

Capítulo 5

Observações e mínimos quadrados

5.1 Introdução

Consideremos uma determinada entidade que pretendemos medir. Para determinar a medida em questão fazemos, em geral, várias observações ou medições e portanto somos conduzidos a uma diversidade de valores para a medida dessa entidade. A variabilidade dos resultados da medição são originados por vários factores e portanto os valores obtidos, são em geral redundantes. Deveremos possuir um critério que nos permita a partir dos valores observados, determinar um valor único para a medida da entidade. O critério que iremos estudar é o critério dos mínimos quadrados e, ao usar tal critério, diremos que estamos a proceder ao ajustamento dos valores observados.

Neste capítulo os termos “observação” e “medição” irão ser usados com o mesmo significado. As medições são caracterizadas pelo seguinte conjunto de propriedades:

1. o acto de medir/observar, que designamos por medição/observação, significa sempre a realização de uma operação física ou uma série de operações;
2. o resultado numérico obtido por medição é usado para representar a medida do objecto e transporta consigo todas as circunstâncias do processo pelo qual foi obtido;
3. as medições são sempre efectuadas com a ajuda de objectos (excepto a operação de contagem simples);
4. ao medir estamos a estabelecer uma comparação com um padrão estabelecido por convenção vindo, portanto, o resultado da medida expresso em unidades;
5. as medições referem-se a conceitos teóricos ou abstrações geométricas (tais como comprimento, área, amplitude de um ângulo, etc) que são usadas para caracterizar uma determinada entidade mas não têm significado físico;
6. o resultado numérico obtido por medição só tem significado quando está associado ao conceito teórico a que a observação diz respeito.

Consideremos, por exemplo, um triângulo para o qual pretendemos determinar a medida da sua área. Para o efeito deveremos determinar a medida do comprimento da base e da altura. A medição é efectuada com o auxílio de uma régua. Ao dizermos que o lado mede 0.10 m estamos a estabelecer uma comparação do comprimento do lado com o metro padrão. Dizer que o lado mede 0.10 m é um abuso de linguagem; deveríamos ser precisos e afirmar: “a medida do comprimento do lado é 0.10 m”.

5.2 Modelo matemático e modelo estocástico

Consideremos um rectângulo $[ABCD]$ para o qual pretendemos determinar a medida da área. Sabemos que a medida da área é dada por $\ell_A = \ell_{\overline{AB}}\ell_{\overline{BC}}$, sendo $\ell_{\overline{AB}}$ e $\ell_{\overline{BC}}$, respectivamente, as medidas dos lados \overline{AB} e \overline{BC} . Atendendo ao seu carácter aleatório, cada observação realizada é encarada como uma concretização de uma variável aleatória. Assim teremos as variáveis aleatórias (v.a.) X_A , $X_{\overline{AB}}$ e $X_{\overline{BC}}$, que estão definidas

no universo $\Omega = \{\text{observações}\}$, também chamado espaço fundamental ou conjunto dos acontecimentos, e que assumem valores reais. As duas últimas v.a. são concretizadas, por exemplo, n e m vezes respectivamente, obtendo-se duas amostras. Estas amostras permitem definir duas variáveis estocásticas (ou estatísticas)

$$x_{\overline{AB}} : \{O_1, \dots, O_n\} \in \Omega \longrightarrow \{x_{\overline{AB},1}, \dots, x_{\overline{AB},n}\}$$

e

$$x_{\overline{BC}} : \{O_1, \dots, O_m\} \in \Omega \longrightarrow \{x_{\overline{BC},1}, \dots, x_{\overline{BC},m}\}.$$

As observações efectuadas não são mais do que concretizações da mesma v.a.. Atendendo a este facto podemos admitir que:

1. as observações são independentes (umas não influenciam as outras);
2. as observações são efectuadas com igual precisão (com o mesmo instrumento de medida).

Na determinação da medida da área do rectângulo foi definido um **modelo matemático**, isto é, um sistema teórico de conceitos abstractos – medida da área, medida do comprimento dos lados – que descreve uma entidade física – a área de um rectângulo. Neste sistema teórico distinguem-se, claramente, duas partes: a primeira (modelo funcional) é constituída pela expressão matemática envolvendo as variáveis do modelo e descreve as propriedades deterministas da entidade em estudo; a segunda (modelo estocástico) é constituída pelo conjunto de hipóteses que são estabelecidas sobre as observações efectuadas.

O modelo estocástico surge uma vez que as observações são concretizações de variáveis aleatórias. Por definição, diz-se que um fenómeno é estocástico ou conjectural sempre que os seus casos particulares dependem do acaso e a respeito dos quais só é possível enunciar probabilidades.

5.3 Alguns conceitos probabilísticos

5.3.1 Distribuições unidimensionais

As v.a. podem dividir-se em dois grupos: discretas e contínuas. As discretas podem tomar um conjunto discreto de valores $x_1, x_2, \dots, x_n, \dots$, finito ou não. Isto é, se X_d for uma v.a. discreta temos que

$$X_d : \Omega \longrightarrow \{x_1, x_2, \dots, x_n, \dots\}.$$

As v.a. contínuas podem tomar quaisquer valores num intervalo real ou mesmo em toda a recta real. Assim, se X_c for uma v.a. contínua

$$X_c : \Omega \longrightarrow \mathbb{R}.$$

Comecemos por considerar o caso discreto. Seja X_d uma v.a. discreta e sejam $x_i, i = 1, 2, \dots$, valores que X_d pode tomar. Seja Ω o universo de X_d . Consideremos os acontecimentos

$$A_i = \{w \in \Omega : X_d(w) = x_i\}, \quad i = 1, 2, \dots,$$

e seja p_i a probabilidade de cada um destes acontecimentos ocorrer, isto é,

$$p_i = P(A_i) = P(X_d = x_i), \quad i = 1, 2, \dots,$$

em que $0 \leq p_i \leq 1$ e $\sum_i p_i = 1$.

Definição 5.1 (Função de probabilidade) *Seja $X_d : \Omega \longrightarrow \{x_1, x_2, \dots, x_n, \dots\}$ uma v.a. com $P(X_d = x_i) = p_i$. A função*

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x \mapsto f(x) = \begin{cases} p_i & x = x_i \\ 0 & x \neq x_i \end{cases}$$

é designada função de probabilidade da v.a. X_d .

No caso contínuo, esta definição não tem significado, sendo substituída pela de função densidade de probabilidade. Porém, antes de definir esse conceito, vamos apresentar a definição de função de repartição (ou de distribuição) que é válida tanto no caso discreto como no caso contínuo.

Definição 5.2 (Função de repartição) *Seja X uma v.a. (discreta ou contínua). A função*

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F(x) = P(X \leq x) \end{aligned}$$

é designada função de distribuição ou função de repartição da v.a. X .

Notamos que, no caso discreto,

$$\begin{aligned} F(x) &= P(X_d \leq x) \\ &= P(\{w \in \Omega : X_d(w) \leq x\}) \\ &= P(\cup_{i: x_i \leq x} \{w \in \Omega : X_d(w) = x_i\}) \\ &= \sum_{i: x_i \leq x} P(\{w \in \Omega : X_d(w) = x_i\}) \\ &= \sum_{i: x_i \leq x} f(x_i) \\ &= \sum_{i: x_i \leq x} p_i \end{aligned}$$

Vamos agora considerar a definição de função densidade de probabilidade. Seja X_c uma v.a. contínua, i.e. $X_c : \Omega \longrightarrow \mathbb{R}$ em que Ω denota o espaço fundamental.

Definição 5.3 (Função densidade) *Dizemos que $f : \mathbb{R} \longrightarrow \mathbb{R}$ é a função densidade (de probabilidade) da v.a. contínua X_c se f é não negativa e $\int_{\mathbb{R}} f(x)dx = 1$. Nesse caso tem-se que*

$$F(x) = P(X_c \leq x) = \int_{-\infty}^x f(t)dt.$$

A definição anterior poderia ter sido usada para definir v.a. contínua. Assim, uma v.a. X_c é contínua se a sua função de repartição admitir a representação expressa na definição anterior.

Observação 5.4 *Antes de prosseguir, façamos uma pequena reflexão sobre o significado da designação de função densidade introduzida na definição anterior. Considere-se, uma vez que $P(X \in \mathbb{R}) = 1$, uma unidade de probabilidade distribuída (ou repartida) por todo o eixo real de tal forma que a quantidade de probabilidade atribuída ao intervalo $(-\infty, x]$ seja igual a $F(x)$, qualquer que seja x . Nestas condições, a quantidade de probabilidade no intervalo $(x, x+\epsilon]$ é dada pela diferença $F(x+\epsilon) - F(x)$. Logo, o quociente $(F(x+\epsilon) - F(x))/\epsilon$ é a quantidade média de probabilidade nesse intervalo. O limite $\lim_{\epsilon \rightarrow 0} (F(x+\epsilon) - F(x))/\epsilon = F'(x)$, se existir, representa a densidade de probabilidade no ponto x . No entanto, nos pontos onde $F'(x)$ existe tem-se que, pela definição anterior, $F'(x) = f(x)$. Assim, convencionando escrever $f(x) = 0$ nos pontos onde $F'(x)$ não existe tem-se que a função densidade determina univocamente a função de repartição.*

O facto de $P(X = x) = F(x) - F(x-0)$, ver Exercício 5.3.1, permite concluir que se a função de repartição F for contínua em todo o $x \in \mathbb{R}$, então $P(X = x) = 0$, i.e., todos os pontos de \mathbb{R} têm probabilidade zero (no entanto $X = x$ não é um acontecimento impossível).

Se a função de repartição tem uma descontinuidade num ponto $x \in \mathbb{R}$, então $P(X = x) = F(x) - F(x-0) > 0$, isto é, o salto de função é a probabilidade da v.a. assumir o valor x .

As funções de densidade e de repartição de uma v.a. são determinadas (e caracterizadas) por um conjunto de parâmetros que são úteis para a compreensão do seu comportamento. O primeiro desses parâmetros expressa a ideia intuitiva de média e é designado por *esperança* ou *valor esperado* ou ainda *valor mais provável* da v.a.. Este valor será “o mais provável de ocorrer” quando se considera a concretização da v.a. em causa.

Definição 5.5 (Esperança) (i) Seja X_d uma v.a. discreta que pode tomar os valores $x_i, i = 1, 2, \dots$, e seja f a sua função de probabilidade. Chamamos *esperança* (média, ou valor mais provável) de X_d a

$$E(X_d) = \sum_i x_i f(x_i).$$

(ii) Seja X_c uma v.a. contínua de função densidade f . Chamamos *esperança* (média, ou valor mais provável) de X_c a

$$E(X_c) = \int_{\mathbb{R}} x f(x) dx.$$

Um outro parâmetro importante é a *variância* pois dá-nos informação relativamente à dispersão dos valores da v.a. em torno do seu valor mais provável.

Definição 5.6 (Variância e desvio padrão) Seja X uma v.a. de esperança $E(X)$. Ao valor de $E[(X - E(X))^2]$ chamamos *variância* da v.a. X que é denotada por σ_X^2 . À raiz quadrada da variância da v.a. X chamamos *desvio padrão* de X e é representado por σ_X .

Note-se que:

1. se X_d é uma v.a. discreta que pode tomar os valores $x_i, i = 1, 2, \dots$, e f a sua função de probabilidade, então

$$\sigma_{X_d}^2 = \sum_i (x_i - E(X_d))^2 f(x_i);$$

2. se X_c é uma v.a. contínua de função densidade f , então

$$\sigma_{X_c}^2 = \int_{\mathbb{R}} (x - E(X_c))^2 f(x) dx.$$

A variância de uma v.a. também pode ser definida por $\sigma_X^2 = E(X^2) - E(X)^2$. De facto,

$$\sigma_X^2 = E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - E(X)^2.$$

5.3.2 Distribuições bidimensionais

Em inúmeras situações o estudo probabilístico ou estatístico envolve $k > 1$ propriedades ou características qualitativas dos elementos w do espaço fundamental Ω . Nesses casos, faz-se corresponder a cada ponto $w \in \Omega$ um ponto (x_1, \dots, x_k) de \mathbb{R}^k , isto é,

$$w \mapsto (X_1(w), \dots, X_k(w)) := X(w).$$

Quando para todo o ponto (x_1, \dots, x_k) de \mathbb{R}^k o conjunto

$$\{w \in \Omega : X_1(w) = x_1, \dots, X_k(w) = x_k\}$$

é um acontecimento, diz-se que $X(w)$ é um vector aleatório ou uma variável aleatória k -dimensional. Daqui por diante o vector $X(w)$ escreve-se $X = (X_1, \dots, X_k)$.

As v.a. bidimensionais são suficientemente importantes para justificar o seu estudo particular. Neste caso (quando $k = 2$), em vez de (X_1, X_2) escreveremos, na maior parte das vezes, (X, Y) .

Dada uma v.a. bidimensional (X, Y) a probabilidade de se obter um ponto na região do plano \mathbb{R}^2 definido pelas desigualdades $X \leq x$ e $Y \leq y$,

$$P(X \leq x, Y \leq y) = P(\{w \in \Omega : X(w) \leq x, Y(w) \leq y\}),$$

existe sempre (por definição) e permite introduzir uma função real de duas variáveis reais de acordo com a seguinte definição.

