Fundamentos de Análise de Dados

José Luís Argaín e Rui Guerra
Departamento de Física - FCT
Universidade do Algarve
jargain@ualg.pt, rguerra@ualg.pt

24 de setembro de 2025

Conteúdo

1 Conceitos fundamentais

- Numa **medida** utiliza-se um **instrumento** para aferir uma dada **quantidade física** de um objeto, a que chamamos **mesurando**.

  Por exemplo, o comprimento (quantidade física) de um parafuso (o mesurando), aferido com uma régua (o instrumento).

- Nesta cadeira fazemos a seguinte distinção, apenas por uma questão de clareza na explicação:
  - **medida**: um ato único, individual, de aferir uma quantidade física diretamente com um instrumento.
  - **medição**: envolve uma ou várias medidas e pode requerer o emprego de estatística e/ou um modelo físico para elaborar uma estimativa da quantidade física.

- O **erro de leitura de uma medida** corresponde a metade da menor escala num **instrumento analógico** e à menor escala num **instrumento digital**.

  Por exemplo, para uma régua milimétrica, a incerteza de leitura é 0,5 mm *e para um cronómetro digital de centésimas de segundo é 0,01* s.

- A **leitura de uma medida** associa sempre o **valor** registado no instrumento ao seu erro de leitura. E deve sempre verificar-se a seguinte regra:

O **número de casas decimas** (nCD) do valor e do erro de leitura devem ser iguais.

  Por exemplo, se a régua tiver uma escala milimétrica, uma medida (L) pode ser L=22 mm. *Como o erro de leitura é $\Delta L = 0,5$ mm, então a leitura da medida escreve-se*

  $$\boxed{L = 22,0 \pm 0,5 \text{ mm.}}$$

  Já se tivermos um cronómetro digital de segundos (o erro de leitura é $\Delta t = 1$ s) e medirmos 22 s, *a leitura é*

  $$\boxed{t = 22 \pm 1 \text{ s.}}$$

- O **número de algarismos significativos (nAS)** da medida é o número de algarismos que se lêem no valor de leitura (zeros à esquerda não contam).

  Por exemplo, no caso da régua, $L = 22,0 \pm 0,5$ mm $\Rightarrow$ nAS(L)=3. No caso do cronómetro, $t = 22 \pm 1$ s $\Rightarrow$ nAS(t)=2. Além disso nAS(022,0)=nAS(22,0)=3. Mas nAS(22,00)=4.

- O nAS representa a precisão do instrumento para cada medida particular. Mais AS implicam mais precisão. Assim, quer individualmente (medida), quer através de um processo de medição, o nAS não pode aumentar quando se escreve o resultado da leitura ou da estimativa. Portanto, a seguinte regra aplica-se sempre:

O nAS associado a uma medição não pode ser superior ao nAS das medidas individuais.

  No entanto pode ser inferior.

  Por exemplo, imagine-se medir o tempo de queda de um grave com um cronómetro manual de milésimos de segundo e com tempos típicos como, por exemplo, 2,345 s. *Cada medida tem 4 AS, mas o observador tem* *um tempo de reação de cerca de um décimo de segundo e obtém valores diferentes a cada medida. Assim, apesar de as medidas terem 4 AS, o resultado da medição (que se explica abaixo) só terá 2 AS.*

- **Regras para a operação algébrica de números com diferentes nCD ou nAS**

  Soma: mantém-se o menor nCD. Exemplos:
  1+2,222=3 (e não 3,222).
  0,3+0,888=1,2 (e não 1,188).
  4,4+6,66=11,1 (e não 11,06).

  Multiplicação: mantém-se o menor nAS. Exemplos:
  1x2,222=2 (e não 2,222).
  1,1x2,222=2,4 (e não 2,4442).
  6,47x8,192=53,0 (e não 53,00224).
  22,2x5,1234=114 (e não 113,73948).

- **Medição por repetição de medidas:** Na maior parte das vezes a medida é repetida várias vezes. A partir da **série de medidas** determina-se uma estatística de forma a caracterizar melhor a quantidade física. Em geral, o valor mais representativo da série de medidas ($x_i, i = 1 \dots, n$, n o número de repetições) é a sua **média** ($\bar{x}$):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \quad (1)$$

  Por exemplo, três medidas do tempo de queda de uma pedra com um cronómetro de décimas, a partir de uma posição fixa numa varanda: t_1 = 2,2 s, t_2 = 2,3 s, t_3 = 2,5 s. Idealmente todas as medidas deveriam ser iguais. Mas a experiência real tem vários fatores que podem variar de forma não controlável, como a posição exata em que se liberta a pedra e o tempo de reação do observador. Neste caso $\bar{t} = 2,3(3)$ s (o 3 entre parêntesis representa uma dízima infinita, ou seja $\bar{t} = 2,3333...$. Para já vamos representar todas as casas decimais).

