

MÉTODOS ESTATÍSTICOS EM FÍSICA EXPERIMENTAL

IFUSP – 2015

1. A Medida

a) Algarismos significativos

Num livro de Física Geral está escrito que a constante gravitacional vale $G = 6,67 \times 10^{-11} \text{ m}^3/(\text{kg}\cdot\text{s}^2)$. Essa constante, no entanto, é conhecida hoje com uma precisão da ordem de uma parte em $7,5 \times 10^5$.

- i) O que o autor do livro está dizendo quando escreve o valor: $6,67 \times 10^{-11} \text{ m}^3/(\text{kg}\cdot\text{s}^2)$?
- ii) O valor experimental da carga do elétron é $1,6021773 \times 10^{-19} \text{ C}$. Escreva esse valor com 6, 5, 4, 3, 2 e 1 algarismo significativo.
- iii) O raio de um círculo é 1,18 cm. Qual a área desse círculo? O raio de outro círculo é 0,564190 cm. Qual sua área?
- iv) Certa grandeza tem o valor 0,333. Escreva este valor na base 3 com um número correto de algarismos significativos.

b) Obtenção de dados em uma medida

Discuta a seguinte afirmação:

"Em uma experiência, é útil medir múltiplas vezes cada grandeza desejada. Isso porque os vários dados obtidos podem não ser iguais. Se acreditamos igualmente em todos eles, a média deve ser a estimativa mais confiável."

Discuta também esta outra afirmação:

"A obtenção de muitos dados correspondentes à mesma grandeza medida só é realmente útil se esses dados forem independentes." (*Procure deixar claro o que significa o termo independente acima. Pense em situações nas quais os dados podem não ser independentes.*)

c) Erros

Neste estágio do curso, o importante é entender as diferenças entre esses dois tipos de erros. Em particular, erros estatísticos são só aqueles originados em variações incontroláveis do sistema de medida (objeto + equipamento + observador) que afetam o valor observado da grandeza.

É possível, entretanto, colocar um pouco de matemática na questão. Chamemos de x_0 o valor verdadeiro da grandeza que se deseja medir e de x_i um dado obtido do valor dessa grandeza. Então,

$$x_i = x_0 + \varepsilon_i,$$

onde ε_i é o erro.

Se $\langle \varepsilon_i \rangle = 0$ (leia-se o valor esperado de ε_i igual a zero), não há erro sistemático devido à medida e os eventuais erros são devidos a flutuações incontroláveis do sistema, ou seja, à flutuação estatística. A grandeza

$$\sigma^2 = \langle \varepsilon_i^2 \rangle$$

mede a "*intensidade*" da flutuação estatística e depende do processo de observação e eventualmente do método de análise. Aliás, adiante, veremos que o método de análise pode também gerar erros sistemáticos.

Você consegue correlacionar a existência de erros sistemáticos com a existência de relações de dependência entre os dados obtidos e as condições em que as medições foram efetuadas?

d) Como representar um resultado

A repetição de uma experiência - como você já viu nos Laboratórios de física geral - leva a resultados diferentes.

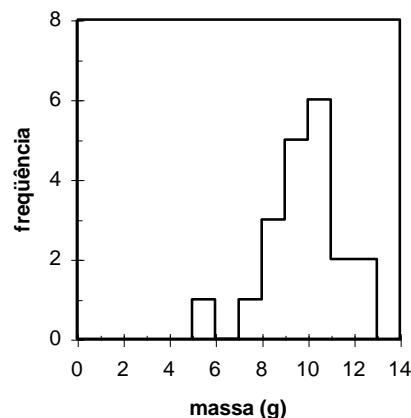
Nessa situação, que é inevitável e inerente ao processo de medida, como podemos representar o resultado? (A resposta está em seguida, mas antes de lê-la, pense e discuta essa questão).

O melhor que podemos fazer é representar a grandeza medida indicando um intervalo onde ela deve estar, com uma certa probabilidade. Por exemplo, o resultado de uma experiência pode levar a uma afirmativa do tipo: "A densidade do material está entre (1,21 e 1,45) g/cm³ com 30% de chance ou está entre (1,217 e 1,440) g/cm³ com 10% de chance ou está entre (1,1 e 1,6) g/cm³ com 80% de chance". Ou seja, podemos optar por intervalos maiores que têm grande chance de conter o valor verdadeiro da grandeza, ou intervalos menores que têm menor chance de conter o valor verdadeiro da grandeza. Note que em qualquer situação o valor da grandeza pode estar ou não no intervalo indicado.