Definição 5.7 (Função de repartição) *Seja (X, Y) uma v.a. bidimensional. A função*

$$\begin{aligned} F : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto F(x, y) = P(X \leq x, Y \leq y) \end{aligned}$$

é designada função de repartição (ou distribuição) da v.a. bidimensional (X, Y) .

Sejam $a < b$, $c < d$ e

$$\begin{aligned} I &= \{w \in \Omega : a < X(w) \leq b, c < Y(w) \leq d\}, \\ A &= \{w \in \Omega : X(w) \leq a, Y(w) \leq c\}, \\ B &= \{w \in \Omega : a < X(w) \leq b, Y(w) \leq c\}, \\ C &= \{w \in \Omega : X(w) \leq a, c < Y(w) \leq d\}. \end{aligned}$$

Então

$$F(b, d) = P(I \cup A \cup B \cup C) = P(I) + P(A) + P(B) + P(C).$$

Por outro lado

$$\begin{aligned} P(A) &= F(a, c), \\ F(b, c) &= P(B) + F(a, c) \implies P(B) = F(b, c) - F(a, c), \\ F(a, d) &= P(C) + F(a, c) \implies P(C) = F(a, d) - F(a, c). \end{aligned}$$

Substituindo obtemos

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(b, c) - F(a, d).$$

Provamos assim o seguinte teorema.

Teorema 5.8 *Seja (X, Y) um v.a. com função de repartição F . Se $a < b$ e $c < d$ então*

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(b, c) - F(a, d).$$

Vamos agora apresentar a noção de função de probabilidade (para v.a. discretas) e a de função densidade (para v.a. contínuas). Seja (X, Y) uma variável que pode tomar os valores $\{(x_i, y_j), i = 1, \dots, j = 1, \dots\}$ e sejam p_{ij} valores não negativos tais que

$$P(X = x_i, Y = y_j) = p_{ij},$$

com $\sum_i \sum_j p_{ij} = 1$.

Definição 5.9 (Função de probabilidade) *Seja (X, Y) uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, \dots, j = 1, \dots\}$ e sejam p_{ij} as probabilidades*

$$P(X = x_i, Y = y_j) = p_{ij}.$$

A função

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto f(x, y) = \begin{cases} p_{ij} & (x, y) = (x_i, y_j) \\ 0 & (x, y) \neq (x_i, y_j) \end{cases} \end{aligned}$$

é designada por função de probabilidade de (X, Y) .

Esta noção permite concluir que, quando se consideram v.a. discretas, a função de repartição é definida por

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j) . \end{aligned}$$

Consideremos agora o caso contínuo e a noção de função densidade.

Definição 5.10 (Função densidade) *Seja (X, Y) um v.a. com função de repartição F . Dizemos que uma função $f(x, y)$ não negativa, integrável e tal que*

$$\int_{\mathbb{R}^2} f(x, y) dx dy = 1$$

é a função densidade da v.a. se

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(t, u) dt du .$$

Se a função de repartição admitir a representação anterior, dizemos que (X, Y) é uma v.a. contínua.

5.3.3 Distribuições marginais

Quando consideramos um vector de duas v.a., (X, Y) , podemos estar interessados em estudar o que se passa com cada uma das variáveis isoladamente. Poderemos pretender

$$P(X \leq x) = P(\{w \in \Omega : X(w) \leq x, Y(w) \text{ qualquer}\}) = P(X \leq x, Y < +\infty).$$

Para isso, vamos definir a função

$$F_1(x) := \lim_{y \rightarrow +\infty} F(x, y)$$

A esta função chamaremos função de repartição marginal da v.a. X . De igual modo, a função

$$F_2(y) := \lim_{x \rightarrow +\infty} F(x, y)$$

é designada por função de repartição marginal de Y .

Consideremos o caso em que (X, Y) é uma v.a. discreta que pode tomar os valores $\{x_{ij}, i = 1, 2, \dots, j = 1, 2, \dots\}$ e seja f a função de probabilidade desta v.a. bidimensional. Assim, as funções de repartição marginais são dadas por

$$\begin{aligned} F_1(x) = P(X \leq x) &= P(\{w \in \Omega : X(w) \leq x, Y(w) \text{ qualquer}\}) \\ &= P(X \leq x, Y < +\infty) \\ &= \sum_{x_i \leq x} \sum_{y_j} f(x_i, y_j) \end{aligned}$$

e

$$\begin{aligned} F_2(y) = P(Y \leq y) &= P(\{w \in \Omega : X(w) \text{ qualquer}, Y \leq y\}) \\ &= P(X < +\infty, Y \leq y) \\ &= \sum_{x_i} \sum_{y_j \leq y} f(x_i, y_j) . \end{aligned}$$

Estas funções podem ainda ser representadas em termos das chamadas funções de probabilidade marginais que são definidas, para o caso da v.a. discreta em causa por

$$f_1(x) = \begin{cases} \sum_{y_j} f(x, y_j) & x = x_i \\ 0 & x \neq x_i \end{cases} ,$$

designada por função de probabilidade marginal de X , e

$$f_2(x) = \begin{cases} \sum_{x_i} f(x_i, y) & y = y_j \\ 0 & y \neq y_j \end{cases}$$

designada por função de probabilidade marginal de Y . Com estas definições temos que

$$F_1(x) = \sum_{x_i \leq x} f_1(x_i)$$

e

$$F_2(y) = \sum_{y_j \leq y} f_2(y_j).$$

Consideremos agora (X, Y) uma v.a. contínua com f a sua função densidade. Então

$$F_1(x) = \int_{-\infty}^x \int_{\mathbb{R}} f(t, y) dy dt = \lim_{y \rightarrow +\infty} \int_{-\infty}^x \int_{-\infty}^y f(t, u) du dt,$$

e

$$F_2(y) = \int_{-\infty}^y \int_{\mathbb{R}} f(x, t) dt dy = \lim_{x \rightarrow +\infty} \int_{-\infty}^y \int_{-\infty}^x f(t, u) dt du,$$

são as funções de repartição marginal de X e Y respectivamente.

Estas funções podem ainda ser representadas em termos das chamadas funções de densidade marginais que são definidas, para o caso da v.a. contínua em causa por

$$f_1(x) = \frac{d}{dx} F_1(x) = \frac{d}{dx} \int_{-\infty}^x \int_{\mathbb{R}} f(t, y) dy dt = \int_{\mathbb{R}} f(x, y) dy,$$

designada por função de densidade marginal de X , e

$$f_2(y) = \frac{d}{dy} F_2(y) = \int_{\mathbb{R}} f(x, y) dx$$

designada por função de densidade marginal de Y . Com estas definições temos que

$$F_1(x) = \int_{-\infty}^x f_1(t) dt,$$

e

$$F_2(y) = \int_{-\infty}^y f_2(u) du.$$

5.3.4 Independência

Começemos por considerar a seguinte definição.

Definição 5.11 (Independência) Consideremos o vector aleatório (X, Y) . Dizemos que X é independente de Y se, para todo $(x, y) \in \mathbb{R}^2$, os acontecimentos

$$A = \{w \in \Omega : X \leq x, Y < +\infty\}, \quad e \quad B = \{w \in \Omega : X < +\infty, Y \leq y\}$$

forem independentes, isto é, se $P(A \cap B) = P(A)P(B)$.

Seja F a função de repartição conjunta da uma v.a. (X, Y) . Então, se X e Y forem independentes, tem-se que, para $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P(\{X \leq x, Y < +\infty\} \cap \{X < +\infty, Y \leq y\}) \\ &= P(\{X \leq x, Y < +\infty\})P(\{X < +\infty, Y \leq y\}) \\ &= F_1(x)F_2(y) \end{aligned}$$

e portanto a função de distribuição conjunta é o produto das distribuições marginais. Reciprocamente, se a distribuição conjunta é o produto das funções distribuições marginais então as v.a. X e Y são v.a independentes. Provamos assim o seguinte teorema.

Teorema 5.12 *Sejam X e Y duas v.a. e seja F a função de distribuição conjunta do vector aleatório (X, Y) . As v.a. X e Y são independentes se e só se*

$$F(x, y) = F_1(x)F_2(y), \quad (x, y) \in \mathbb{R}^2,$$

em que F_1 e F_2 são as função de distribuição marginal de cada uma das componentes.

Vamos agora demonstrar duas condições necessárias e suficientes para a independência de v.a., uma para o caso discreto e outra para o caso contínuo.

Teorema 5.13 *Seja (X, Y) uma v.a. discreta que podem tomar os valores*

$$\{(x_i, y_j), i = 1, 2, \dots, j = 1, \dots\}.$$

As v.a. X e Y são independentes se e só se

$$f(x_i, y_j) = f_1(x_i)f_2(y_j),$$

em que f é a função de probabilidade conjunta do vector aleatório (X, Y) e f_1, f_2 são as funções de probabilidade marginais de X e Y respectivamente.

Demonstração: Suponhamos que X e Y são independentes. Então

$$P(X = x_i, Y = y_j) = P(X = x_i, Y < +\infty)P(X < \infty, Y = y_j),$$

ou seja

$$f(x_i, y_j) = f_1(x_i)f_2(y_j).$$

Suponhamos agora que $f(x_i, y_j) = f_1(x_i)f_2(y_j)$ e provemos que X e Y são independentes. De facto,

$$\begin{aligned} P(X \leq x_i, Y \leq y_j) &= \sum_{\ell=1}^i \sum_{k=1}^j P(X = x_\ell, Y = y_k) \\ &= \sum_{\ell=1}^i \sum_{k=1}^j f(x_\ell, y_k) \\ &= \sum_{\ell=1}^i \sum_{k=1}^j f_1(x_\ell)f_2(y_k) \\ &= \left(\sum_{\ell=1}^i f_1(x_\ell) \right) \left(\sum_{k=1}^j f_2(y_k) \right) \\ &= P(X \leq x_i, Y < +\infty)P(X < +\infty, Y \leq y_j). \end{aligned}$$

o que prova o pretendido. \square

Teorema 5.14 *Seja (X, Y) um vector aleatório contínuo e seja f a sua função densidade. Sejam f_1 e f_2 as funções densidade marginais. As v.a., X e Y são independentes se e só se*

$$f(x, y) = f_1(x)f_2(y), \quad (x, y) \in \mathbb{R}^2.$$

Demonstração: Suponhamos que X e Y são independentes. Então

$$F(x, y) = F_1(x)F_2(y)$$

e portanto

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial x} F_1(x) \frac{\partial}{\partial y} F_2(y)$$

o que permite concluir que

$$f(x, y) = f_1(x)f_2(y).$$

Suponhamos, reciprocamente, que as funções densidade verificam $f(x, y) = f_1(x)f_2(y)$. Então

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(t, u) dt du \\ &= \int_{-\infty}^x \int_{-\infty}^y f_1(t) f_2(u) dt du \\ &= \int_{-\infty}^x f_1(t) dt \int_{-\infty}^y f_2(u) du \\ &= F_1(x) F_2(y) \end{aligned}$$

e portanto as v.a. X e Y são independentes. \square

5.3.5 Funções de variáveis aleatórias

Seja X uma v.a. com função de repartição F_X e consideremos $g : \mathbb{R} \rightarrow \mathbb{R}$. Admita-se que $Y = g(X)$ é a v.a. que assume o valor $y = g(x)$ quando $X = x$. Pretende-se determinar a função de repartição F_Y da v.a. Y . Vamos começar por considerar o seguinte exemplo.

Exemplo 5.15 Consideremos dois casos.

1. Seja $g(x) = ax + b$. Então

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right), \quad \text{se } a > 0. \end{aligned}$$

Se $a < 0$ então

$$F_Y(y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a} - 0\right).$$

2. Seja $g(x) = x^2$. Então

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y} - 0), \quad \text{se } y > 0. \end{aligned}$$

e para $y \leq 0$, tem-se $F_Y(y) = 0$.

No caso geral, se X for uma v.a. contínua temos que

$$F_Y(y) = P(Y \leq y) = P(X \in g^{-1}(]-\infty, y]),$$

onde $g^{-1}(]-\infty, y])$ representa a imagem recíproca por g do intervalo $] - \infty, y]$. Nos casos mais correntes, $g^{-1}(]-\infty, y])$ é um intervalo ou uma união de intervalos donde resulta que a probabilidade do último membro se pode exprimir facilmente em termos de F_X .

Se a v.a. for discreta, a v.a. $Y = g(X)$, com g uma função real de variável real, é também discreta. Neste caso, o nosso objectivo pode ser atingido se determinarmos a função de probabilidade f_Y da v.a. Y a partir da função de probabilidade da v.a. X , que denotamos por f_X . Assim, se X for uma v.a. discreta que pode tomar os valores $D = \{x_i, i = 1, 2, \dots\}$, a v.a. $Y = g(X)$ toma valores $g(D) = \{y_i = g(x_i), i = 1, 2, \dots\}$. Então a função de probabilidade da v.a. Y é dada por

$$f_Y(y_j) = P(Y = y_j) = P(g(X) = y_j) = \sum_{x_i: g(x_i)=y_j} f_X(x_i)$$

e, para $y \neq y_j$, $f_Y(y) = 0$.