- Os **erros aleatórios** originam a dispersão dos valores numa série de medidas. A sua causa é a variabilidade de fatores não controláveis (por exemplo, a ação do observador ou a própria variabiliadade do fenónemo físico).

  1. Se $n \geq 10$ estima-se a dispersão dos valores através do **desvio padrão** (σ):

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

  2. Se $n < 10$ estima-se a dispersão dos valores através do desvio médio (δ)

$$\delta = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|. \quad (3)$$

  Por exemplo, no caso anterior, os três desvios são $\Delta 1 = |2,2-2,33(3)| = 0,133(3)$, $\Delta 2 = |2,3-2,33(3)| = 0,033(3)$ e $\Delta 3 = |2,5 - 2,33(3)| = 0,166(6)$. Assim, $\delta = 0,111$. Suponhamos agora que são feitas 10 medidas com um cronómetro de centésimas: x_i = 2,33, 2,35, 2,44, 2,38, 2,28, 2,55, 2,44, 2,42, 2,31, 2,40 s. Neste caso a média é 2,39 s e o desvio padrão é σ = 0,07846 s, que deve ser arredondado ao mesmo nCD: σ = 0,07846 $\rightarrow$ 0,08 s.

- Os **erros sistemáticos** correspondem a um enviesamento sistemático e constante das leituras devido a algum desajuste no instrumento ou no processo de medição.

  Por exemplo, se o cronómetro anterior demorasse sistematicamente 0,2 s a parar após a ação do operador, então todos os resultados seriam sistematicamente translacionados de 0,2 s e t_1 = 2,2+0,2=2,4 s, t_2 = 2,3+0,2=2,5 s, t_3 = 2,5+0,2=2,7 s.

- O **erro observacional** é o resultado conjunto dos erros aleatório (i.e. σ ou δ consoante o número de medições) e sistemático.

- A **incerteza da medição**, Δx é o máximo entre os erros de leitura e observacional:

$$\Delta x = \max(\text{erro de leitura}, \text{erro observacional}). \quad (4)$$

  Se não houver erro observacional, no caso de haver apenas uma medida, então a incerteza de medição é igual ao erro de leitura. O erro observacional descreve a **dispersão** dos valores medidos. Assim, quando se pretende descrever essa dispersão, como acontece com os gráficos de barras (ver mais à frente), é a incerteza da medição que se deve usar.

- Mas o facto de se saber que os valores medidos estão contidos num certo intervalo (dado pela incerteza da medição, Δx), quer dizer que há um intervalo de confiança, mais estreito, para a localização do valor médio. Esse intervalo de confiança é dado pelo **desvio padrão da média**, σ_m, que também é conhecido por **erro padrão**:

$$\sigma_m = \frac{\Delta x}{\sqrt{n}}. \quad (5)$$

- O **resultado da medição** apresenta-se finalmente na forma

$$\boxed{X = \bar{x} \pm \sigma_m.} \quad (6)$$

  Esta expressão tem também uma interpretação em termos de intervalo de probabilidade. Para isso é necessário assumir que as medidas se distribuem em torno do valor médio de acordo com uma probabilidade de distribuição normal (gaussiana), o que em geral é verdade. Assim, a probabilidade que o valor real da média esteja dentro do intervalo

$$[\bar{x} - \sigma_m; \bar{x} + \sigma_m] \quad (7)$$

  é de aproximadamente 68%.

- Nesta cadeira assume-se um princípio conservador, que simplifica os cálculos:

O erro padrão é arredondado ao primeiro algarismo significativo.

  No caso das 10 medidas com o cronómetro de centésimas, o erro de leitura é 0,01 s*, enquanto o erro observacional é de 0,07846* s*. Portanto a incerteza de medição é*
  $$\Delta x = 0,07846\,\text{s} \tag{8}$$
  e o erro padrão é
  $$\sigma_m = \frac{0,07846}{\sqrt{10}} = 0,024811231\,\text{s} \to 0,02\,\text{s}. \tag{9}$$

- O resultado da medição deve ainda satisfazer a seguinte regra:

O número de casas decimais do valor médio deve ser igual ao número de casas decimais da incerteza de medição.