Convença-se de que isso é o melhor que podemos fazer.

e) Histogramas

Histogramas como o da figura abaixo são usados para representar dados obtidos numa medida. A cada intervalo de histogramação (abscissa) chamamos canal.



A escolha dos intervalos afeta o histograma, alterando detalhes - qual é o canal mais alto, quais são o primeiro e o último canais ocupados, etc. Mas não altera certas grandezas de caráter geral, em particular a localização do histograma e a dispersão dos dados.

f) Redução dos Dados

Podemos conhecer o comportamento dos erros estatísticos, o que podemos é "reduzir" os dados, ou seja, a partir da observação de vários dados podemos determinar alguns parâmetros que contenham todo o conhecimento que havia no conjunto de dados original.

Isso é, de certa forma, feito quando você constrói um histograma a partir dos dados de uma tabela. Embora o histograma não permita mais o conhecimento do valor de cada dado que lhe deu origem, ele contém praticamente toda a informação que os dados originais continham.

g) Significado de probabilidade

Não vamos aqui ocupar os 20 anos que J. Bernoulli precisou no início do século XVIII para chegar ao resultado abaixo, que apresentaremos aqui, algo impropriamente, como uma definição de probabilidade.

A probabilidade p é

$$p = \lim_{N \rightarrow \infty} n / N$$

onde n é o número de vezes que se obteve o resultado e N é o número de observações efetuadas. O "resultado" citado acima pode ser: obter um dado num canal do histograma (como exemplo concreto, obter um valor para a massa no intervalo $[8, 9[$ g, referindo-se ao histograma da Fig. 1)

Exercícios

1. Dois experimentadores mediram a frequência da rede elétrica num mesmo local e obtiveram os seguintes dados (em Hz).

Observador 1: (58,50; 56,82; 61,32; 60,30; 59,22; 59,94; 60,48)

Observador 2: (50,22; 51,66; 50,60; 50,64; 52,08; 52,20)

O que você pode dizer sobre a existência de erros estatísticos nas duas medidas efetuadas? E sobre erros sistemáticos?

2. Suponha um experimento no qual o valor de certa grandeza foi observado várias vezes, sendo os resultados os seguintes:

TABELA 1.1

21,002	23,732	24,627	24,903	22,052
23,266	21,995	23,150	22,488	23,390
24,611	22,737	28,367	21,774	24,115
21,497	24,340	22,644	24,299	24,848

Construa um histograma utilizando um intervalo de histograma de uma unidade.

Todos os dados da Tabela 1.1 correspondem a uma mesma grandeza. Suponha que o experimento seja tal que os dados não sejam dependentes uns dos outros. (A dependência a que nos referimos aqui, quer dizer, por exemplo,

que se um dado for, por acaso, maior que o valor verdadeiro da grandeza, então o dado seguinte pode tanto ser maior quanto menor que o valor verdadeiro.) Discuta e responda as seguintes questões:

- i) Olhando para a tabela, você acredita que um ou alguns dados sejam melhores do que outros?
- ii) Se todos os dados correspondem à mesma grandeza, por que então a variação?
- iii) Se todos os dados são equivalentes, qual o valor que melhor representaria a grandeza medida?
- iv) Calcule a média dos 20 dados da Tabela 1.1. Em seguida, considerando apenas o histograma, calcule a média dos dados.

3. Faça os gráficos dos valores tabelados abaixo, no tipo de papel indicado, e obtenha os parâmetros pedidos em cada caso.

