x_2 \pm \ldots a_n x_n$$

$$u_c^2(y) = a_1^2 u^2(x_1) + a_2^2 u^2(x_2) + \ldots a_n^2 u^2(x_n)$$

Atenção: As incertezas nunca se subtraem!

Multiplicação Multiplicação de quantidades X_i:

$$y = x_1 \times x_2$$

$$u_c(y)^2 = x_2^2 . u^2(x_1) + x_1^2 . u^2(x_2)$$

[2]Veja-se, por exemplo, o caso do volume de um paralelepípedo, determinado a partir das dimensões das arestas, sendo a relação funcional $V = l_1.l_2.l_3$, em que l_1, l_2 e l_3 são as dimensões das arestas.

Divisão Divisão de quantidades X_i:

$$y = \frac{x_1}{x_2}$$

$$u_c^2(y) = \left(\frac{u(x_1)}{x_2}\right)^2 + \left(\frac{x_1.u(x_2)}{x_2^2}\right)^2$$

Potências Produto de quantidades X_i, elevadas às potências $a, b, \ldots p$ multiplicadas por uma constante A:

$$y = Ax_1^a x_2^b \ldots x_n^p$$

$$u_c^2(y) = a^2u^2(x_1) + b^2u^2(x_2) + \ldots p^2u^2(x_n)$$

No caso mais geral, fórmulas complexas tem que ser avaliadas por utilização directa da Eq. 7 ou pela sua decomposição numa sequência dos casos mais simples como os anteriores.

Exemplo: No caso simples de um quadrado $y = x^2$, $u_c^2(y) = 2^2u^2(x) \Rightarrow u_c(y) = 2u(x)$

2.6 Como reportar os resultados

A incerteza é normalmente expressa com dois algarismos significativos (vêr [1] capítulo 7). A estimativa do mesurando é reportada com o número de casa decimais iguais à da incerteza. Exemplo:

$$m_S = (100,021 \pm 0,035)g$$

Este resultado significa isto que, dadas as condições em que foi efectuada a medição, o experimentador considera como provável que a massa tenha um valor qualquer compreendido entre 99,986 g e 100,056 g.

3 Histogramas

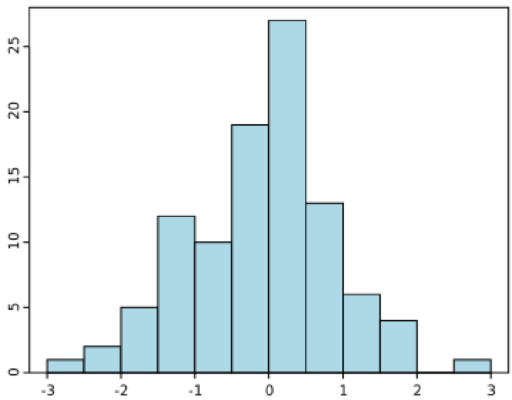

Figura 2: Exemplo de um histograma com 6 classes, entre -3 e +3.

Sabe-se já que, em virtude dos erros aleatórios, se se repetir a medição de uma mesma grandeza física em condições supostas idênticas, não se obtém sempre o mesmo resultado, mas sim um conjunto de resultados diferentes. Sabe-se também que na maioria das situações, numa série de medidas da mesma grandeza, os desvio das medidas em relação ao valor médio têm um distribuição de probabilidades do tipo Normal ou Gaussiana. Por outro lado, as expressões empregues para estimar a incerteza nas medidas assumem para estas este tipo de distribuição. Assim, um processo que permita estimar a distribuição de probabilidades é útil porque permite aferir se os resultados da experiência são "bem comportados" e se se pode utilizar as expressões para estimar as incertezas.

Um processo gráfico de exprimir os diferentes resultados obtidos consiste em desenhar um *histograma*. Para construir um histograma procede-se do seguinte modo:

1. Marcam-se no eixo da abcissa os valores máximo e mínimo das leituras obtidas;

2. Divide-se o intervalo obtido num número arbitrário de subintervalos iguais;

3. Tendo por base cada um destes subintervalos constroem-se rectângulos cujas alturas sejam proporcionais ao número de vezes que se obteve uma leitura de valor compreendido no subintervalo em causa. Considera-se que um valor pertence a um determinado intervalo se for igual ou maior que o estremo esquerdo do intervalo e menor que o estremo direito do referido intervalo.