  Assim, no caso das 10 medidas temos que a **dispersão dos resultados** *é caracterizada pela* **incerteza da medição** *(σ no exemplo), arredondada ao primeiro AS,*
  $$\boxed{t = 2,39 \pm 0,08\,\text{s},} \tag{10}$$
  mas o **resultado da medição***, que exprime a incerteza em relação à média das medidas, é caracterizado pelo* **erro padrão***, σ_m:*
  $$\boxed{t = 2,39 \pm 0,02\,\text{s}.} \tag{11}$$

2 Incerteza padrão combinada

Até agora considerou-se o caso de apenas uma medição. Mas há muitos casos em que é necessário realizar duas ou mais medições e combiná-las numa fórmula para obter a estimativa de uma grandeza composta. Por exemplo, para estimar a densidade, ρ, de um sólido precisamos de ter previamente duas medições, uma para a massa, m, e outra para o volume, V, pois

$$\rho = \frac{m}{V}. \tag{12}$$

Se m tiver uma incerteza, Δm, e V tiver uma incerteza ΔV, como se calcula a incerteza na densidade, $\Delta\rho$?

2.1 Ilustração simplificada do problema

Nesta secção faz-se uma análise muito simples do problema para perceber o que quer dizer a incerteza combinada. O resultado a que se vai chegar é uma sobrestimação do erro combinado, mas ilustra de forma intuitiva os cálculos mais rigorosos da secção 2.2.

Considere-se ainda o caso da densidade. Tendo em conta as incertezas na massa e na densidade, qual é o valor máximo que ainda é razoável esperar para a densidade? Esse valor, que se vai representar por ρ^+, será obtido assumindo:

- o valor superior expectável para m, que é $m + \Delta m$ (porque m está no numerador);
- o valor inferior expectável para V, que é $V - \Delta V$ (porque V está no numerador).

Da mesma forma, o valor inferior expectável para ρ, representado por ρ^-, será obtido assumindo:

- o valor inferior expectável para m, que é $m - \Delta m$ (porque m está no numerador);
- o valor superior expectável para V, que é $V - \Delta V$ (porque V está no numerador).

Então

$$\rho^+ = \frac{m + \Delta m}{V - \Delta V} \tag{13}$$
$$\rho^- = \frac{m - \Delta m}{V + \Delta V}. \tag{14}$$

Podemos definir um intervalo de incerteza para a densidade com base nestas estimativas extremas, e que será

$$\rho \in [\rho^-; \rho^+]. \tag{15}$$

Por exemplo, se $m = 501 \pm 5$ g e se $V = 828 \pm 8$ cm^3, então

$$\rho = \frac{501}{828} = 0,605\ \text{g/cm}^3 \tag{16}$$
$$\rho^+ = \frac{501 + 5}{828 - 8} = 0,617\ \text{g/cm}^3 \tag{17}$$
$$\rho^- = \frac{501 - 5}{828 + 8} = 0,593\ \text{g/cm}^3. \tag{18}$$

Então
$$\rho \in [0,593; 0,617]\ \text{g/cm}^3 \tag{19}$$
ou seja,
$$\rho = 0,605 \pm 0,012\ \text{g/cm}^3 \tag{20}$$
que, na perspetiva conservadora adoptada na cadeira, se deve escrever simplesmente na forma
$$\boxed{\rho = 0,61 \pm 0,01\ \text{g/cm}^3.} \tag{21}$$

Este exemplo permitiu compreender como é que a incerteza das parcelas se materializa numa incerteza para a grandeza composta. O intervalo de incerteza obtido é, no entanto, sobrestimado. Neste caso a estimativa da incerteza composta é cerca de 40% superior à estimativa realista[1]. Apesar de tudo é perfeitamente razoável estimar o intervalo de confiança desta forma e, com duas variáveis, é uma excelente forma de ter rapidamente uma ideia do erro composto. No entanto, com mais variáveis, este tipo de estimativa simples pode degradar-se e não é aconselhável. A próxima secção dá a forma correta de estimar o erro composto.