O que poderemos dizer quanto à esperança da v.a. $Y = g(X)$? No caso discreto, se X é uma v.a. que pode tomar os valores $\{x_i, i = 1, 2, \dots\}$ e f é a sua função de probabilidade, então a esperança de Y é

$$E(Y) = \sum_{x_i} g(x_i)f(x_i).$$

No caso contínuo, se f for a função densidade de X , a esperança de Y é

$$E(Y) = \int_{\mathbb{R}} g(x)f(x)dx.$$

5.3.6 Esperança matemática de um vector aleatório

Definição 5.16 (Esperança) *Seja (X, Y) uma v.a. bidimensional, $E(X)$ a esperança de X e $E(Y)$ a esperança de Y . Ao vector $(E(X), E(Y))$ chamamos esperança matemática ou valor esperado do vector aleatório (X, Y) e que é denotado por $E(X, Y)$.*

Se (X, Y) uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$ e f é a sua função de probabilidade conjunta então

$$\begin{aligned} E(X, Y) &= \left(\sum_{x_i} x_i f_1(x_i), \sum_{y_j} y_j f_2(y_j) \right) \\ &= \left(\sum_{x_i} \sum_{y_j} x_i f(x_i, y_j), \sum_{x_i} \sum_{y_j} y_j f(x_i, y_j) \right). \end{aligned}$$

Se (X, Y) uma v.a. contínua e f é a sua função densidade conjunta então

$$\begin{aligned} E(X, Y) &= \left(\int_{\mathbb{R}} x f_1(x) dx, \int_{\mathbb{R}} y f_2(y) dy \right) \\ &= \left(\int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) dy dx, \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) dx dy \right). \end{aligned}$$

Consideremos agora a função g , real de duas variáveis reais, e a v.a. $Z = g(X, Y)$ que assume o valor $z = g(x, y)$ quando $(X, Y) = (x, y)$. Temos então que

$$E(Z) = \int_{\mathbb{R}^2} g(x, y)f(x, y)dx dy.$$

Se (X, Y) é uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, 2, \dots, j = 1, \dots\}$ e f é a função de probabilidade conjunta então

$$E(Z) = \sum_{x_i} \sum_{y_j} g(x_i, y_j) f(x_i, y_j).$$

No caso contínuo temos que, se f é a função densidade da v.a. X , então

$$E(Z) = \int_{\mathbb{R}} g(x, y) f(x, y) dx dy.$$

Vamos demonstrar alguns resultados importantes.

Teorema 5.17 *Seja (X, Y) uma v.a. e $Z = X + Y$. Então $E(Z) = E(X) + E(Y)$.*

Demonstração: Vamos fazer a demonstração apenas para o caso contínuo. O caso discreto fica como exercício. Temos, sucessivamente, que

$$\begin{aligned} E(X + Y) &= \int_{\mathbb{R}^2} (x + y) f(x, y) dx dy \\ &= \int_{\mathbb{R}^2} x f(x, y) dx dy + \int_{\mathbb{R}^2} y f(x, y) dx dy \\ &= E(X) + E(Y), \end{aligned}$$

o que prova o resultado. \square

Teorema 5.18 *Seja (X, Y) uma v.a. e $Z = XY$. Então, se X e Y forem independentes $E(Z) = E(X)E(Y)$.*

Demonstração: Vamos fazer a demonstração apenas para o caso contínuo. O caso discreto fica como exercício. Temos, sucessivamente, que

$$E(XY) = \int_{\mathbb{R}^2} xy f(x, y) dx dy.$$

Como X e Y são independentes, temos que

$$\int_{\mathbb{R}^2} xy f(x, y) dx dy = \int_{\mathbb{R}^2} x f_1(x) f_2(y) dx dy.$$

Logo $E(X, Y) = E(X)E(Y)$, o que prova o pretendido. \square

5.3.7 Variância, covariância e coeficiente de correlação

Sejam X e Y duas v.a.. Se X é independente de Y então $E(XY) = E(X)E(Y)$. Consideremos a diferença

$$\sigma_{XY} = E(XY) - E(X)E(Y).$$

Esta diferença é nula se as v.a. são independentes (o recíproco pode não ser verdadeiro, i.e. pode suceder que a diferença seja nula sem que haja independência das v.a.). Temos ainda que

$$\sigma_{XY} = E((X - E(X))(Y - E(Y))).$$

Consideremos um referencial com origem no ponto $(E(X), E(Y))$ e a quantidade definida por $(x - E(X))(y - E(Y))$. Esta quantidade é positiva se x e y estiverem no primeiro e terceiro quadrantes e é negativa se x e y estiverem no segundo e quarto. Assim, se a probabilidade de ocorrerem valores de X e Y acima das respectivas médias ou abaixo das médias é grande então σ_{XY} é positivo e tem um valor elevado. Por outro lado, se a probabilidade de ocorrerem valores de X acima (ou abaixo) da média e de Y abaixo (ou acima) da média é grande então σ_{XY} é negativo mas é, em valor absoluto, grande. Deste modo concluímos que o coeficiente σ_{XY} mede a dependência das v.a. bem como o seu desvio em relação à média.

Definição 5.19 (Covariância) *Sejam X e Y duas v.a. de esperanças $E(X)$ e $E(Y)$. O coeficiente*

$$\sigma_{XY} = E[(X - E(X))(Y - E(Y))]$$

é designado por covariância entre X e Y e também é denotado por $\text{cov}(X, Y)$.

Vamos demonstrar o seguinte resultado.

Teorema 5.20 *Sejam X e Y duas v.a. e $Z = X + Y$. Então*

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}.$$

Demonstração: Temos, sucessivamente, que

$$\begin{aligned} \sigma_Z^2 &= E[(Z - E(Z))^2] \\ &= E[(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))] \\ &= E[(X - E(X))^2] + E[(Y - E(Y))^2] + 2E[(X)(Y - E(Y))] \\ &= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}, \end{aligned}$$

o que prova o pretendido. \square

À matriz

$$\begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

chamamos matriz de covariância da v.a. (X, Y) .

Os conceitos apresentados para v.a. bidimensionais poderiam ser generalizados para v.a. de dimensão $n \in \mathbb{N}$. Vamos apenas, para já, generalizar o conceito de matriz de covariância.

Definição 5.21 *Consideremos n v.a. X_i e $i = 1, 2, \dots, n$. A matriz*

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \sigma_{X_1 X_3} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \sigma_{X_2 X_3} & \cdots & \sigma_{X_2 X_n} \\ \sigma_{X_3 X_1} & \sigma_{X_3 X_2} & \sigma_{X_3}^2 & \cdots & \sigma_{X_3 X_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \sigma_{X_n X_3} & \cdots & \sigma_{X_n}^2 \end{bmatrix}$$

é designada por matriz de covariância do vector aleatório (X_1, \dots, X_n) .

Para finalizar, vamos definir um conceito de grande interesse prático. Consideremos a quantidade

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

em que σ_{XY} , σ_X e σ_Y são, respectivamente, a covariância de (X, Y) e os desvios padrão de X e Y , respectivamente. A ρ_{XY} chamaremos coeficiente de correlação da v.a. (X, Y) . Como é evidente, se as v.a. X e Y forem independentes $\sigma_{XY} = \rho_{XY} = 0$. Neste caso, e sempre que o coeficiente de correlação for nulo, dizemos que as variáveis X e Y são não correlacionadas. Por outro lado,

$$|\rho_{XY}| \leq 1.$$

De facto, uma vez que, pela desigualdade de Hölder,

$$\sigma_{XY} = E((X - E(X))(Y - E(Y))) \leq [E((X - E(X))^2)]^{1/2} [E((Y - E(Y))^2)]^{1/2} = \sigma_X \sigma_Y.$$

Temos que ρ_{XY} é máximo, em valor absoluto, quando $\sigma_{XY} = \sigma_X \sigma_Y$. Neste caso (quando o coeficiente de correlação $\rho_{XY} = \pm 1$, ou próximo desse valor) dizemos que X e Y são fortemente correlacionadas.

5.3.8 Exercícios

Exercício 5.3.1 Seja F uma função de repartição. Mostre que os seguintes resultados.

1. Para todo o $x \in \mathbb{R}$ tem-se que $0 \leq F(x) \leq 1$.
2. A função F é não decrescente.
3. $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$ e $F(+\infty) := \lim_{x \rightarrow +\infty} F(x) = 1$.
4. Para valores de a e b finitos tais que $a < b$, tem-se que $P(a < X \leq b) = F(b) - F(a)$.
5. A função F é contínua à direita, i.e., $F(x+0) := \lim_{\epsilon \rightarrow 0^+} F(x+\epsilon) = F(x)$.
6. $P(X = x) = F(x) - \lim_{\epsilon \rightarrow 0^-} F(x+\epsilon) := F(x) - F(x-0)$.

Exercício 5.3.2 Seja F uma função de repartição. Mostre que os seguintes resultados.

1. $P(X \leq x) = F(x)$ (por definição).
2. $P(X < x) = F(x-0)$.
3. $P(X > x) = 1 - F(x)$.
4. $P(X \geq x) = 1 - F(x-0)$.
5. $P(a < X \leq b) = F(b) - F(a)$ (exercício anterior).
6. $P(a < X < b) = F(b-0) - F(a)$.
7. $P(a \leq X < b) = F(b-0) - F(a-0)$.
8. $P(a \leq X \leq b) = F(b) - F(a-0)$.

Exercício 5.3.3 Um determinado estabelecimento comercial tem capacidade de vender entre zero e quatro teodolitos num mês. Seja X a v.a. que indica o número de teodolitos vendidos num mês e

k	0	1	2	3	4
$P(X = k)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{8}$

1. Determine a função de distribuição da v.a. X .
2. Qual probabilidade do estabelecimento vender entre 1 e 3 teodolitos num mês?

Exercício 5.3.4 A duração, em milhares de horas, da componente de um tipo de radar, é uma v.a. X com função densidade f tal que $f(x) = 0.1e^{-0.1x}$, se $x > 0$, e $f(x) = 0$, caso contrário. Pretende-se determinar a probabilidade para que uma componente escolhida ao acaso: dure menos de 4×10^3 horas; dure entre 5×10^3 e 10×10^3 horas; dure mais de 15×10^3 horas.

Exercício 5.3.5 Seja (X, Y) uma v.a. cuja função de probabilidade é dada na tabela

x	y	2	4
1		0.2	0.3
2		0.1	0.1
3		0.2	0.1

1. Determine a função de distribuição da v.a. (X, Y) .
2. Calcule $P(X \leq 3, Y \leq 4)$, $P(X \leq 3, 2 < Y \leq 4)$ e $P(1 < X \leq 3, 2 < Y \leq 4)$.

Exercício 5.3.6 Seja (X, Y) uma v.a. que tem por função densidade

$$f(x, y) = \begin{cases} e^{-(x+y)}, & (x, y) \in]0, +\infty[\times]0, +\infty[, \\ 0, & \text{caso contrário.} \end{cases}$$

1. Determine a função de distribuição da v.a. (X, Y) .
2. Calcule $P(X \leq 3, Y \leq 4)$, $P(X \leq 3, 2 < Y \leq 4)$ e $P(1 < X \leq 3, 2 < Y \leq 4)$.

Exercício 5.3.7 Seja (X, Y) uma v.a. cuja função de probabilidade é dada na tabela

x	y	0	1
0		$\frac{1}{3}$	$\frac{4}{15}$
1		$\frac{4}{15}$	$\frac{2}{15}$

1. Determine a função de probabilidade das distribuições marginais de X e Y .
2. Determine a função de distribuição marginal de X e Y .

Exercício 5.3.8 Vai ser atribuído a uma família, através de um sorteio, um apartamento num edifício com três andares, havendo dois tipos de apartamentos em cada andar (tipo A e B). Pretendendo o dono do edifício que a referida família fique, de preferência, com um no primeiro andar e do tipo A , viciou o sorteio de modo que a função probabilidade fosse a seguinte:

	1º andar	2º andar	3º andar
apartamento do tipo A	$\frac{20}{30}$	$\frac{1}{30}$	$\frac{1}{30}$
apartamento do tipo B	$\frac{6}{30}$	$\frac{1}{30}$	$\frac{1}{30}$

1. Determine as distribuições marginais das variáveis em questão.
2. Qual é a probabilidade de sair um apartamento do tipo A ? E qual é a probabilidade de sair o primeiro andar?
3. Diga se as variáveis são dependentes ou independentes.

Exercício 5.3.9 Seja (X, Y) uma v.a. e g e h duas funções reais de duas variáveis reais. Mostre que

$$E(\alpha g(X, Y) + \beta h(X, Y)) = \alpha E(g(X, Y)) + \beta E(h(X, Y)), \quad \alpha, \beta \in \mathbb{R}.$$

Exercício 5.3.10 Considere a v.a. (X, Y) que tem por função densidade

$$f(x, y) = \begin{cases} x + y, & (x, y) \in]0, 1[\times]0, 1[, \\ 0, & \text{caso contrário.} \end{cases}$$

1. Determine a esperança do vector aleatório (X, Y) .
2. Determine $E(X)$, $E(X^2)$, σ_X^2 e σ_{XY} .
3. Diga se v.a. X e Y são ou não independentes.