4 Traçado de Gráficos

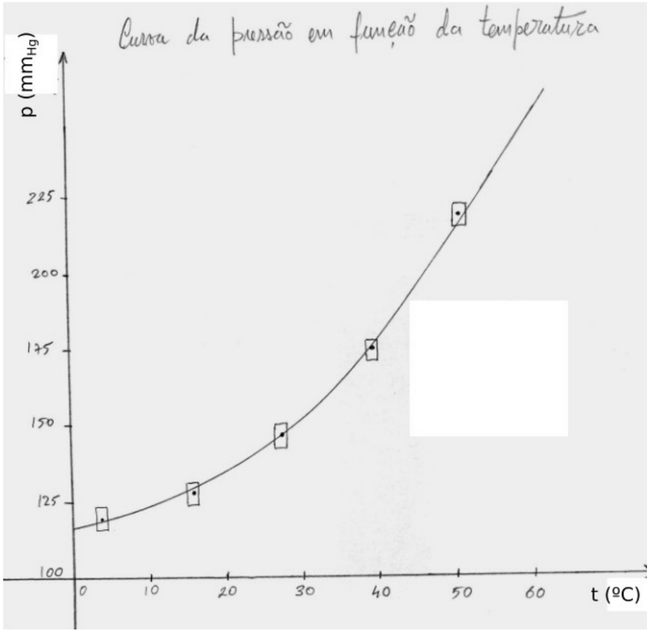

Figura 3: Exemplo de um gráfico corretamente feito à mão. Note-se a marcação das barras de erro horizontais e verticais formando um retângulo.

Ao pretender-se tirar conclusões, de natureza qualitativa e/ou quantitativa sobre a dependência relativa das duas grandezas, há em geral o maior interesse em traduzir os resultados numéricos de que se disponha sob a forma de gráficos. Com efeito, a representação gráfica dos valores experimentais (ou calculados, eventualmente), além de evidenciar os aspectos particulares da dependência entre as grandezas com maior nitidez do que o correspondente conjunto de valores numéricos, possibilita uma análise numérica rápida e relativamente precisa de muitos problemas.

A informação que se pode obter de um gráfico é tanto mais completa e significativa quanto mais funcional e objectivo for o gráfico. Para conseguir que um gráfico desempenhe convenientemente a sua finalidade, torna-se necessário seguir determinadas normas. Seguidamente apresentam-se as regras mais importantes a respeitar no traçado de gráficos. Estas regras devem ser utilizadas também quando o gráfigo é produzido por um programa de computador. Um exemplo de um gráfico bem traçado é apresentado na Fig. 3.

4.1 Normas

Suponhamos que se tem um conjunto de valores numéricos respeitantes à variação de y com x (x é a variável independente e y é a variável dependente), e que se pretende representá-los graficamente. As normas gerais mais importantes a que se deve obedecer no traçado do gráfico são as seguintes:

1. É conveniente marcar os valores de x em abcissas e os valores de y em ordenadas. Junto dos respectivos eixos deve caracterizar-se as grandezas em causa (mediante uma palavra ou conjunto de palavras e/ou símbolo da grandeza) e deve indicar-se também as unidades em que estão expressas essas grandezas.

2. As escalas devem ser escolhidas de acordo com a gama de valores das variáveis, sem esquecer no entanto o que se pretende com o gráfico (por exemplo, as escalas lineares não têm que começar necessariamente em zero, mas se se pretender verificar se os pontos experimentais definem uma linha passando pela origem do referencial, isto é, pelo ponto $(0, 0)$, então é óbvio que este ponto deve figurar no gráfico). Deve ainda estabelecer-se um compromisso entre o número de algarismos a considerar na marcação dos pontos experimentais e o tamanho do gráfico. Por outro lado, a escolha das escalas deve ser feita de modo a permitir uma leitura directa fácil dos valores.

3. Nos eixos deve indicar-se exclusivamente os valores que caracterizam a escala (não se deve jamais indicar nos eixos os valores numéricos dos pontos a marcar, nem tão pouco desenhar as linhas em cujo cruzamento se situa o ponto experimental a assinalar).

4. Para marcar um par de valores (x, y) num gráfico, basta assinalá-lo mediante um pequeno símbolo (cruz, circunferência, quadrado, triângulo, etc.). Tornando-se necessário traçar mais do que uma curva num mesmo gráfico, os pontos de cada conjunto de valores numéricos devem ser assinalados com símbolo diferentes.

5. Todo o gráfico deve ter uma legenda que o identifique e esclareçaa completamente (neste particular, é preferível pecar por excesso de pormenores do que por defeito...). É habitual colocar a legenda sob o eixo das abcissas ou num espaço (suficientemente) livre do próprio gráfico.

Atenção: Não se deve construir uma curva ligando os diferentes pontos experimentais por segmentos de recta. A linha quebrada obtida não teria significado físico dado que as funções normalmente representadas têm variações suaves (derivadas finitas e contínuas).