[1] Isto é fácil de compreender a partir da interpretação probabilística do intervalo de confiança para a média. $m+\Delta m$ e $V-\Delta V$ (ou $m-\Delta m$ e $V+\Delta V$) são os valores do extremos superiores dos intervalos de confiança de 68% para as respetivas médias, ou seja, já tocam na gama de medidas "pouco prováveis". Calcular ρ a partir destes pontos extremos implica assumir simultaneamente duas medidas "pouco prováveis", o que já é "muito improvável".

2.2 O caso geral: fórmula de propagação dos erros

Considere-se um mesurando Y que não é medido diretamente, 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]

$$Y = f(X_1, X_2, \ldots, X_n). \tag{22}$$

A estimativa do mesurando Y, designada por y, é obtida a partir da equação (23), 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:

$$y = f(x_1, x_2, \ldots, x_n). \tag{23}$$

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

$$\Delta_c^2(y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \Delta^2(x_i). \tag{24}$$

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

2.2.1 Formas simplificadas

A equação (24) reduz-se a formas mais simples em casos de interesse prático. Por exemplo:

Soma ou subtração Se

$$y = a_1 x_1 \pm a_2 x_2 \pm \ldots \pm a_n x_n, \tag{25}$$

então

$$\Delta^2 y = a_1^2 \Delta^2(x_1) + \ldots + a_n^2 \Delta^2(x_n). \tag{26}$$

Multiplicação Se

$$y = x_1 \times x_2, \tag{27}$$

então

$$\Delta^2 y = x_2^2 \Delta^2(x_1) + x_1^2 \Delta^2(x_2). \tag{28}$$

Divisão Se

$$y = x_1 / x_2, \tag{29}$$

então

$$\Delta^2 y = \left(\frac{\Delta(x_1)}{x_2}\right)^2 + \left(\frac{x_1 \cdot \Delta(x_2)}{x_2^2}\right)^2. \tag{30}$$

Potências Se

$$y = A x_1^a x_2^b \ldots x_n^p, \tag{31}$$

então

$$\frac{\Delta^2 y}{y^2} = a^2 \frac{\Delta^2(x_1)}{x_1^2} + b^2 \frac{\Delta^2(x_2)}{x_2^2} + \ldots + p^2 \frac{\Delta^2(x_n)}{x_n^2}. \tag{32}$$

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

O resultado final para a medição, $y \pm \Delta y$, será escrito tendo em conta que: a) o nAS de y não pode ser superior ao menor nAS das parcelas; b) o erro $\Delta(y)$ é arredondado ao primeiro AS.

Vamos repetir o exemplo da densidade

$$\Delta^2 \rho = \left(\frac{\Delta m}{V}\right)^2 + \left(\frac{m.\Delta V}{V^2}\right)^2 = \tag{33}$$

$$\left(\frac{5}{828}\right)^2 + \left(\frac{501 \times 8}{828^2}\right)^2 = \tag{34}$$

$$= 0,000070642 \Rightarrow \Delta\rho = 0,00840488 \text{ g/cm}^3. \tag{35}$$

Neste caso obteríamos o resultado final

$$\rho = 0,605 \pm 0,008 \text{ g/cm}^3. \tag{36}$$

Comparando com o que obtemos com o método simples (equação 21), vemos que os resultados são parecidos, mas que se verifica o que anteriormente foi explicado: a estimativa da secção 2.1 é aproximadamente 40% superior à que agora se obteve. Com efeito 0,0084 × 1,4 = 0,0118 → 0,012, o valor da secção anterior.

3 Traçado de gráficos

Quando se pretende 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 aspetos 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 gráfico é produzido por um programa de computador. Um exemplo de um gráfico bem traçado é apresentado na Fig. 1.

3.1 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-lo 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 respetivos 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.

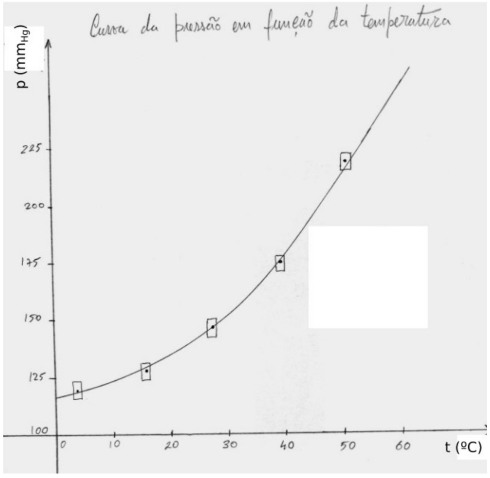

Figura 1: Exemplo de um gráfico corretamente executado à mão. Note-se a marcação das barras de erro horizontais e verticais formando um retângulo.