5.4 Alguns conceitos estatísticos

5.4.1 Introdução

A estatística é a ciência que se preocupa com a colecta, análise, interpretação e apresentação dos dados e tem como objectivo fundamental o estudo de uma população. Pode dividir-se em duas disciplinas: a estatística decritiva que se preocupa com a colecta, análise, interpretação dos dados estatísticos; e a estatística indutiva, também chamada amostral ou inferencial, que é aquela que, partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões, fundamentando-se na teoria das probabilidades.

Pode dizer-se que os objectivos da teoria das probabilidades e da estatística indutiva são, de certo modo, inversos: na primeira, parte-se de um determinado esquema ou modelo e procura calcular-se a probabilidade de certos resultados ou observações; na segunda, parte-se dos resultados ou observações e procura saber-se alguma coisa sobre o esquema ou modelo.

O nosso objectivo nesta secção será, dada uma v.a. X , calcular a sua média e matriz de covariância. Pelo que foi visto, estes parâmetros só podem ser determinados se conhecermos a sua função de probabilidade (no caso discreto) ou a função densidade (no caso contínuo). Não sendo o caso, o processo usado para tal fim consiste em considerar várias concretizações da v.a. X e construir o que chamamos uma amostra de tamanho relevante. Esta amostra permite construir uma variável estatística e o seu estudo permite inferir uma caracterização para a v.a. a que a amostra diz respeito.

5.4.2 Distribuição de frequências

Consideremos uma v.a. X e uma amostra de tamanho n relativamente grande. Sejam $x_i, i = 1, 2, \dots, n$, os valores observados, i.e., seja $A = \{x_i : i = 1, \dots, n\}$ uma amostra. A frequência absoluta de x_i , $F(x_i)$, é o número de vezes que x_i ocorre nos n valores observados na amostra A . A frequência relativa de x_i , $F_r(x_i)$, é igual ao quociente entre a frequência absoluta e o tamanho da amostra, i.e.

$$F_r(x_i) = \frac{F(x_i)}{n}, \quad i = 1, \dots, n.$$

A distribuição das frequências relativas permite determinar aproximações para a probabilidade de um determinado acontecimento. Suponhamos, por exemplo, que pretendemos calcular $P(X = x_i)$. Este valor é aproximado por $F_r(x_i)$. Se pretendemos calcular $P(X \leq x)$ então tomamos para este valor a aproximação $\sum_{x_i \leq x} F_r(x_i)$.

Suponhamos agora que possuímos uma v.a. bidimensional (X, Y) e uma amostra de tamanho $n + m$ $\{(x_i, y_j), i = 1, \dots, n, j = 1, \dots, m\}$. Sejam $F_r(x_i, y_j)$ as frequências relativas de (x_i, y_j) , para todo o i e todo o j . Tal como no caso unidimensional, $P(X = x_i, Y = y_j) \approx F_r(x_i, y_j)$ e $P(X \leq x, Y \leq y) \approx \sum_{x_i \leq x} \sum_{y_j \leq y} F_r(x_i, y_j)$. No entanto, para calcular $P(X = x_i, Y \in \mathbb{R})$ ou $P(X \in \mathbb{R}, Y = y_j)$ necessitamos de determinar as frequências relativas marginais. Assim, para $i = 1, \dots, n$, temos que

$$F_{r,x}(x_i) = F_{r,1}(x_i) = \sum_{j=1}^m F_r(x_i, y_j)$$

é a frequência relativa marginal de x_i e, para $j = 1, \dots, m$,

$$F_{r,y}(y_j) = F_{r,2}(y_j) = \sum_{i=1}^n F_r(x_i, y_j)$$

é a frequência relativa marginal de y_j . Então:

- $P(X = x_i, Y = y_j) \approx F_r(x_i, y_j)$;
- $P(X \leq x, Y \leq y) \approx \sum_{x_i \leq x} \sum_{y_j \leq y} F_r(x_i, y_j)$;
- $P(X = x_i, Y \in \mathbb{R}) \approx F_{r,1}(x_i)$;
- $P(X \in \mathbb{R}, Y = y_j) \approx F_{r,2}(y_j)$;

- $P(X \leq x, Y \in \mathbb{R}) \approx \sum_{x_i \leq x} F_{r,1}(x_i)$;
- $P(X_2 \leq y) \approx \sum_{y_j \leq y} F_{r,2}(y_j)$.

Os dados da amostra podem ser classificados de acordo com alguma propriedade dando assim origem a classes ou categorias. Por exemplo, na contagem das alturas da população portuguesa pode ser conveniente agrupar os dados em classes com, por exemplo, cinco centímetros de amplitude.

A frequência absoluta de cada classe é igual ao número de elementos dessa classe. A frequência relativa de uma classe é o quociente entre a frequência absoluta da classe e o número total de observações. A distribuição de frequências relativas para dados classificados pode ser dada graficamente por um gráfico de barras onde no eixo das abcissas se consideram as classes e no eixo das ordenadas as frequências relativas da classe. Este gráfico é designado por **histograma**.

Suponhamos agora que temos uma amostra de uma v.a. bidimensional (X, Y) ,

$$A = \{(x_i, y_j), i = 1, \dots, n, j = 1, \dots, m\}$$

e que os dados estão agrupados por classes

$$\begin{aligned} & [x_{a_1}, x_{a_2}[, [x_{a_2}, x_{a_3}[, \dots, [x_{a_p}, x_{a_{p+1}}[, \\ & [y_{b_1}, y_{b_2}[, [y_{b_2}, y_{b_3}[, \dots, [y_{b_q}, y_{b_{q+1}}[. \end{aligned}$$

Seja

$$P_{a_i, b_j} = F_r([x_{a_i}, x_{a_{i+1}}[\times [y_{b_j}, y_{b_{j+1}}[).$$

Tal como no caso unidimensional, a distribuição de frequências relativas das diversas classes pode ser dada pelo histograma.

Suponhamos que pretendemos determinar a frequência relativa (marginal) da classe $[x_{a_i}, x_{a_{i+1}}[$ ou da classe $[y_{b_j}, y_{b_{j+1}}[$. Temos então que definir frequências relativas marginais. Assim

$$F_{r,1}([x_{a_i}, x_{a_{i+1}}[) := F_r([x_{a_i}, x_{a_{i+1}}[\times \mathbb{R}) = \sum_j p_{a_i, b_j}$$

é designada por frequência relativa marginal da classe $[x_{a_i}, x_{a_{i+1}}[$ e

$$F_{r,2}([y_{b_j}, y_{b_{j+1}}[) := F_r([y_{b_j}, y_{b_{j+1}}[\times \mathbb{R}) = \sum_i p_{a_i, b_j}$$

é designada por frequência relativa marginal da classe $[y_{b_j}, y_{b_{j+1}}[$.

Podemos assim determinar a distribuição de frequências relativas de uma classe $[x_{a_i}, x_{a_{i+1}}[$ sabendo que o valor observado para Y está na classe $[y_{b_j}, y_{b_{j+1}}[$. Esta distribuição de frequências é designada por distribuição de frequência relativa condicionada e pode ser usada para medir o grau de dependência das v.a.. Esta distribuição pode ser determinada considerando apenas a coluna da distribuição de frequências relativa em que j é fixo.

5.4.3 Média, variância, covariância e coeficiente de correlação de uma amostra

Seja X uma v.a. e consideremos a amostra de tamanho n , $\{x_i, i = 1, \dots, n\}$. A

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

chamamos média (aritmética) da amostra.

Vejamus se este valor é ou não uma boa aproximação para a esperança da v.a. X . Consideremos as n v.a. $X_i, i = 1, \dots, n$, com a mesma distribuição de X e, com estas v.a., definamos

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Como é evidente, esta nova v.a. pode tomar como valor a média da amostra \bar{x} . Mais ainda,

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

e, portanto, \bar{X} é um estimador cêntrico para $E(X)$ e, como tal, qualquer valor que esta variável assumir pode ser tomado como estimativa para $E(X)$. Por este motivo podemos dizer que \bar{X} é também um estimador consistente para $E(X)$ uma vez que, para $\epsilon > 0$, arbitrariamente pequeno,

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - E(X)| \leq \epsilon) = 1$$

e, portanto, quando $n \rightarrow +\infty$, a probabilidade de \bar{X} tomar um valor na proximidade de $E(X)$ é aproximadamente igual a 1. Como \bar{x} é um valor que \bar{X} pode tomar, para n suficientemente grande, \bar{x} é uma boa aproximação de $E(X)$.

Outro parâmetro importante a calcular é variância de uma amostra. Para isso, vamos definir a variância (corrigida) da amostra na forma

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

onde \bar{x} é a média da amostra e n o seu tamanho.

A razão de usarmos s_x^2 em vez de

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

pode ser justificada se mostrarmos que a esperança da v.a.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

com \bar{X} o estimador cêntrico da esperança e X_i , $i = 1, \dots, n$, uma v.a. com a mesma distribuição de X , é igual à variância de X , isto é, $E(S_x^2) = \sigma_X^2$. Neste caso dizemos que S_x^2 é um estimador cêntrico da variância de X e, deste modo, que a variância da amostra s_x^2 é uma estimativa cêntrica σ_X^2 .

Temos sucessivamente

$$\begin{aligned}
E(S_x^2) &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^n ((X_i - E(X)) - (\bar{X} - E(X)))^2\right] \\
&= \frac{1}{n-1} \sum_{i=1}^n E[(X_i - E(X))^2] - 2E[(X_i - E(X))(\bar{X} - E(X))] \\
&\quad + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
&= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{2}{n-1} E\left[\sum_{i=1}^n X_i - nE(X)\right](\bar{X} - E(X)) \\
&\quad + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
&= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{2n}{n-1} E[\bar{X} - E(X)]^2 + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
&= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{n}{n-1} E[\bar{X} - E(X)]^2 \\
&= \frac{n}{n-1} \sigma_X^2 - \frac{n}{n-1} E[(\bar{X} - E(X))^2]
\end{aligned}$$

e, atendendo a que as v.a. são independentes,

$$\begin{aligned}
E[(\bar{X} - E(X))^2] &= E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - E(X)\right)^2\right] \\
&= \frac{1}{n^2} \sum_{i=1}^n E[(X_i - E(X))^2] \\
&\quad + \frac{1}{n^2} \sum_{i \neq j} E[(X_i - E(X))(X_j - E(X))] \\
&= \frac{1}{n} \sigma_X^2
\end{aligned}$$

e portanto

$$E(S_x^2) = \sigma_X^2.$$

Observação 5.22 A estimativa \bar{s}_x^2 , embora sendo uma estimativa consistente com a variância da v.a. a que a amostra diz respeito, não é uma estimativa cêntrica para este parâmetro.

Consideremos agora uma v.a. bidimensional (X, Y) e, para esta v.a., uma amostra $A = \{(x_i, y_i) : i = 1, \dots, n\}$. As médias \bar{x} e \bar{y} são estimativas cêntricas, respectivamente, para $E(X)$ e $E(Y)$. Mais ainda, as variâncias das v.a. marginais X e Y são estimadas considerando s_x^2 e s_y^2 . Atendendo a que se trata de um vector aleatório poderemos estar interessados estudar a covariância existente entre X e Y devendo-se para este efeito calcular σ_{XY} .

Consideremos a v.a.

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

em que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Exercício 5.4.1 Prove que S_{XY} é um estimador cêntrico para a covariância entre X e Y , isto é, prove que $E(S_{XY}) = \sigma_{XY}$.

Como a v.a. S_{XY} pode tomar o valor

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

concluimos que s_{xy} é uma estimativa cêntrica para a covariância e portanto este valor é uma boa estimativa para aquele parâmetro. A s_{xy} chamamos covariância da amostra.

Finalmente, temos que uma estimativa cêntrica para o coeficiente de correlação ρ_{XY} pode ser dado por

$$\frac{s_{xy}}{s_x s_y}$$

sendo s_{xy} a covariância da amostra A e s_x e s_y são, respectivamente, os desvio padrão das amostras $\{x_i : i = 1, \dots, n\}$ e $\{y_i : i = 1, \dots, n\}$.

5.4.4 Exercícios

Exercício 5.4.2 Mediu-se várias vezes a altura de um poste com uma fita de aço obtendo-se os seguintes resultados:

Nº da medição	1	2	3	4	5	6	7	8	9	10
Valor obtido	31.29	31.24	31.27	31.26	31.36	31.25	31.26	31.27	31.28	31.24

Determine a média da amostra recolhida bem como a sua variância e desvio padrão.

Exercício 5.4.3 Determine a capacidade esperada de um depósito cilíndrico, sabendo que se fizeram várias medições, todas nas mesmas condições e independentes, da sua altura h e do raio da base r , tendo-se obtido os seguintes resultados (em metros):

altura h	10.30	10.33	10.31
raio da base r	3.62	3.60	3.64

Exercício 5.4.4 Para determinar o comprimento de uma linha poligonal mediram-se independentemente os seus três segmentos tendo-se obtido os seguinte resultados (em metros):

x_i	1.00	1.12	1.01	1.25	1.10
y_i	3.00	3.36	3.03	3.75	3.50
z_i	4.00	4.36	4.03	4.75	4.50

Determine uma estimativa para o comprimento da linha poligonal e para o seu desvio padrão.