4.2 Tipos de papel. Linearização de gráficos

No parágrafo anterior chamou-se a atenção para a conveniência de escolher as escalas em função da gama dos valores numéricos a representar graficamente. Isto pressupõe complementarmente uma escolha prévia do tipo de papel mais adequado ao traçado do gráfico em causa: com duas escalas lineares (*papel milimétrico*), com uma escala logarítmica e outra linear (*papel semilog ou log-lin*), com ambas as escalas logarítmicas (*papel log-log*), etc, quando o gráfico é traçado à mão, ou da escolha de tipo de eixos, quando é gráfico é traçado no computador. As folhas de papel gráfico podem ter vários formatos: A5, A4 , A3, etc. As escalas logarítmicas podem ter várias décadas, completas ou não. De notar que uma escala logarítmica nunca pode começar em zero, pois $\log 0 = \ln 0 = -\infty$.

O tipo de papel de gráfico que se usa frequentemente é o papel milimétrico, mas em certos casos convém utilizar outros. Vejamos alguns casos típicos, a título de exemplo.

4.2.1 Papel semi-logarítmico

Quando a relação entre as variáveis x e y é de tipo exponencial (actividade de uma fonte radioactiva "versus"tempo, absorção de uma radiação "versus"espessura do filtro, etc.), i.e.

$$y = y_0 e^{\alpha x}, (\alpha \neq 0) \tag{9}$$

e se pretende, a partir de um conjunto de valores experimentais, determinar α e/ou y_0, o gráfico deve ser feito em papel semi-log ou em papel milimétrico em que se marca no eixo do yy o valor do logarítmo de y. Com efeito, logaritmizando a expressão 9 aplicando logaritmos neperianos tem-se

$$\ln y = \ln y_0 + \alpha x. \tag{10}$$

A mesma expressão pode ser linearizada aplicando-se logaritmos de base 10 (log), ficando nesse caso

$$\log y = \log y_0 + (\log e)\alpha x. \tag{11}$$

Portanto, em papel semi-log (e marcando y na escala logarítmica), o gráfico de 9 é uma recta cujo declive vale $\alpha \log e$ e cuja ordenada na origem vale y_0. Se o gráfico for feito em papel milimétrico, e se na ordenadas se marcar o logarítimo de y, então como mostram as eq. 10 e 11, o gráfico será tambem uma recta.

O valor de α determina-se a partir da recta ajustada ao gráfico da seguinte forma:

Se o gráfico tiver sido feito em papel semi-log

1. Tomam-se dois pares de pontos sobre a recta ajustadas aos pontos experimentais e lê-se nos eixos do gráfico os valores de (x_1, y_1) e (x_2, y_2),

2. Calcula-se $\ln y_1$ e $\ln y_2$ e usa-se a expressão seguite para determinar α

$$\alpha = \frac{\ln y_2 - \ln y_1}{x_2 - x_1} \tag{12}$$

Se o gráfico tiver sido feito em papel milimétrico com valores de $\ln y$

1. Tomam-se dois pares de pontos sobre a recta ajustadas aos pontos experimentais e lê-se nos eixos os valores de $(x_1, \ln y_1)$ e $(x_2, \ln y_2)$ (não se esqueça que agora os valores que foram marcados no gráfico foram x e $\ln y$.

2. Usa-se a eq. 12 como no ponto anterior

Se o gráfico tiver sido feito em papel milimétrico com valores de $\log y$

1. Tomam-se dois pares de pontos sobre a recta ajustadas aos pontos experimentais e lê-se nos eixos os valores de $(x_1, \log y_1)$ e $(x_2, \log y_2)$ (não se esqueça que agora os valores que foram marcados no gráfico foram x e $\log y$.

2. Determina-se α usando a expressão

$$\alpha = \frac{\log y_2 - \log y_1}{\log(e)(x_2 - x_1)} \tag{13}$$

Em papel milimétrico o gráfico de 9 seria evidentemente um troço de exponencial, e a determinação de α ou y_0 seria menos rápida e menos precisa. O interesse em linearizar um gráfico reside precisamente no facto de ser mais fácil "trabalhar"com uma linha recta do que com uma linha curva.

Convém também utilizar papel semilog quando é muito vasta a gama de valores a marcar num dos eixos coordenados. Se, por exemplo, y representar o fluxo de neutrões térmicos (com energias da ordem de três dezenas de meV) num ponto de um meio como a água e x designar a distância desse ponto á fonte de neutrões, os valores de y podem variar de 4 ordens a grandeza (de 10^5 a 10 neutrões $cm^{-2}s^{-1}$, por exemplo) enquanto x varia apenas de 1 a 25 cm. Num caso como este, os valores de y devem ser marcados numa escala logarítmica e os valores de x numa escala linear. Se se fizesse o gráfico em papel milimétrico, mesmo fazendo corresponder a 1 mm 100 neutrões $cm^{-2}s^{-1}$ seria necessária uma folha de papel com 1 m de comprimento, o que é pouco prático.