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 direta 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 esclareça 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 reta. A linha quebrada obtida não teria significado físico dado que as funções normalmente representadas têm variações suaves (derivadas finitas e contínuas).

3.2 Barras de erro e retângulos de precisão

Como se viu, o resultado de uma medição, x, tem sempre associada uma incerteza $\Delta(x)$ (limite superior do erro, erro de leitura, erro padrão, erro provável, etc.), que se representa simbolicamente por

$$X = x \pm \Delta x. \tag{37}$$

Se $\Delta(x)$ designar o limite superior do erro de x, o significado desta representação é o seguinte: o valor da grandeza está situado no intervalo $[x - \Delta(x), x + \Delta(x)]$; se $\Delta(x)$ designar a incerteza estatística de x (desvio padrão), então a expressão tem outro significado: pode afirmar-se com aproximadamente 68% de certeza que o valor da grandeza está situado no referido intervalo. Para representar graficamente a margem de erro $\Delta(x)$, desenha-se no gráfico a correspondente barra de erro, isto é, um segmento de reta de "comprimento" $2\Delta(x)$ centrado no ponto x. No caso geral,há incertezas tanto em x como em y, caracterizadas por $\Delta(x)$ e $\Delta(y)$, respetivamente. Então a cada ponto do gráfico estão associadas duas barras de erro, uma paralela ao eixo dos yy e de "comprimento" $2\Delta(y)$, e outra paralela ao eixo dos xx e de "comprimento" $2\Delta(x)$, e ambas centradas no ponto experimental. Tem-se assim uma margem de erro a duas dimensões, definindo-se aquilo a que se chama o retângulo de precisão do ponto experimental. Quando os erros de x e/ou de y são muito pequenos em relação à(s) escala(s) do gráfico, um retângulo de precisão reduz-se a um "ponto"ou a uma barra de erro, conforme o caso.

4 Determinação da reta que melhor se ajusta ao gráfico

Após o gráfico construído, pretende-se, geralmente, 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. No caso mais simples, que é também o mais frequente, a relação é linear, o que se traduz numa reta. Iremos debruçar-nos sobre o ajuste de uma reta. A reta 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 a 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.

4.1 Método manual

O método manual consiste no traçado directo sobre o gráfico de uma reta que se julgue melhor descrever a tendência revelada pelos pontos experimentais. 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 reta obtida, que tem a equação geral bem conhecida,

$$y = mx + b, \tag{38}$$

são determinados por leitura direta do gráfico. Assim, o declive é dado por

$$m = \frac{\Delta y}{\Delta x}, \tag{39}$$

em que $\Delta(x)$ e $\Delta(y)$ são as variações observadas entre dois pares de pontos. Depois basta escolher um par de pontos específicos, (x_0, y_0), para obter b:

$$b = y_0 - mx_0. \tag{40}$$

4.2 Regressão linear

A regressão linear pelo método dos mínimos quadrados pode ser feita nas máquinas de calcular e nos programas de gráficos, nomeadamente o Excel. O procedimento relativo às máquinas de calcular depente, naturalmente, de cada modelo de máquina, pelo que se recomenda a consulta dos respetivos manuais de instruções. Quanto ao procedimento a ser usado quando o ajuste se faz no computador, ir-se-á exemplificar com o Microsoft Excel, por ser o programa mais difundido. Além disso, os "clones" do Excel trabalham de forma semelhante. No Excel, a regressão linear pode ser feita diretamente a partir do gráfico ou usando a função *proj.lin*. Para se fazer o ajuste a partir do gráfico deve proceder-se da seguinte forma:

1. Selecionar os dados (a forma mais fácil é ter as colunas x e y lado a lado e selecionar as duas).

2. Adicionar linha de tendência, i.e., construir a reta de regressão linear (na maior parte das versões, em particular as mais recentes, esta opção aparece num menu situado no canto superior direito do gráfico, onde se escolhe *Adicionar linha de tendência*).

3. Para que a equação da reta apareça no gráfico, selecione *Opções da linha de tendência* (no mesmo menu, mas num segundo nível de opções). Aparece um painel lateral onde pode marcar a caixa *Mostrar equação no gráfico*.