5.5 Propriedades dos erros de observação

5.5.1 Os tipos de erros de observação

Quando se efectua uma medição/observação podemos distinguir 3 fases: calibragem do objecto usado na medição; colocação do objecto em posição dese efectuar a medição; observação do valor marcado no objecto. Associados a este processo aparecem, inevitavelmente, erros. As teoria clássicas apresentam os erros como sendo de três tipos.

1. **Erros acidentais ou ocasionais.** Têm origem em acidentes ou descuidos aquando da medição. Por exemplo, apontar ao alvo errado, engano na leitura do resultado, etc. São erros, em geral, de valor significativo e podem ser minimizados atendendo a cuidados simples que podem passar por: planejar as observações com cuidado; efectuar várias medições; usar técnicas simples de validação dos resultados (lógica e bom-senso); testar (e conhecer bem) o material utilizado; repetir a experiência com técnicas diferentes.
2. **Erros sistemáticos.** São os erros que se repetem sempre que a medição é efectuada nas mesmas condições físicas. Não podem, por isso, ser detectados pela repetição sucessiva de experiências. Pode dizer-se que quando existem apenas erros sistemáticos não há nada de errado no processo de observação. Estes erros podem ter origem no observador (problemas de visão, etc), na má calibragem do objecto usado na medição (cada vez menos frequentes), nas condições físicas ou na má escolha do modelo. Neste último caso, o modelo pode ser “corrigido” no sentido de se ter em conta os erros. Para isso, os estudos efectuados na disciplina de Análise Numérica I poderão ser úteis.
3. **Erros aleatórios.** São os erros que permanecem quando são eliminados os erros pertencentes aos dois primeiros grupos. São de pequena amplitude e têm origem desconhecida. Apresentam um comportamento análogo ao das v.a..

Como já foi dito, ao efectuar uma medição, o nosso objectivo é determinar um valor que caracteriza a entidade em estudo. Atendendo ao carácter aleatório da observação, a medição que se realiza é a concretização de uma v.a. que está associada ao conceito teórico que se pretende quantificar. Uma vez que o valor pretendido (valor mais provável da v.a.) é desconhecido realizamos, em geral, várias observações e construímos uma amostra. A média da amostra é uma boa estimativa para a esperança da v.a. e a variância da amostra é um bom indicador do erro cometido ao substituir o valor mais provável da v.a. pela média da amostra.

Como é evidente, a construção da amostra é fundamental para a determinação das estimativas anteriores. A construção dessas amostras é feita realizando um conjunto de observações que podem ser de três tipos.

1. **Directas.** O conceito teórico em estudo a que está associada uma v.a. é quantificado por observação directa da entidade a que o conceito se refere (por exemplo, determinação da amplitude de um ângulo, determinação da medida do comprimento de um lado de um triângulo, etc.).
2. **Indirectas.** O conceito teórico que se pretende quantificar está associada uma v.a. que é função de outras v.a.. Os conceitos teóricos que estão subjacentes a estas últimas v.a. são quantificados por observação directa. Assim obtemos para o primeiro conceito teórico um valor determinado por observação indirecta (por exemplo, mede-se o valor da velocidade e determina-se o deslocamento.).
3. **Condicionadas.** Sobre as observações realizadas existem relações de compatibilidade que são traduzidas, em geral, por relações matemáticas. Assim, estas observações são condicionadas pela referida relação matemática (por exemplo, observação da amplitude dos ângulos internos de um triângulo existe uma relação: a soma das amplitudes é igual a 2π .).

5.6 Precisão: cofactor e peso

Seja $X = (X_1, \dots, X_n)$ uma v.a.. Em geral, a sua média $E(X)$ é tomada em substituição de qualquer valor que a v.a. possa tomar. Ao efectuar essa aproximação, a matriz de covariância (ou a variância, no caso unidimensional) dá-nos um indicador do erro cometido. O conhecimento das estimativas para os erros cometidos é importante para avaliar da precisão do resultado e também para proceder a ajustamentos.

A variância de uma v.a. e a covariância entre duas v.a. são valores que têm significado relativo pois dependem da v.a. a que dizem respeito. Atendendo a este facto, quando se consideram ajustamentos, é geralmente introduzida uma normalização dessas quantidades que é efectuada à custa de uma quantidade denominada variância de referência, que será denotada por σ_0^2 (a raiz quadrada da variância de referência é chamada desvio padrão de referência). Surge assim a noção cofactor.

Consideremos um vector aleatório (X_1, \dots, X_n) e seja σ_{ij} a covariância entre X_i e X_j (se $i = j$ então consideramos a variância). Seja σ_0^2 a variância de referência. A

$$q_{ij} = \frac{\sigma_{ij}}{\sigma_0^2}$$

chamamos cofactor entre X_i e X_j e à matriz

$$Q = \frac{1}{\sigma_0^2} C,$$

sendo $C = [\sigma_{ij}]$ a matriz de covariância, chamamos matriz cofactor. Se a matriz cofactor for invertível, a sua inversa é designada matriz peso e é denotada por W , isto é,

$$W = [w_{ij}] = Q^{-1} = \sigma_0^2 C^{-1}.$$

Observação 5.23 *Atendendo às definições anteriores, podemos fazer as observações seguintes.*

1. O termo “peso” deve-se ao facto de, em geral, um peso “grande” significar “grande precisão” e um peso “pequeno” significar “pouca precisão”. Este facto só é totalmente verdade se as matrizes peso e cofactor forem diagonais.
2. Se nada é dito em contrário, tomamos a variância de referência igual a 1.
3. Atendendo a que a matriz de covariância é uma matriz simétrica então a matriz cofactor e a matriz peso (se existir) são também simétricas.
4. Se as v.a. são não correlacionadas então a matriz de covariância é uma matriz diagonal e, como tal, existe a matriz peso, que é também uma matriz diagonal, de elementos da diagonal principal iguais ao produto da variância de referência pelo inverso da variância de cada v.a., isto é,

$$w_{ii} = \frac{\sigma_0^2}{\sigma_{ii}}.$$

5. Se as v.a. têm coeficientes de correlação iguais a ± 1 então a covariância é calculada facilmente (basta calcular os desvio padrão) e portanto facilmente podemos determinar a matriz cofactor e matriz peso.

5.7 Propagação da média, da variância e covariância

Seja $X = (X_1, \dots, X_n)$ uma v.a. para a qual se conhecem as suas propriedades estocásticas (função densidade ou função de probabilidade) e $Y = (Y_1, \dots, Y_m)$ uma v.a. que depende de X pela relação funcional $Y = g(X)$. O nosso objectivo consiste em determinar as propriedades estocásticas da v.a. Y .

Em aplicações práticas este problema surge, geralmente, de forma mais simplificada. Nesses casos, consideram-se três situações: propagação das médias, isto é, determinar $E(Y)$ conhecendo $E(X)$; propagação dos erros aleatórios, isto é, determinar a matriz de covariância de Y sabendo a matriz de covariância de X ; propagação dos erros sistemáticos, situação estudada em Análise Numérica I com recurso à fórmula de propagação do erro. Vamos apenas estudar os dois primeiros casos.

5.7.1 Propagação das médias

Sabemos que, se $Y = g(X)$ então

$$E(y) = E(g(Y)) \Leftrightarrow E(Y_i) = E(g_i(X_1, \dots, X_n)), \quad i = 1, \dots, m.$$

Temos então (no caso contínuo)

$$E(Y_i) = \int_{\mathbb{R}^n} g_i(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

sendo f a função densidade de $X = (X_1, \dots, X_m)$.

No caso de g ser uma função linear a esperança de Y resulta imediatamente. De facto, consideremos $Y = AX + b$, com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Assim sendo, como a esperança é uma funcional linear,

$$Y = AX + b \Leftrightarrow E(Y) = AE(X) + b$$

e então

$$E(Y_i) = \sum_{j=1}^n a_{ij}E(X_j) + b_i, \quad i = 1, \dots, m.$$

Consideremos agora o caso não linear. Iremos particularizar a apresentação para o caso em que $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$, isto é, $m = n = 2$. A generalização para o caso geral é imediata.

Quando a função $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ é não linear, procede-se à sua “linearização” usando polinómios de Taylor. Como se sabe, dada uma função $y = g(x)$ o seu polinómio de Taylor de grau um em torno de $\bar{x} = (\bar{x}_1, \bar{x}_2)$ é dado por

$$\begin{cases} y_1 \approx g_1(\bar{x}) + \frac{\partial g_1}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \frac{\partial g_1}{\partial x_2}(\bar{x})(x_2 - \bar{x}_2) \\ y_2 \approx g_2(\bar{x}) + \frac{\partial g_2}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \frac{\partial g_2}{\partial x_2}(\bar{x})(x_2 - \bar{x}_2) \end{cases}$$

Temos então, no caso de considerarmos v.a.

$$\begin{cases} Y_1 \approx g_1(E(X)) + \frac{\partial g_1}{\partial x_1}(E(X))(X_1 - E(X_1)) + \frac{\partial g_1}{\partial x_2}(E(X))(X_2 - E(X_2)) \\ Y_2 \approx g_2(E(X)) + \frac{\partial g_2}{\partial x_1}(E(X))(X_1 - E(X_1)) + \frac{\partial g_2}{\partial x_2}(E(X))(X_2 - E(X_2)) \end{cases}$$

Temos então que

$$Y \approx g(E(X)) + J(X - E(X)),$$

onde J é a matriz de Jacobi de g calculada no ponto \bar{x} , isto é,

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(E(X)) & \frac{\partial g_1}{\partial x_2}(E(X)) \\ \frac{\partial g_2}{\partial x_1}(E(X)) & \frac{\partial g_2}{\partial x_2}(E(X)) \end{bmatrix}. \quad (5.1)$$

Podemos então concluir que, aplicando a função esperança,

$$E(Y) \approx g(E(X)). \quad (5.2)$$

Observação 5.24 Notemos que o erro cometido na aproximação (5.2) é da ordem da variância da v.a. X .

5.7.2 Propagação das variâncias e covariâncias

Iremos considerar o caso em que $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$, isto é, $m = n = 2$. A generalização para o caso geral é imediata.

Começemos por considerar o caso linear, isto é,

$$Y = AX + b \Leftrightarrow \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Então,

$$\begin{aligned} \sigma_{Y_1}^2 &= E[(Y_1 - E(Y_1))^2] \\ &= E[(a_{11}(X_1 - E(X_1)) + a_{12}(X_2 - E(X_2)))^2] \\ &= a_{11}^2 E[(X_1 - E(X_1))^2] + 2a_{11}a_{12}E[(X_1 - E(X_1))(X_2 - E(X_2))] + a_{12}^2 E[(X_2 - E(X_2))^2] \\ &= a_{11}^2 \sigma_{X_1}^2 + 2a_{11}a_{12}\sigma_{X_1 X_2} + a_{12}^2 \sigma_{X_2}^2. \end{aligned}$$

De igual modo podemos concluir que

$$\sigma_{Y_2}^2 = a_{21}^2 \sigma_{X_1}^2 + 2a_{21}a_{22}\sigma_{X_1X_2} + a_{22}^2 \sigma_{X_2}^2.$$

Relativamente à covariância $\sigma_{X_1X_2}$ temos que

$$\sigma_{Y_1Y_2}^2 = E[(Y_1 - E(Y_1))(Y_2 - E(Y_2))],$$

e, por um raciocínio análogo, temos que (prove)

$$\sigma_{Y_1Y_2}^2 = a_{11}a_{21}\sigma_{X_1}^2 + (a_{11}a_{22} + a_{21}a_{22})\sigma_{X_1X_2}.$$

Atendendo aos resultados anteriores não é difícil concluir que, sendo

$$C_x = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma_{X_2}^2 \end{bmatrix}, \quad C_Y = \begin{bmatrix} \sigma_{Y_1}^2 & \sigma_{Y_1Y_2} \\ \sigma_{Y_1Y_2} & \sigma_{Y_2}^2 \end{bmatrix}$$

as matrizes de covariância de X e Y respectivamente, se tem

$$C_Y = AC_X A^T.$$

Para o caso geral temos o seguinte teorema.

Teorema 5.25 *Seja X uma v.a. de dimensão n da qual se conhecem as suas propriedades estocásticas e Y a v.a. de dimensão m dada pela relação $Y = AX + b$, com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Sejam C_X e C_Y , respectivamente, as matrizes de covariância de X e Y . Então*

$$C_Y = AC_X A^T.$$

Por outro lado, sendo Q_X e Q_Y as matrizes cofactor das v.a. X e Y , respectivamente,

$$Q_Y = AQ_X A^T.$$

Observação 5.26 *Notemos que:*

1. *mesmo que C_X seja diagonal, isto é, mesmo que as v.a. X_i , $i = 1, \dots, n$, sejam não correlacionadas, a matriz C_Y é, em geral, uma matriz densa, ou seja, as v.a. Y_i , $i = 1, \dots, n$, estão, geralmente, correlacionadas;*
2. *a determinação da propagação das matrizes peso é deduzida imediatamente do teorema anterior considerando as inversas (caso existam).*

Consideremos agora o caso não linear. Tal como no caso linear, comecemos por considerar a particularização $n = m = 2$. Linearizando a função g temos que

$$Y \approx g(E(X)) + J(X - E(X)),$$

com J a matriz de Jacobi de g calculada em $E(X)$, ou seja, a matriz (5.1). Podemos então aplicar o teorema anterior e demonstrar o seguinte resultado.