4.2.2 Papel log-log

Se as gamas de valores a marcar compreenderem várias ordens de grandeza tanto em ordenadas como em abcissas, vê-se agora facilmente que há conveniência em utilizar papel com ambas as escalas logarítmicas. É o que acontece, por exemplo, quando se pretende representar graficamente a secção eficaz de absorção de certos núcleos para neutrões com energias compreendidas entre uma dezena de meV e alguns MeV (espectro neutrónico de um reactor nuclear térmico).

Existe uma outra situação, de natureza diferente, em que se deve empregar papel log-leg. É necessário com frequência verificar experimentalmente relações do tipo

$$y = k.x^{\beta} \tag{14}$$

em que β e ou k são constantes (reais, quaisquer) a determinar. Logaritmizando a expressão em 14, tem-se

$$\log y = \log k + \beta \log x \tag{15}$$

Assim, em papel log-log, y varia linearmente com x, sendo β o valor do declive da recta. De notar que, como as escalas são idênticas no papel log-log, os seus eixos coordenados são do tipo dos do chamado "círculo trigonométrico", e β pode ser determinado mediante a razão dos comprimentos dos catetos de um triângulo rectângulo desenhado convenientemente sobre o gráfico. A outra maneira de calcular é análoga à indicada na alínea a), tendo agora em conta a expressão 15.

É possível ainda linearizar o gráfico correspondente á expressão (14) por outra via: marcando, em papel milimétrico, y em ordenadas e x^β em abcissas. Neste caso β funciona como parâmetro conhecido e k representa o declive da recta.

Como se vê, dado um conjunto de valores experimentais e conhecida a forma da lei de variação das grandezas em causa, pela conjugação dos dois processos de linearização indicados acima é possível inferir os valores dos parâmetros da lei, β e k no caso da expressão (14)

4.3 Produção de gráficos no Microsoft Office Excel

O programa de computador *Microsoft Excel*, que faz parte do pacote de programas *Microsoft Office*, é uma folha de cálculo especialmente adequada para contabilidade, mas que pode ser usada para produzir gráficos científicos elementares, servindo como alternativa à produção manual ou por recurso a programas mais especializados. Para se construir um gráfico no *Excel* deve-se proceder da seguinte forma:

1. Introduzir os dados das abcissas (x) e ordenadas (y) em 2 colunas da folha de cálculo do Excel.
2. Selecionar com o rato as 2 colunas de dados.
3. No menu *Inserir* escolher *Dispersão* e a seguir *Dispersão com apenas marcadores* (ou símbolos).

Depois do gráfico desenhado pode-se alterar o seu aspecto por forma a seguir as recomendações dadas neste documento. Selecionando a série de dados e clicando no botão direito do rato, pode-se, entre outras coisas,

1. Selecionar dados (pode substituir/modificar os dados atuais e/ou adicionar outros)
2. Adicionar linha de tendência, i.e., construir a reta de regressão linear (vêr mais adiante) .
3. Em *Formatar Série de Dados* pode escolher o tipo e tamanho dos marcadores (ou símbolos) para modificar os que aparecem por defeito.

4.4 Barras de erro e retângulos de precisão

Como se viu, o resultado de uma medição, x, tem sempre associado um certo erro, $u(x)$ (limite superior do erro, erro de leitura, erro padrão, erro provável, etc.), que se representa simbolicamente como se viu anteriormente por

$$X = x \pm u(x)$$

Se $u(x)$ designar o limite superior do erro de x, o significado desta representação é o seguinte: o valor da grandeza está situado no intervalo $[x - u(x), x + u(x)]$; se $u(x)$ designar a incerteza estatística de x, então a expressão tem outro significado: pode-se afirmar com aproximadamente 68% de certeza que o valor da grandeza está situado no referido intervalo. Para representar graficamente a margem de erro $u(x)$, desenha-se no gráfico a correspondente *barra de erro*, isto é, um segmento de recta de "comprimento" $2.u(x)$ centrado no ponto x.

No caso geral um ponto experimental corresponde a um par de valores, x e y, cada um dos quais com um certo, $u(x)$ e $u(y)$. Então a cada ponto do gráfico estão associadas duas barras de erro, uma paralela ao eixo dos yy e de "comprimento" $2.u(y)$, e outra paralela ao eixo dos xx e de "comprimento"$2.u(x)$, e ambas centradas no ponto experimental. Tem-se assim uma margem de erro a duas dimensões, definindo-se aquilo a que se chama $r$ectângulo de precisão do ponto experimental. Quando os erros de x e/ou de y são desprezáveis (em si mesmo(s) e/ou em relação à(s) escala(s) do gráfico), um rectângulo de precisão reduz-se a um "ponto"ou a uma barra de erro, conforme o caso.