4. 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.

5. No mesmo painel, mas no separador *Formatar Série de Dados*, pode escolher o tipo e tamanho dos marcadores (ou símbolos) se quiser modificar os que aparecem por defeito (automaticamente).

Em vez de usar a opção Adicionar linha de tendência, para fazer regressão linear, é recomendável usar a função *proj.lin*, que é explicada a seguir.

4.3 Usando a função proj.lin

Os dados estão dispostos em duas colunas, correspondentes aos dados x e y. O procedimento é como se segue:

1. Seleccionar uma célula livre e em branco, em qualquer ponto da folha de cálculo e inserir a função proj.lin, como indicado na figura 2.

2. Selecionar os dados da seguinte forma

Figura 2: Passo 1: inserir proj.lin numa célula.

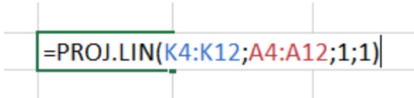

Figura 3: Passo 2: inserir os dados.

   (a) Selecionar a coluna y (no exemplo, PROJ.LIN(K4:K12 e adicionar, na barra da fórmula, um ponto e vírgula (;), ou seja, neste exemplo, PROJ.LIN(K4:K12; (em algumas versões o separador é a vírgula)

   (b) Selecionar a coluna x e adicionar ponto e vírgula (ou vírgula) PROJ.LIN(K4:K12;A4:A12;

   (c) Escrever 1 e adicionar ponto e vírgula (ou vírgula) PROJ.LIN(K4:K12;A4:A12;1;

   (d) Escrever outro 1 e fechar o parêntesis PROJ.LIN(K4:K12;A4:A12;1;1)

   (e) Fazer ENTER.

   Após estes passos deve ver-se uma fórmula como a da figura 3

3. Após o ENTER, o Excel faz aparecer todos os resultados (alguns de que não precisamos), tal como se mostra na figura 4.

Um dos números devolvidos pelo proj.lin é a **incerteza no declive**. Isto será útil para determinar o erro de uma medição baseada no declive.

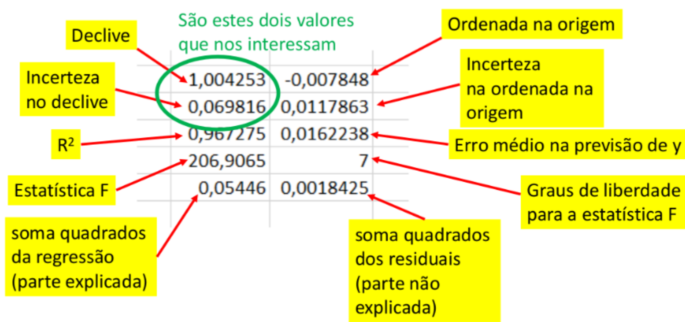

Figura 4: Passo 2: os resultados dados pela função proj.lin.

4.4 Determinação das incertezas nos parâmetros de uma reta ajustada a pontos experimentais

Como o declive, m, se determina a partir dos pontos experimentais, e estes estão sujeitos a incerteza, então o próprio declive também tem uma incerteza. Por exemplo, se repetíssemos a experiência iríamos obter um declive ligeiramente diferente. Então associa-se a m uma incerteza $\Delta(m)$. Da mesma forma, também à ordenada na origem b está associada uma incerteza $\Delta(b)$. A determinação destas incertezas também se faz de forma diferente, consoante o gráfico tenha sido traçado no computador ou à mão.

A função proj.lin do Excel fornece imediatamente $\Delta(m)$ e $\Delta(b)$, como se pode ver na figura 4. Estes valores estão dados na segunda linha do quadro de resultados. Escrevemos então que

$$\begin{aligned} m &= \bar{m} \pm \Delta m && (41) \\ b &= \bar{b} \pm \Delta b, && (42) \end{aligned}$$

em que $\bar{m}$ e $\bar{b}$ são os valores centrais da determinação, lidos da primeira linha dos resultados de proj.lin, e Δm Δb são as respetivas incertezas, retiradas da segunda linha dos resultados. Por exemplo, usando os resultados da figura 4, o declive e a ordenada na origem são

$$\begin{aligned} m &= 1,00 \pm 0,07 && (43) \\ b &= -0,01 \pm 0,01, && (44) \end{aligned}$$

A Utilização to Excel no traçado e análise de gráficos

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 (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 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.