Teorema 5.27 *Seja X uma v.a. de dimensão n da qual se conhecem as suas propriedades estocásticas e Y a v.a. de dimensão m dada pela relação $Y = g(X)$, com $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sejam C_X e C_Y , respectivamente, as matrizes de covariância de X e Y . Então*

$$C_Y = JC_X J^T,$$

onde J é a matriz de Jacobi de g calculada em $E(X)$. Por outro lado, sendo Q_X e Q_Y as matrizes cofactor das v.a. X e Y , respectivamente,

$$Q_Y = JQ_X J^T.$$

5.7.3 Exercícios

Exercício 5.7.1 Para a estrela de Rigel (β da constelação de Oriente) efectuaram-se as seguintes observações relativas à sua magnitude aparente m e à sua distância à Terra r (em parsec):

magnitude aparente m	2.00	2.01	2.09	1.99	1.97
distância à Terra r	279.0	279.5	278.9	278.8	279.1

Supondo que as observações são todas independentes, determine:

1. a magnitude absoluta M de Rigel sabendo que

$$2.512^{(M-m)} = \left(\frac{10}{r}\right)^2;$$

2. uma estimativa para o erro com que vem afectado o valor determinado na alínea anterior.

Exercício 5.7.2 Para a execução de uma carta usou-se a seguinte projecção cartográfica (ϕ e λ em radianos; x e y em quilómetros):

$$\begin{cases} x = (\cos \phi_0)\lambda, \\ y = (\cos \phi_0) \ln \tan \left(\frac{\pi}{4} + \frac{\phi}{2}\right), \end{cases}$$

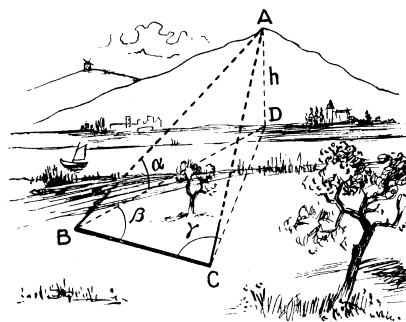
sendo (ϕ, λ) as coordenadas geográficas – latitude e longitude – e ϕ_0 a latitude do paralelo central. Para um paralelo central a 40° , fizeram-se as seguintes determinações (considerando-se independentes), com um receptor GPS, para um mesmo vértice geodésico:

ϕ	$40^\circ 11' 52''.57$	$40^\circ 11' 52''.54$	$40^\circ 11' 52''.54$	$40^\circ 11' 52''.61$
λ	$-8^\circ 26' 37''.11$	$-8^\circ 26' 36''.99$	$-8^\circ 26' 37''.20$	$-8^\circ 26' 37''.15$

Determine as coordenadas rectangulares x e y mais prováveis para esse vértices e os respectivos desvios padrão.

Exercício 5.7.3 Para determinar a diferença de nível h entre os pontos A e B (ver figura), efectuaram-se as seguintes observações do comprimento \overline{BC} (em metros) e dos ângulos α , β e γ (em graus)

\overline{BC}	100.0	100.1	100.1	100.3	100.2
α	25.1	25.1	25.0	25.0	24.9
β	58.5	58.3	58.3	58.4	58.3
γ	74.3	74.5	74.4	74.3	74.4



Supondo que as observações são não correlacionadas, determine:

1. a diferença de nível pretendida sabendo que

$$h = \overline{BD} \tan \alpha;$$

com

$$\overline{BD} = \frac{\overline{BC} \sin \gamma}{\sin (180 - \beta - \gamma)}.$$

2. uma estimativa para o erro com que vem afectado o valor determinado na alínea anterior.

Exercício 5.7.4 Para se determinar uma distância D (metros) em Topografia usa-se muitas vezes a relação

$$D = G \sin^2 z, \quad \text{com } G = 100(l_s - l_i),$$

sendo z (grados) o ângulo zenital e l_s e l_i (metros) leituras numa régua graduada (mira). Suponhamos que foram efectuadas as seguintes medições:

z (grados)	101.2733	101.2736	101.2732
l_s (metros)	1.756	1.754	1.757
l_i (metros)	1.000	0.997	1.001

Determine o valor mais provável para a distância D bem como o respectivo desvio padrão.

Exercício 5.7.5 Observando-se o cometa Zork obtiveram-se, num sistema de coordenadas polares, os seguintes valores para o raio (em U.A.) e o ângulo polar (em graus)

raio ρ	2.7	2.0	1.61	1.2	1.02
ângulo θ	48°	67°	83°	108°	126°

A primeira lei de Kepler estabelece uma relação entre o raio ρ , o ângulo θ e a excentricidade e da elipse através da relação

$$\rho = \frac{p}{1 - e \cos \theta},$$

em que p é um parâmetro real. Determine uma estimativa para o valor mais provável para a excentricidade bem como para o desvio padrão desse valor considerando $p = 1$.

Exercício 5.7.6 Para determinar a medida do perímetro da Terra (p), Eratóstenes (276-195 a.C.) usou a relação

$$p = \frac{2\pi}{\alpha} s,$$

sendo α o ângulo (em radianos) que os raios solares faziam com a vertical do lugar em Alexandria, no solstício de Verão, ao meio-dia, e s a distância de Alexandria a Siena (local onde se sabia que os raios solares, nessa data e nessa hora, incidiam verticalmente).

Suponhamos, por hipótese, que Eratóstenes efectuou as seguintes observações (na realidade ele só efectuou a primeira)

distância s (em estádios)	5000	5020	4980	5007	4991
ângulo α (em graus)	7.2°	7.1°	7.0°	7.3°	7.2°

Considerando as observações não correlacionadas, determine uma estimativa para o valor mais provável do perímetro da Terra obtido por Eratóstenes bem como uma estimativa para o seu desvio padrão (1 estádio = 157.5 m = 0.1575 km).

5.8 O princípio dos mínimos quadrados

5.8.1 Introdução

Na primeira parte deste capítulo introduzimos os conceitos básicos que estão associados ao processo de medição/observação bem como a noção de modelo matemático e alguns tópicos sobre ajustamento. Em particular, foi dito que o ajustamento só tem sentido quando se efectuam observações redundantes.

No sentido estatístico, o ajustamento é um método para determinar estimativas para as variáveis estocásticas e os seus parâmetros de distribuição a partir de uma amostra obtida por observação. De entre os métodos de ajustamento, o dos mínimos quadrados é o mais divulgado.

Desde a sua primeira aplicação a um problema de astronomia por Carl F. Gauss, o método dos mínimos quadrados tem vindo a ser aplicado num vasto conjunto de situações tanto no campo da ciência como no da engenharia. A sua importância prática foi potenciada com o aparecimento dos computadores.

Antes de planear as observações, há que especificar o o modelo funcional que descreve o fenómeno em estudo. Esse modelo é determinado por um certo número de variáveis (parâmetros ou observações) e por relações entre elas. É claro que existe sempre um número mínimo de variáveis que são estritamente necessárias para descrever uma situação física, um acontecimento ou um conjunto de acontecimentos. Para ilustrar este facto consideremos o seguinte exemplo.

Exemplo 5.28 Suponhamos um triângulo que pretendemos identificar através dos seus ângulos. Atendendo a que a soma dos seus ângulos internos é 2π , conhecendo apenas dois ângulos ficamos com a caracterização estabelecida. No entanto podemos considerar como variáveis do modelo funcional os três ângulos internos.

Denotemos por n_0 o número mínimo de variáveis (independentes) estritamente necessárias ao modelo e por n o número total de observações (encaradas como variáveis do modelo). É manifesto que cada variável pode ser concretizada mais do que uma vez, no entanto, iremos supor que cada variável do modelo é apenas observada uma vez. As observações devem ser funcionalmente independentes, isto é, nenhuma das n observações pode ser obtida das restantes $n - 1$. Em geral temos $n > n_0$ e então $r = n - n_0$ é designado por redundância ou número de graus de liberdade. Atendendo a que $r > 0$ é necessário utilizar algum processo que permita determinar um valor ajustado para cada variável.

5.8.2 O ajustamento dos mínimos quadrados

Devido às propriedades estocásticas inerentes às observações, as observações redundantes não são, usualmente, compatíveis com o modelo funcional. Assim, cada conjunto mínimo de variáveis conduz a um resultado diferente. Temos assim que introduzir um critério adicional para obter um resultado único (o “melhor resultado”) que se ajuste ao modelo.

Denotemos por ℓ o vector dos valores observados (que inclui as observações redundantes e inconsistentes com o modelo) e por $\hat{\ell}$ o vector dos valores ajustados e que satisfazem o modelo. Temos que $\dim \ell = \dim \hat{\ell} = n$ mas, em geral, $\ell \neq \hat{\ell}$. À diferença

$$v = \hat{\ell} - \ell$$

chamamos vector dos resíduos ou vector das correcções.

Devido à redundância, pode existir uma infinidade de estimativas para $\hat{\ell}$ (e, conseqüentemente, para v) que satisfaçam o modelo. De entre estas estimativas existe uma que, para além de ser consistente com o modelo, satisfaz o princípio dos mínimos quadrados. Este princípio assegura que os valores de $\hat{\ell}$ estão “o mais próximo possível”, no sentido da distância euclidiana, dos valores da amostra observada ℓ tomando as suas propriedades estocásticas em consideração.

Exemplo 5.29 Consideremos as observações $\ell = (\ell_1, \dots, \ell_n)$ que correspondem aos pontos do plano (x_i, ℓ_i) , $i = 1, \dots, n$, e que pretendemos ajustar a uma recta. A questão que se coloca é a de determinar os valores $\hat{\ell}_i = a + bx_i$, $i = 1, \dots, n$, por forma a que

$$\sum_{i=1}^n (\hat{\ell}_i - \ell_i)^2$$

seja mínimo. Se considerarmos $v = \hat{\ell} - \ell$, com $\hat{\ell} = (\hat{\ell}_1, \dots, \hat{\ell}_n)$ temos que a função a minimizar se pode escrever na forma $v^T v$.

O princípio dos mínimos quadrados consiste em determinar $\hat{\ell}$ (ou v , uma vez que $\hat{\ell} = \ell + v$) por forma a que

$$\phi(v) = v^T W v$$

seja mínima, sendo $W := W_\ell$ a matriz peso das observações. Uma vez que $W = Q^{-1}$ temos que a matriz peso é quadrada de ordem n e os seus elementos traduzem as propriedades estocásticas (tais como a correlação) de todas as observações.

Observação 5.30 *Note-se que a aplicação dos mínimos quadrados não necessita do conhecimento das distribuições associadas às observações. Temos apenas que conhecer a matriz peso W ou a matriz cofactor Q .*

Podemos considerar os seguintes casos particulares:

1. se as observações são não correlacionadas então W é uma matriz diagonal e portanto

$$\phi(v) = \sum_{i=1}^n w_{ii} v_i^2;$$

2. se as observações são não correlacionadas e consideradas com igual precisão (ou igual peso) então $W = I$ e

$$\phi(v) = \sum_{i=1}^n v_i^2 = v^T v.$$

As variáveis do modelo podem não ser (e geralmente não são) caracterizadas nas observações realizadas. Atendendo a este facto, as variáveis que não são observadas são designadas **parâmetros** para as distinguir das observações para as quais sabemos os valores dados pela amostra. Por exemplo, no Exemplo 5.29 as variáveis a e b são parâmetros.

Os valores dos parâmetros são geralmente desconhecidos no início e as suas estimativas são determinadas pelo processo de ajustamento. O vector dos parâmetros é denotado por Δ e $\dim \Delta = \mu$ denota o número de parâmetros.

Relativamente às equações matemáticas entre as variáveis do modelo estas podem ser de dois tipos.

1. **Equações de condição:** são designadas equações de observação as equações envolvendo pelo menos uma observação.
2. **Equações de restrição:** são designadas equações de condição as equações envolvendo apenas parâmetros e constantes.

Neste curso iremos apenas considerar modelos onde aparecem apenas equações de condição. Iremos ainda supor que os parâmetros, tal como as observações, são funcionalmente independentes, isto é, nenhum parâmetro pode ser obtido dos restantes $\mu - 1$.

Observação 5.31 *O vector Δ pode também aparecer nas equações de condição. Assim, o número de condições é maior ou igual a μ .*

As equações de condição ou de restrição podem ser lineares ou não lineares. Iremos apenas estudar o caso linear pois quando as equações são não lineares podemos reduzir este caso ao caso linear, como foi visto na secção anterior.

5.8.3 Ajustamento com equações de condição

Caso geral

Ao iniciar o processo de ajustamento temos que ter em conta os seguintes passos.