5 Determinação da recta que melhor se ajusta ao gráfico

Após o gráfico construido, normalmente está-se interessado em descrever matematicamente a relação existente entre as duas variáveis. Para isso, começa-se por determinar o tipo de curva que melhor se ajusta ao comportamento dos dados, por exemplo e muito frequentemente, uma recta, e depois determina-se a sua equação. Iremos debruçar-nos sobre o ajuste de uma recta.

A recta que melhor se ajusta a um gráfico pode ser determinada *manualmente* ou por recurso a um *método estatístico*. O mais utilizado destes métodos estatísticos é o chamado *regressão linear (pelo método dos mínimos quadrados)*, que se encontra disponível na maioria das calculadoras científicas e nos programas de computador para elaboração de gráficos, como sejam o *Microsoft Excel* ou o *LibreOffice Calc*.

5.1 Método manual

O método manual, consiste no traçado directo sobre o gráfico de uma recta que se julgue melhor descrever a tendência revelada pelos pontos experimentai. Ao traçar a linha que melhor se ajusta aos pontos experimentais, não se deve "pretender" que ela passe necessariamente por todos os pontos. Deve haver apenas a preocupação de traçar a linha que melhor traduza a dependência global relativa das grandezas em causa. Os parâmetros da recta obtida, de equação geral $y = mx + b$, são determinados por leitura directa do gráfico ($m = \Delta y/\Delta x$).

5.2 Regressão linear

O método estatístico, também chamado método de regressão linear, consiste em minimizar a soma do quadrado da distância entre cada um dos pontos experimentais e a recta que se pretende ajustar, de equação $y = mx + b$. Depois de alguma manipulação matemática, chegam-se às seguintes equações que permitem calcular o declive (m) e a ordenada na origem (b) da recta ajustada:

$$m = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum (x_i)^2 - \left(\sum x_i\right)^2}$$

$$b = \frac{\sum y_i \sum (x_i)^2 - \sum x_i \sum x_i y_i}{N \sum (x_i)^2 - \left(\sum x_i\right)^2}$$

A regressão linear pelo método dos mínimos quadrados pode ser feita nas máquinas de calcular e em programas do tipo folha de cálculo, nomeadamente o Excel, que permitem calcular automáticamente os parâmetros da recta quando se faz o gráfico.

O procedimento relativo às máquinas de calcular depende, naturalmente, de cada modelo de máquina, pelo que se recomenda a consulta dos respectivos manuais de instruções.

Quanto ao procedimento a ser usado quando o ajuste se faz no computador, ir-se-á exemplificar como o Microsoft Excel. No *Excel*, a regressão linear pode ser feita directamente a partir do gráfico ou usando um pacote de funções extra chamado *Analysis ToolPak*. Para se fazer o ajuste a partir do gráfico deve-se proceder da seguinte forma:

1. Selecionar dados (pode substituir/modificar os dados atuais e/ou adicionar outros)

2. Adicionar linha de tendência, i.e., construir a reta de regressão linear. Para que a equação da reta apareça no gráfico, depois de escolher *Adicionar linha de tendência*, selecione *Opções da linha de tendência* e marque *Mostrar equação no gráfico*. Para mudar o formato dos coeficientes da reta, coloque o cursor em cima da equação da reta e clique no botão direito do rato. No menu que aparece, escolha *Formatar Rótulo da Linha de Tendência*. Depois na *Categoria* selecione *Número* e coloque o número de casas decimais que deseja. Em vez de usar a opção *Adicionar linha de tendência*, para fazer regressão linear, é recomendável usar o Analysis ToolPak, que será descrita a seguir.

3. Em *Formatar Série de Dados* pode escolher o tipo e tamanho dos marcadores (ou símbolos) se quiser modificar os que aparecem por defeito (automaticamente).

	Coeficientes	Erro-padrão
Intercetar	-0,433333333	1,792612506
Variável X1	0,978787879	0,028890566

Tabela 1: Tabela produzida pelo *Analysis ToolPack*. Se se considerar a recta com equação $y = ax + b$, então os valores apresentados na tabela em cima têm o seguinte significado: -0.4333333 é ordenada na origem (b), 1.17926 é a incerteza estimada para b, $u(b)$; 0,978787879 é o declive da recta (a) e 0,028890566 é a incerteza de a, $u(a)$.