TABELA 1.2

x	y	x	y
3	1240	8	2390
4	1470	9	2620
5	1700	10	2850
6	1930	11	4000
7	2160	12	5150

papel milimetrado, $y = ax + b$, $a=?$, $b=?$

TABELA 1.3

x	y	x	y
5	75	25	1875
10	300	30	2700
15	675	35	3675
20	1200	40	4800

papéis milimetrado e di-log, $y = ax^2$, $a=?$

TABELA 1.4

x	y	x	y
1000	$3,68 \times 10^{-1}$	6000	$2,48 \times 10^{-3}$
2000	$1,353 \times 10^{-1}$	7000	$9,12 \times 10^{-4}$
3000	$4,98 \times 10^{-2}$	8000	$3,36 \times 10^{-4}$
4000	$1,832 \times 10^{-2}$	9000	$1,234 \times 10^{-4}$
5000	$6,74 \times 10^{-3}$	10000	$4,54 \times 10^{-5}$

papéis milimetrado e mono-log, $y = e^{ax}$, $a=?$

TABELA 1.5

x	y	x	y
1000	1,649	6000	20,09
2000	2,718	7000	33,12
3000	4,482	8000	54,60
4000	7,389	9000	90,02
5000	12,18	10000	148,8

papéis milimetrado e mono-log, $y = e^{ax}$, $a=?$

4. Faça o mesmo que no Exercício 3, escolhendo o papel conveniente para cada caso.

TABELA 1.6

x	y	x	y
1000	25,7	2500	77,25
1500	41,85	3000	96,2
2000	59,10	3500	115,7

$$y = y_0(x/x_0), y_0 = ?, x_0 = ?, y_0/x_0 = ?$$

TABELA 1.7

x	y	x	y
0	1,000	120	0,405
24	0,835	144	0,335
48	0,695	168	0,280
72	0,580	192	0,235
96	0,485		

$$y = e^{-ax}, a = ?$$

2. A Função Densidade de Probabilidade

a) O significado da probabilidade

Apresentamos no roteiro anterior uma "definição" para probabilidade, baseada na frequência média em que um certo evento é observado. Na prática, é impossível medir-se uma probabilidade através dessa definição, mesmo nos casos mais simples. *Você teria paciência de jogar uma moeda milhões de vezes?*

Se pensarmos em um experimento de Física, há duas questões que limitam severamente a possibilidade de repetir a medida: (a) o tempo necessário para observar um fenômeno e (b) o custo da experiência. Assim, estaremos sempre forçados a realizar um número finito e, provavelmente, pequeno de observações e *induzir o resultado* que, portanto, será apenas *o valor mais provável da grandeza* em que estamos interessados. Isso nos obriga a usar intervalos onde há uma probabilidade conhecida de estar o valor - *verdadeiro!* - da grandeza. Não percamos de vista que pode também acontecer do valor verdadeiro estar fora do intervalo que apresentamos. É uma questão de chance obter um valor próximo ou afastado do valor verdadeiro, sendo obrigação do físico relatar seu resultado como pode ser deduzido dos dados que obteve, sem selecionar dados ou modificar o resultado em função de suas crenças, por exemplo, na Teoria que está testando ou nos resultados de outros experimentadores.

Esperamos que até o fim do curso fique clara a idéia subjacente de *flutuação estatística*. Esta flutuação determina que o procedimento acima é o único correto e que pode produzir bons resultados para o conjunto dos experimentos realizados no mundo, embora possa às vezes nos fazer crer que nos equivocamos (apenas na aparência!).

b) Erros estatísticos e sistemáticos

Em um experimento em Física, há sempre uma quantidade razoável de fontes de erro. Se essas fontes atuam independentemente e cada uma delas pode provocar um erro "para mais" ou "para menos", ao acaso, e ainda se o fato de uma das fontes ter ocasionado um erro "para mais" não afetar a probabilidade que ela venha a dar, na sua próxima manifestação, um erro também "para mais", então estamos numa situação tal que o experimento está sujeito apenas a erros estatísticos.

No entanto, não é rara a ocorrência de erros sistemáticos: erros sempre presentes e sempre no mesmo sentido. As causas dos erros sistemáticos são variadas: uso incorreto de algum aparelho, instrumento descalibrado, tendência sistemática do experimentador em sempre superestimar a observação de uma escala (por exemplo, por causa da paralaxe na leitura de um equipamento com mostrador e ponteiro), etc.

Não podemos identificar os erros sistemáticos apenas observando a flutuação dos dados. É, por isso, um dever do experimentador identificá-los e buscar reduzi-los ao máximo.

c) Propriedades da *f.d.p.*

Os histogramas "buscam" (ou parecem "buscar") imitar uma determinada forma, que é a da função que governa a obtenção de dados durante a medida. A essa função chamamos **função densidade de probabilidade (*f.d.p.*)**.