1. Especificar número mínimo de observações n_0 .
2. Considerar as n observações no sentido de assegurar que não ocorrerão deficiências.

Neste passo calculamos imediatamente a redundância $r = n - n_0$. Este valor de r significa que entre as n observações deverão existir r condições que deverão ser verificadas.

3. Considerar os parâmetros que se pretendem estimar.

Para cada parâmetro introduzido devemos considerar mais uma condição. Assim sendo, considerando μ parâmetros devemos considerar um total de $c = r + \mu$ condições. Notemos que $0 \leq \mu \leq n_0$ pois, caso contrário, o número de condições excedia o número mínimo de observações o que contradiz o facto dos parâmetros serem funcionalmente independentes. Assim sendo, $r \leq c \leq n$ e $0 \leq \mu \leq n_0$.

4. Especificar as equações de condição que definem o modelo.

Seja Δ o vector dos parâmetros e ℓ o vector das observações. O modelo funcional que iremos considerar é constituído por $c = r + \mu$ condições independentes podem ser lineares ou “linearizadas”. O vector das observações deveria verificar estas equações mas, de facto, o sistema é verificado pelo vector das observações ajustadas $\hat{\ell}$, isto é,

$$A\hat{\ell} + B\Delta = d,$$

ou ainda, sendo v o vector dos resíduos,

$$Av + B\Delta = f$$

com $f = d - A\hat{\ell}$. Notemos que, relativamente ao sistema estabelecido, $A \in \mathbb{R}^{c \times n}$ e $B \in \mathbb{R}^{c \times \mu}$ sendo as suas características iguais a c e μ , respectivamente. Assim sendo, o sistema de equações

$$[A \ B] \begin{bmatrix} v \\ \Delta \end{bmatrix}$$

é indeterminado (tem uma infinidade de soluções) uma vez que a matriz $[A \ B]$ tem característica c .

5. Determinar a melhor solução do sistema $Av + B\Delta = f$ no sentido dos mínimos quadrados.

Atendendo aos passos descritos, o problema que temos que resolver é

$$\begin{cases} \min v^T W v \\ \text{s.a } Av + B\Delta = f \end{cases} .$$

Este problema vai ser resolvido utilizando o método dos multiplicadores de Lagrange. Assim, sendo λ o vector dos multiplicadores de Lagrange ($\lambda \in \mathbb{R}^c$), o problema a minimizar é equivalente a

$$\min \psi(v, \lambda, \Delta) = v^T W v - 2\lambda^T (Av + B\Delta - f).$$

Para resolver este último problema há que determinar as derivadas parciais da função a minimizar em ordem a cada uma das variáveis. Para simplificar a notação, consideremos

$$\frac{\partial \psi}{\partial v}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial v_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial v_n}(v, \lambda, \Delta) \right]^T ,$$

$$\frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial \lambda_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial \lambda_c}(v, \lambda, \Delta) \right]^T ,$$

$$\frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial \Delta_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial \Delta_\mu}(v, \lambda, \Delta) \right]^T .$$

Assim, a função $\psi : \mathbb{R}^{n+c+\mu} \rightarrow \mathbb{R}$ tem um extremo (prova-se que é mínimo) no ponto (v, λ, Δ) que satisfaz as chamadas equações normais

$$\begin{cases} \frac{\partial \psi}{\partial v}(v, \lambda, \Delta) = 0 \\ \frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) = 0 \\ \frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) = 0 \end{cases} .$$

Uma vez que W é uma matriz simétrica, tem-se que (prove)

$$\frac{\partial \psi}{\partial v}(v, \lambda, \Delta) = 2v^T W - 2\lambda^T A.$$

Além disso, também se pode concluir que (prove)

$$\frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) = -2(Av + B\Delta - f)$$

e que

$$\frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) = -2\lambda^T B.$$

Atendendo a estes resultados temos que as equações normais são da forma

$$\begin{cases} -Wv + A^T \lambda = 0 \\ Av + B\Delta = f \\ B^T \lambda = 0 \end{cases}$$

ou, na forma matricial,

$$\begin{bmatrix} -W & A^T & 0 \\ A & 0 & B \\ 0 & B^T & 0 \end{bmatrix} \begin{bmatrix} v \\ \lambda \\ \Delta \end{bmatrix} = \begin{bmatrix} 0 \\ f \\ 0 \end{bmatrix}.$$

Note-se que a matriz do sistema de equações normais é quadrada de ordem $n + c + \mu$. Além disso é uma matriz simétrica e não singular. De facto,

$$\left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & 0 & B & f \\ 0 & B^T & 0 & 0 \end{array} \right] \xrightarrow{L_2 := AW^{-1}L_1 + L_2} \left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & B^T & 0 & 0 \end{array} \right].$$

A matriz AQA^T é quadrada de ordem c e pode ser vista como uma matriz cofactor resultante da propagação de Q pelas equações de condição. Como, além disso, é invertível (W é invertível e a característica de A é c), temos que, denotando a sua inversa por W_c ($W_c := (AQA^T)^{-1}$),

$$\left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & B^T & 0 & 0 \end{array} \right] \xrightarrow{L_3 := B^T W_c L_2 - L_3} \left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & 0 & N & B^T W_c f \end{array} \right]$$

sendo $N = B^T W_c B$. A matriz N é uma matriz quadrada de ordem μ invertível pois W_c é invertível e a matriz B tem característica igual a μ . Como tal, o sistema de equações normais é possível e determinado e a sua solução pode ser obtida do seguinte modo:

1. Resolver $N\Delta = B^T W_c f$;
2. Determinar $\lambda = W_c[f - B\Delta]$;
3. Determinar $v = QA^T \lambda$.

O valor das observações ajustadas é então dado por $\hat{\ell} = v + \ell$. Temos então o seguinte algoritmo para resolver o problema dos mínimos quadrados.

Algoritmo 4.1 Mínimos quadrados: equações do tipo $A\hat{\ell} + B\Delta = d$.

1. $f = d - A\ell$
2. $W_c = (AQA^T)^{-1}$
 $N = B^T W_c B$
3. $\Delta = N^{-1} B^T W_c f$
4. $\lambda = W_c[f - B\Delta]$
5. $v = QA^T \lambda$
6. $\hat{\ell} = \ell + v$

Condições apenas com observações

Suponhamos que o número de parâmetros é $\mu = 0$. Neste caso, o número de condições c é igual à redundância r e as equações (lineares) de condição são da forma

$$A\hat{\ell} = d \Leftrightarrow Av = f,$$

com $f = d - A\ell$. Uma vez que, neste caso, $B = 0$, as equações normais são da forma

$$\begin{bmatrix} -W & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

É fácil concluir, atendendo ao que foi feito no caso geral, que a solução deste problema pode ser dada pelo seguinte algoritmo.

Algoritmo 4.2 Mínimos quadrados: equações do tipo $A\hat{\ell} = d$.

1. $f = d - A\ell$
2. $W_c = (AQAT)^{-1}$
3. $\lambda = W_c f$
4. $v = QA^T \lambda$
5. $\hat{\ell} = \ell + v$

Observações indirectas

Vamos agora considerar o caso em que o ajustamento é efectuado com observações e parâmetros mas com a restrição de que cada condição contém apenas uma observação. Neste caso, como n é o número de observações, temos que $c = n$. A forma geral das equações de condição é

$$\hat{\ell} + B\Delta = d,$$

com $B \in \mathbb{R}^{n \times \mu}$. Uma vez que $\hat{\ell} = v + \ell$, as equações de condição podem ser escritas na forma

$$v + B\Delta = f,$$

com $f = d - \ell$.

As equações normais podem ser obtidas como no caso geral considerando $A = I$. No entanto, neste caso, temos que $v = f - B\Delta$ e assim o problema de minimização fica reduzido a

$$\min \psi(\Delta) = (f - B\Delta)^T W (f - B\Delta).$$

Como

$$\psi(\Delta) = f^T W f - 2f^T W B \Delta + \Delta^T B^T W B \Delta,$$

temos que

$$\frac{\partial \psi}{\partial \Delta}(\Delta) = -2f^T W B + 2\Delta^T B^T W B \Delta.$$

Assim sendo, as equações normais

$$\frac{\partial \psi}{\partial \Delta}(\Delta) = 0$$

são da forma

$$B^T W B \Delta = B^T W f$$

que é um sistema linear possível e determinado de ordem μ . Logo, considerando $N = B^T W B$, temos que $\Delta = N^{-1} B^T W f$. O algoritmo para obter a solução do problema dos mínimos quadrados é, neste caso, o seguinte.

Algoritmo 4.3 Mínimos quadrados: equações do tipo $\hat{\ell} + B\Delta = d$.

1. $f = d - \ell$
2. $N = B^T W_c B$
3. $\Delta = N^{-1} B^T W_c f$
4. $v = f - B\Delta$
5. $\hat{\ell} = \ell + v$

5.8.4 Estimativas para a precisão

Determinação da variância de referência

No final do processo de ajustamento, a variância de referência pode ser dada pela estimativa

$$\sigma_0^2 \approx s_0^2 = \frac{v^T W v}{r},$$

sendo v o vector dos resíduos, W a matriz peso das observações e r a redundância. Note-se que quando considerámos m observações independentes da mesma quantidade, a primeira observação estabelece um valor para a incógnia e as restantes observações são redundantes. Nesse caso, como vimos,

$$s_x^2 = \frac{v^T v}{m - 1}.$$

Determinação da matriz cofactor

Vamos determinar estimativas para as matrizes cofactor tanto das observações ajustadas $\hat{\ell}$ como dos valores dos parâmetros Δ .

Vamos considerar o caso geral, isto é, o caso em que se consideram equações do tipo $A\hat{\ell} + B\Delta = f$. Atendendo ao Algoritmo 4.1 temos o seguinte processo de cálculo.

1. Determinar Q_f . Pelo Teorema 5.25 temos que

$$Q_f = (-A)Q(-A)^T = AQA^T.$$

2. Determinar Q_Δ . Temos, sucessivamente,

$$Q_\Delta = (N^{-1} B^T W_c) Q_f (N^{-1} B^T W_c)^T = N^{-1} B^T W_c Q_f W_c B N^{-1} = N^{-1} B^T W_c B N^{-1}$$

Atendendo à definição da matriz N , temos que $Q_\Delta = N^{-1}$.

3. Determinar Q_λ . Temos que

$$\lambda = W_c [f - B\Delta] = W_c (I - B N^{-1} B^T W_c) f$$

Assim

$$\begin{aligned} Q_\lambda &= W_c (I - B N^{-1} B^T W_c) Q_f (I - B N^{-1} B^T W_c)^T W_c \\ &= (I - W_c B N^{-1} B^T - W_c B N^{-1} B^T + W_c B N^{-1} B^T W_c B N^{-1} B^T) W_c \\ &= (I - W_c B N^{-1} B^T) W_c, \end{aligned}$$

pois $Q_f W_c = I$ e $N^{-1} B^T W_c B = I$. Como tal,

$$Q_v = W_c (I - B N^{-1} B^T W_c).$$

4. Determinar Q_v . Como se pode ver facilmente

$$Q_v = QA^T Q_\lambda (QA^T)^T = QA^T Q_\lambda AQ,$$

ou, o que é equivalente,

$$Q_v = QA^T W_c AQ - QA^T W_c B N^{-1} B^T W_c AQ.$$

5. Determinar $Q_{\hat{\ell}}$. Temos que

$$\begin{aligned} \hat{\ell} &= \ell + v \\ &= \ell + QA^T \lambda \\ &= \ell + QA^T W_c (f - B\Delta) \\ &= \ell + QA^T W_c (I - B N^{-1} B^T W_c) f \\ &= (I - Q_v W) \ell + QA^T Q_\ell d. \end{aligned}$$

Assim,

$$Q_{\hat{\ell}} = (I - Q_v W) Q_\ell (I - Q_v W)^T = Q - 2Q_v + Q_v W Q_v.$$

Ora, como (prove)

$$Q_v W Q_v = Q_v$$

sai que

$$Q_{\hat{\ell}} = Q - Q_v.$$

Podemos resumir o efectuado no seguinte algoritmo.

Algoritmo 4.4 Propagação da matriz cofactor: equações do tipo $A\hat{\ell} + B\Delta = d$.

1. $Q_\Delta = N^{-1}$
2. $Q_v = QA^T W_c AQ - QA^T W_c B N^{-1} B^T W_c AQ$
3. $Q_{\hat{\ell}} = Q - Q_v$

Para os casos particulares, conclui-se, de forma similar, que os algoritmos de propagação da matriz cofactor são os seguintes.

Algoritmo 4.5 Propagação da matriz cofactor: equações do tipo $A\hat{\ell} = d$.

1. $Q_v = QA^T W_c AQ$
2. $Q_{\hat{\ell}} = Q - Q_v$

Algoritmo 4.6 Propagação da matriz cofactor: equações do tipo $\hat{\ell} + B\Delta = d$.