5.2.1 Usando o Analysis ToolPak

O *Analysis ToolPak* é um suplemento do *Microsoft Office Excel* que permite, entre outras coisas, fazer a regressão linear a um conjunto de pontos, determinando os coeficientes da recta que melhor se ajusta a esse conjunto e os respectivos erros estatísticos, o que não é possivel quando se ajusta a recta directamente a partir do gráfico. Para o fazer, deve-se proceder da seguinte forma:

1. Activar o pacote, como explicado no apêndice.
2. Depois de carregar o *Analysis ToolPak*, o comando ficará disponível no separador *Dados* em *Análise de dados.*
3. Uma vez em *Análise de dados*, selecionar *Regressão.* Uma vez na caixa denominada Regressão, deverá escolher a colunas de x (intervalo x) e y (intervalo y), usando o cursor. Depois de fazer OK, aparecerá numa nova folha de cálculo um conjunto de tabelas, das quais só é importante a informação mostrada na tabela 1.

6 Determinação das incertezas nos parâmetros de uma recta ajustada a pontos experimentais

Frequentemente, após ajustar uma recta aos pontos experimentais, esta-se interessado em determinar os parâmetros dessa recta, isto é, as constantes m e b, respectivamente o declive e a ordenada na origem, pois a partir destes será possível calcular outras grandezas. Por exemplo, há uma experiência chamada *queda livre*, em que o valor local da aceleração da gravidade g se determina a partir do conhecimento do coeficiente angular da recta m através de $g = 2 \times m$. É de senso comum que, se m se determina a partir do gráfico de pontos experimentais que estão sujeitos a erro, m também deve ter associada uma incerteza $u(m)$, tal como b tem associado uma incerteza $u(b)$. A determinação destas incertezas faz-se de forma diferente, consoante o gráfico tenha sido traçado no computador ou à mão.

6.1 Gráfico traçado no computador

Para gráficos traçados no computador, a função *LINEST* ou pacote *Analysis ToolPack* do Excel fornece imediatamente $u(m)$ e $u(b)$. O procedimento é explicando no apêncice.

6.2 Gráfico traçado à mão

Quando a gráfico e a recta são traçados à mão, estão disponíveis vários métodos de estimação de $u(m)$ e $u(b)$, consoante a recta passe ou não pela origem, e os pontos estejam ou não bem alinhados com a recta.

6.2.1 Recta que passa pela origem

Suponha-se que se obtém numa realização experimental n pontos experimentais (x_i, y_i), que obedecem a uma relação do tipo

$$y = mx$$

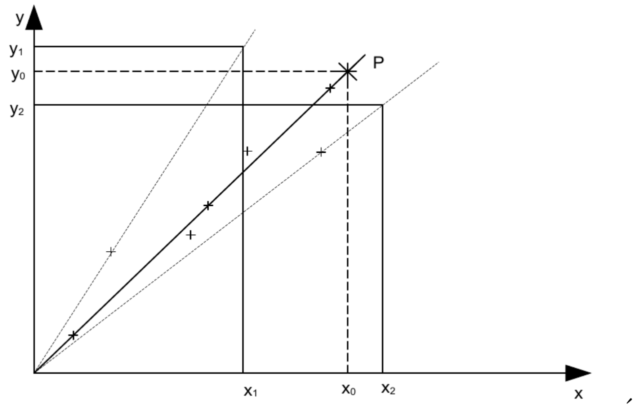

Figura 4: Gráfico experimental com recta que passa na origem.

com m constante, isto é, uma recta com ordenada na origem nula. Marque-se estes pontos num gráfico. Em geral, devido aos erros que afectam as medidas, os pontos não se distribuirão exactamente sobre uma linha recta. A recta a considerar então deverá ser a que "melhor" traduzir a lei de variação de y com x (os pontos experimentais que caiam fora dessa recta "média" e que estejam acima dela deverão "compensar"os que se encontram abaixo).

Considere-se um ponto P qualquer desta recta, de coordenadas x_0 e y_0 lidas no gráfico. O declive da recta é dado pelo quociente $m_0 = y_0/x_0$ (É conveniente não escolher o ponto P perto da origem). A partir do gráfico obtido anteriormente, considere-se as duas rectas definidas pela origem e pelos pontos experimentais mais afastados para baixo e para cima da recta "média"(note-se que pontos anormalmente afastados da recta "média"não deverão ser considerados, por corresponderem certamente a medições incorrectas). Os respectivos declives são $m_1 = y_1/x_1$ e $m_2 = y_2/x_2$ respectivamente. Toma-se Δm como o maior dos valores das diferenças $|m_1 - m_0|$ e $|m_2 - m_0|$. O limite superior do erro de m_0, $u(m_0)$, será dado por:

$$u(m_0) = \frac{\Delta m}{\sqrt{n-1}} \tag{16}$$

6.2.2 Recta que não passa pela origem

Este método usa os pontos extremos do gráfico como limites possíveis da variação do declive e ordenada na origem da recta ajustada. Este método é mais fácilmente aplicado para gráficos traçados à mão, com a recat ajustada também à mão ou por um método numérico. O método pode também ser aplicado em gráficos feitos no computador, sendo neste caso bastante complicado de implementar.