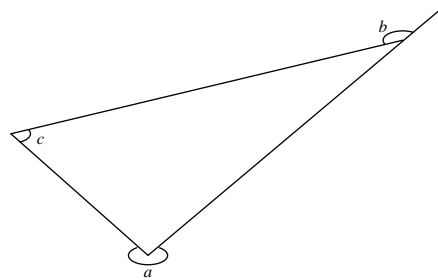
1. $Q_\Delta = N^{-1}$
2. $Q_v = Q - B N^{-1} B^T$
3. $Q_{\hat{\ell}} = Q - Q_v$

5.8.5 Exercícios

Exercício 5.8.1 Mediram-se, com a mesma precisão, os três ângulos internos de um triângulo tendo-se obtido $\alpha = 40^\circ 19' 02''$, $\beta = 70^\circ 30' 0''$ e $\gamma = 69^\circ 11' 01''$. Supondo que as medições são não correlacionadas, calcule:

1. o valor ajustado dos três ângulos;
2. a matriz cofactor;
3. uma estimativa para a variância de referência.

Exercício 5.8.2 Observaram-se os seguintes valores para os ângulos a , b e c do triângulo representados na figura: $a = 320^\circ 19' 40''$, $b = 129^\circ 14' 37''$ e $c = 89^\circ 34' 20''$, sendo dado às observações os pesos $p_a = 3$, $p_b = 4$ e $p_c = 2$. Supondo que as observações são não correlacionadas determine o valor ajustado dos três ângulos

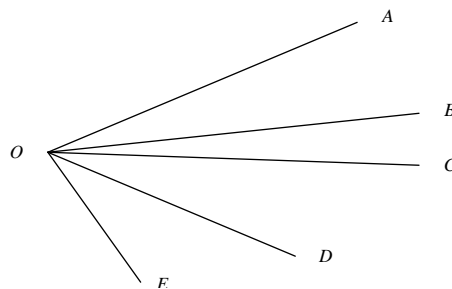


internos do triângulo e a matriz cofactor.

Exercício 5.8.3 Mediram-se os ângulos $\angle AOB$, $\angle BOC$, $\angle AOC$, $\angle COD$ e $\angle COE$ indicados na figura obtendo-se os valores

$\angle AOB$	$\angle BOC$	$\angle AOC$	$\angle COD$	$\angle COE$
$30^\circ 15' 1''$	$20^\circ 00' 00''$	$50^\circ 15' 18''$	$30^\circ 00' 00''$	$70^\circ 00' 01''$

não correlacionados e com igual peso.



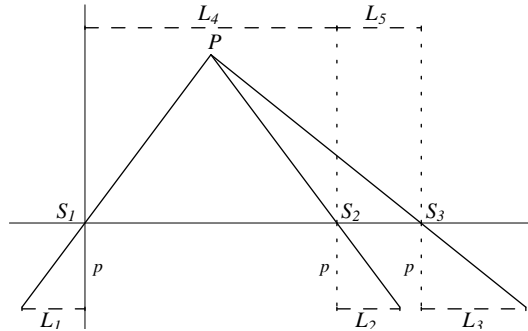
Determine os valores ajustados dos ângulos referidos, a matriz cofactor e uma estimativa para a variância de referência.

Exercício 5.8.4 Para determinar a distância vertical ajustada dos pontos B, C e D a uma plataforma horizontal, situada abaixo destes, mediram-se as seguintes distâncias (em metros):

	$A - B$	$A - C$	$A - D$	$B - C$	$C - D$
Distância vertical	10.216	12.384	15.869	2.138	3.453
Distância horizontal	16	30	18	14	12

Sabendo que a distância vertical do ponto A (ponto mais baixo) a essa plataforma é de 15.000 metros e que o peso de cada distância vertical medida é inversamente proporcional à distância horizontal entre os pontos, determine as distâncias pretendidas.

Exercício 5.8.5 Considere o seguinte esquema que pode representar três câmeras fotográficas colocadas nas posições S_1 , S_2 e S_3 alinhadas sobre o eixo das abscissas e que fotografam um ponto P de coordenadas (x_1, x_2) . Determine os valores ajustados de x_1 e x_2 supondo que é conhecido $p = 100$ mm (sem erro) e ainda as observações



$L_i, i = 1, \dots, 5$, supostas não correlacionadas, e os correspondentes desvios padrão e cujos valores são dados na seguinte tabela:

Observação	Valor	Desvio padrão
L_1	16.5 mm	0.10 mm
L_2	3.8 mm	0.10 mm
L_3	20.4 mm	0.10 mm
L_4	10.0 m	0.05 m
L_5	8.0 m	0.05 m

Exercício 5.8.6 Num nivelamento trigonométrico de alta precisão entre quatro marcos, foram feitas as seguintes medições:

lado	dist. (km)	dN (m)	lado	dN (m)
AB	2.25	+39.274	BA	-39.243
BC	4.00	-94.848	CB	+94.892
CD	1.75	+20.052	DC	-20.032
DA	3.00	+35.619	AD	-35.587

Determine a cota ajustada de cada marco, conhecendo $N_A = 87.631m$.

Exercício 5.8.7 Por forma a estabelecer a cota de 3 marcos, B , C e D , fizeram-se dois nivelamentos, $ABCD$ e $ABDA$; foram registados os seguintes resultados (em metros) :

nivelamento 1: $dN_{AB} = +3.753$, $dN_{BC} = +5.548$, $dN_{CD} = +10.427$, $dN_{DA} = -19.721$;

nivelamento 2: $dN_{AC} = +9.280$, $dN_{CB} = -5.540$, $dN_{BA} = -3.755$;

Determine os valores mais prováveis para as cotas dos marcos sabendo que $N_A = 169.721$ m.

Exercício 5.8.8 Num nivelamento fechado, $ABCD$, mediram-se as seguintes diferenças de nível :

lado	AB	BC	CD	DA
dN (m)	+5.216	+2.394	+1.055	-8.690

Calcule as cotas mais prováveis para B , C e D , sabendo que $N_A = 100.000$ m e considerando as seguintes condições alternativas :

1. Todas as observações têm igual precisão;
2. Os troços BC e CD têm o dobro do comprimento dos troços AB e DA ;

3. Os troços BC e CD foram medidos duas vezes, sendo os valores dados as médias das diferenças de nível obtidas.

Exercício 5.8.9 Numa poligonal foram medidas as seguintes coordenadas relativas :

lado	ΔM (m)	ΔP (m)
AB	1105.362	1346.542
BC	964.547	-965.426
CD	-892.513	-882.492
DA	-1177.341	501.334
BD	72.084	-1847.982

Considerando que todas as medições foram feitas com igual precisão, calcule as coordenadas ajustadas para as estações B , C e D , sabendo que as coordenadas de A são $M_A = 1000.000$ m, $P_A = 1000.000$ m.

Exercício 5.8.10 Mediram-se com a mesma precisão os três lados de um triângulo tendo-se obtido $a = 324.53$ m, $b = 541.41$ m, $c = 483.22$ m, todos com um desvio padrão de 5 cm, e o ângulo oposto ao lado a , $\hat{a} = 40^{\circ}4'25''$ com um desvio padrão de 5''.

Determine:

- o valor ajustado dos ângulos internos;
- a matriz cofactor;
- uma estimativa para a variância de referência.

Exercício 5.8.11 A lei de Hooke estabelece que a força F aplicada a uma mola é directamente proporcional ao deslocamento provocado de acordo com a seguinte relação

$$F = k(e - e_0),$$

onde k é a constante da mola, e o comprimento da mola quando sujeita à força F e e_0 o comprimento inicial da mola.

No sentido de determinar a constante da mola usaram-se diferentes forças (conhecidas) tendo sido observados os comprimentos resultantes, dados na seguinte tabela

força F (em gramas)	3	5	8	10
comprimento e (em milímetros)	13.3	16.3	19.4	20.9

Sabendo que o comprimento inicial da mola é $e_0 = 10$ mm e considerando as medições (não correlacionadas) com precisão inversamente proporcional ao comprimento observado, determine a melhor estimativa para a constante da mola, usando o algoritmo dos mínimos quadrados.

Exercício 5.8.12 Suponhamos que l_1 , l_2 e l_3 são três observações independentes de uma distância Δ . Se as observações tiverem pesos w_1 , w_2 e w_3 , respectivamente, mostre que a melhor estimativa para a distância pretendida, segundo o critério dos mínimos quadrados, é igual à média pesada das observações, i.e.,

$$\Delta = \frac{w_1 l_1 + w_2 l_2 + w_3 l_3}{w_1 + w_2 + w_3}.$$

Exercício 5.8.13 As coordenadas de um ponto y do plano foram medidas por dois métodos distintos tendo sido obtidos os seguintes resultados:

- método 1: $y_1 = \begin{bmatrix} 1.1 \\ 2.0 \end{bmatrix}$, com a matriz peso $W_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$;
- método 2: $y_2 = \begin{bmatrix} 1.0 \\ 2.1 \end{bmatrix}$, com a matriz peso $W_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

Supondo que não existe correlação entre y_1 e y_2 , determine:

- a melhor estimativa para as coordenadas do ponto, usando o algoritmo dos mínimos quadrados;
- a matriz cofactor dos valores obtidos na alínea anterior e uma estimativa para a variância de referência.

Referências bibliográficas

R.L. Burden e J.D. Faires (1988), *Numerical Analysis*, 4th ed., PWS-Kent, Boston.

B. Fornberg (1988), *Generation of finite difference formulas on arbitrarily spaced grids*, *Math. Comp.*, 51, 699-706.

E.M. Mikhail e F.A. Ackerman, (1976), *Observations and Least Squares*, John Wiley & Sons, Washington DC.

B. Murteira (1979), *Probabilidades e Estatística*, Mc-Graw Hill, Lisboa.

H. Pina (1995), *Métodos Numéricos*, Mc Graw-Hill, Lisboa.

P.M. Prenter (1976), *Splines and Variational Methods*, John Wiley & Sons, New York.

M. Rosa (1992), *Tópicos de Análise Numérica*, Dep. Matemática, Univ. Coimbra.

Conteúdo

1	Preliminares	3
1.1	Introdução	3
1.2	Erro absoluto e erro relativo	3
1.3	Condicionamento de matrizes	4
1.4	O polinómio de Taylor	6
1.5	Exercícios	8
2	Interpolação polinomial de funções de uma variável	10
2.1	Introdução	10
2.2	Interpolação polinomial de Lagrange	10
2.2.1	Fórmula de Lagrange	11
2.2.2	Fórmula de Newton	13
2.2.3	Erro de interpolação	14
2.2.4	Zeros dos polinómios de Chebyshev	17
2.2.5	Exercícios	18
2.3	O polinómio interpolador de Lagrange segmentado	21
2.3.1	Caso linear	21
2.3.2	Caso quadrático	23
2.3.3	Caso geral	26
2.3.4	Exercícios	26
2.4	Interpolação de Hermite	28
2.4.1	Existência e unicidade	28
2.4.2	Erro de interpolação	30
2.4.3	Interpolação de Hermite segmentada	31
2.4.4	Polinómios osculadores	32
2.4.5	Exercícios	35
2.5	Interpolação com funções <i>spline</i>	37
2.5.1	Abordagem clássica	37
2.5.2	Funções de base	38
2.5.3	Estudo do erro	42
2.5.4	Exercícios	42
2.6	Derivação numérica	43
2.6.1	Aproximação da primeira derivada	43
2.6.2	Aproximação da segunda derivada. Algumas fórmulas	46
2.6.3	Aproximação de derivadas de ordem superior	47
2.6.4	Exercícios	47
3	Interpolação de funções definidas em \mathbb{R}^2	48
3.1	Introdução	48
3.2	Polinómio interpolador de Lagrange	48
3.3	Interpolação de Lagrange segmentada	52
3.4	O polinómio interpolador de Hermite	54

4	Interpolação de curvas paramétricas	56
4.1	Interpolação paramétrica	56
4.2	Interpolação cúbica de Hermite segmentada	57
4.3	Curvas de Bézier	58
4.4	Exercícios	59
5	Observações e mínimos quadrados	61
5.1	Introdução	61
5.2	Modelo matemático e modelo estocástico	61
5.3	Alguns conceitos probabilísticos	62
5.3.1	Distribuições unidimensionais	62
5.3.2	Distribuições bidimensionais	64
5.3.3	Distribuições marginais	66
5.3.4	Independência	67
5.3.5	Funções de variáveis aleatórias	69
5.3.6	Esperança matemática de um vector aleatório	70
5.3.7	Variância, covariância e coeficiente de correlação	71
5.3.8	Exercícios	73
5.4	Alguns conceitos estatísticos	75
5.4.1	Introdução	75
5.4.2	Distribuição de frequências	75
5.4.3	Média, variância, covariância e coeficiente de correlação de uma amostra	76
5.4.4	Exercícios	79
5.5	Propriedades dos erros de observação	79
5.5.1	Os tipos de erros de observação	79
5.6	Precisão: cofactor e peso	80
5.7	Propagação da média, da variância e covariância	81
5.7.1	Propagação das médias	81
5.7.2	Propagação das variâncias e covariâncias	82
5.7.3	Exercícios	84
5.8	O princípio dos mínimos quadrados	85
5.8.1	Introdução	85
5.8.2	O ajustamento dos mínimos quadrados	86
5.8.3	Ajustamento com equações de condição	87
5.8.4	Estimativas para a precisão	91
5.8.5	Exercícios	93