Suponha-se que se obtém, numa realização experimental, n pontos experimentais (x_i, y_i) que obedecem a uma relação do tipo

$$y = mx + b$$

com m e b constantes, isto é, uma recta que não intersecta a origem dos eixos. Marque-se estes pontos num gráfico. Para isso, procede-se do seguinte modo [8, 5], supondo um conjunto de n pontos experimentais (x_i, y_i):

1. Faz-se o gráfico, traçando a recta que melhor se ajusta aos pontos experimentais, R_0, de equação $y = m_0x + b_0$.

2. Desenham-se duas linhas paralelas a R_0 passando pelos pontos experimentais mais afastados de R_0 para cima e para baixo, *não incluindo as barras de erro* [5].

3. Desenham-se duas linhas paralelas ao eixo do yy, uma passando pelo primeiro ponto experimental à esquerda e outro passando pelo último ponto experimental à direita. Estas quatro linhas definem quatro pontos (1, 2, 3, 4, na figura), que são os vértices do chamado *paralelogramo de incerteza*

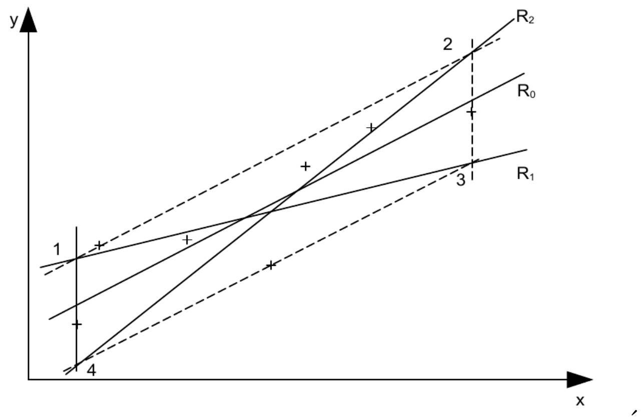

Figura 5

4. Pelos vértices opostos do paralelogramo fazem-se passar as rectas R_1 (definida pelos ponto 1 e 3) e R_2 (definida pelos pontos 2 e 4).
5. Determinam-se os valores dos parâmetros m e b das três rectas R_0, R_1 e R_2, respectivamente m_0, m_1, m_2 e b_0, b_1, b_2.
6. Calculam-se $\Delta m = |m_1 - m_0| = |m_2 - m_0|$[3] .
7. Calculam-se as diferenças $|b_1 - b_0|$ e $|b_2 - b_0|$. Toma-se Δb como a maior destas diferenças.
8. O limite superior dos erros $u(m_0)$ e $u(b_0)$ será dado por:

$$u(m_0) = \frac{\Delta m}{\sqrt{n-2}}, \quad u(b_0) = \frac{\Delta b}{\sqrt{n-2}}, \tag{17}$$

6.2.3 Pontos coincidentes com a recta

Pode acontecer que os pontos experimentais estejam tão bem alinhados sobre a recta experimental que não seja aplicável nenhum dos métodos anteriores. Então, o cálculo do limite superior do erro de declive é feito a partir do limite superior dos erros de leitura, no gráfico, dos valores de x e de y, e que são iguais a metade da menor divisão avaliável do gráfico. Isto corresponde a tratar o papel gráfico como um instrumento de medida analógico. De notar que o valor do limite superior do erro do declive da recta calculado por esta via, depende das escalas adoptadas no traçado do gráfico, diminuindo o valor do erro quando se ampliam as escalas. Esta incerteza é depois propagada ao cálculo do declive. Note que as incertezas nos dados experimentais não são aqui utilizadas e que as barras de erros são normalmente demasiadamente pequenas para serem visíveis no gráfico.

Uma vez que $m = \Delta y/\Delta x$, $\Delta y = y_2 - y_1$ e que $\Delta x = x_2 - x_1$ é tomado como constante, por aplicação da fórmula de propagação das incertezas vem que:

$$u_c(m) = \frac{\sqrt{u^2(y_1) + u^2(y_2)}}{x_2 - x_1} \tag{18}$$

A Apêndice: Utilização to Excel no traçado e análise de gráficos

A.1 Produção de gráficos no Microsoft Office Excel

O programa de computador *Microsoft Excel*, que faz parte do pacote de programas *Microsoft Office*, é uma folha de cálculo especialmente adequada para contabilidade, mas que pode ser usada para produzir gráficos científicos elementares, servindo como alternativa á produção manual ou por recurso a programas mais especializados. Para se construir um gráfico no *Excel* deve-se proceder da seguinte forma:

[3] Pode-se demonstrar que estas duas diferenças são iguais

1. Introduzir os dados das abcissas (x) e ordenadas (y) em 2 colunas da folha de cálculo do Excel.
2. Selecionar com o rato as 2 colunas de dados.
3. No menu *Inserir* escolher *Dispersão* e a seguir *Dispersão com apenas marcadores* (ou símbolos).

Depois do gráfico desenhado pode-se alterar o seu aspecto por forma a seguir as recomendações dadas neste documento. Selecionando a série de dados e clicando no botão direito do rato, pode-se, entre outras coisas,

1. Selecionar dados (pode substituir/modificar os dados atuais e/ou adicionar outros)
2. Adicionar linha de tendência, i.e., construir a reta de regressão linear (ver mais adiante) .
3. Em *Formatar Série de Dados* pode escolher o tipo e tamanho dos marcadores (ou símbolos) para modificar os que aparecem por defeito.

A.1.1 Barras de erros no Microsoft Excel

O *Microsoft Excel* não permite marcar no gráfico os retângulos de precisão, devendo-se, em vez disso, marcar as duas barras de erro, vertical e horizontal, formando uma cruz. Para o fazer deve-se proceder da seguinte forma:

1. Clique na série de dados.
2. No menu selecione *Esquemas* → *Barras de Erros* → *Mais opções de Barras de Erros.*
3. Formatar, clique em Série de dados selecionada.
4. Nos separadores *Barras de erro em x* ou *Barras de erro em y*, selecione as opções desejadas.

A.2 Como carregar no Excel o pacote *Analysis ToolPak*

O Analysis ToolPak é um suplemento do Microsoft Office Excel que fica disponível a partir do momento em que o Microsoft Office ou o Excel são instalados. Para o utilizar no Excel, no entanto, primeiro tem de ser carregado. Para o fazer:

1. Clique no Botão do Microsoft Office e, em seguida, clique em *Opções do Excel.*
2. Clique em *Suplementos* e, em seguida, na caixa *Gerir*, selecione *Suplementos do Excel.*
3. Clique em *Ir.*
4. Na caixa *Suplementos* disponíveis, marque a caixa de verificação Analysis ToolPak e clique em OK.

Atenção: se o Analysis ToolPak não aparecer na lista *Suplementos* disponíveis, clique em *Procurar* para o localizar. Se aparecer uma mensagem com a indicação de que o Analysis ToolPak não está instalado no computador, clique em *Sim* para o instalar.

B Literatura

Para o aluno interessado em aprofundar o tema da análise de dados tratamento de erros, aqui ficam algumas referências bibliográficas.

Os livros [3], [7] e [6] contêm introduções genérica sobre o tratamento de erros e fornecem uma prespectiva geral sobre o trabalho de laboratório. O livro [9] é a referência na análise e tratamento de erros e o livro [4] discute de forma aprofundada os métodos de ajuste de curvas.

A organização internacional *Joint Committee for Guides in Metrology (JCGM)* publica dois textos, disponíveis on-line, que são considerados as referências na forma de expressar incerteza de uma medição [1], e no vocabulário adoptado para se fazer referência a essas incertezas [2].

As referências [8, 5] discutem os médodos de estimar as incerezas associadas ao traçado manual de rectas. Este assunto é também abordado no cap. 3 da ref. [6].

Referências

[1] *Guide to the expression of uncertainty in measurement (GUM)*, volume ISO draft guide DGUIDE99998. Joint Committee for Guides in Metrology (JCGM), Geneva, 2004.

[2] *International Vocabulary of Metrology (VIM): Basic and General Concepts and Associated Terms.* Joint Committee for Guides in Metrology (JCGM), 2012.

[3] M. C. Abreu, L. Matias, and L. Peralta. *Física experimental - uma introdução.* Editorial Presença, 1994.

[4] R. B. D'Agostino. *Goodness of Fit Techniques.* 1986.

[5] C. James. Rapid methods for analysing errors in straight line graphs. *Physics Education*, 4(3):151, may 1969.

[6] Mike Pentz, Milo Shott, and Francis Aprahamian. *Handling experimental data.* Open University Press, 1 edition, 1988.

[7] Daryl W. Preston and Eric R. Dietz. *The Art of Experimental Physics.* 1991.

[8] John L. Safko. Error analysis of straight line plots in the undergraduate physics laboratory. *American Journal of Physics*, 33(5):379–382, 05 1965.

[9] J. R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements.* University Science Books, 2 edition, 1997.