O significado mensurável da *f.d.p.* é que

$$P(x \in [x_1, x_2]) = \int_{x_1}^{x_2} f(x) dx,$$

onde $P(x \in [x_1, x_2])$ é a probabilidade de se observar a grandeza com valor no intervalo $[x_1, x_2]$.

Note a propriedade

$$P(x \in [x_1, x_3]) = P(x \in [x_1, x_2]) + P(x \in [x_2, x_3]).$$

Para variáveis aleatórias discretas,

$$P(x = x_1 \text{ ou } x = x_2 \text{ ou } x = x_3 \text{ ou } \dots \text{ ou } x = x_n) = F(x_1) + F(x_2) + F(x_3) + \dots + F(x_n),$$

com $F(x_i)$ sendo a **função probabilidade** associada à variável aleatória discreta.

Existem apenas duas restrições sobre a função f que representa uma *f.d.p.*:

$$1) \int_{-\infty}^{+\infty} f(x) dx = 1$$

$$2) f(x) > 0, \text{ para } x \in U,$$

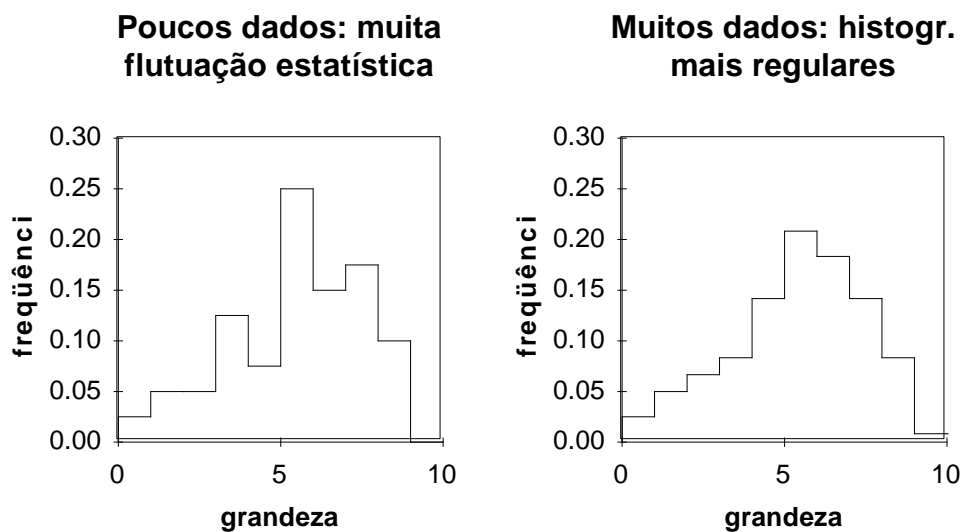
onde U é o espaço amostral dos valores dos dados.

d) A flutuação estatística dos dados

Ao resolver o Exercício 3 deste roteiro, você fez dois histogramas correspondentes a dados de duas medidas da mesma grandeza com o mesmo arranjo experimental. Você deve ter notado que eles têm em comum a região em torno da qual os dados estão distribuídos (afinal correspondem a medidas da mesma grandeza) e a dispersão dos dados (pois o arranjo experimental era o mesmo).

- i) Repetindo-se essa medida outras vezes, ainda com 20 dados e o mesmo equipamento, como você esperaria que fossem os histogramas?
- ii) E se a medida fosse feita com 200 dados? E com 2000 dados?
- iii) Você esperaria que, quanto maior fosse o número de dados, menor seria a irregularidade do histograma?

Vamos discutir essa questão de "irregularidade" de histogramas quando estudarmos a função de probabilidade binomial. Esperamos que você tenha dado a resposta afirmativa ao item *iii*) e que acredite na figura a seguir. Note que o que estamos tratando por "irregularidade" é uma irregularidade relativa. Na realidade, a variação absoluta do número de eventos em um intervalo do histograma aumenta com o número total de eventos. Porém, a variação aumenta mais lentamente que o número total de eventos, fazendo com que a razão diminua com o número de dados.



Por exemplo, se você joga 10 moedas para cima, o número de caras-coroas pode ser (0-10), (1-9), (2-8), (3-7), (4-6), (5-5), (6-4), (7-3), (8-2), (9-1) e (10-0). Se você já jogou moedas (ou outros jogos do mesmo tipo), deve saber que os casos intermediários (com 4, 5 ou 6 caras) ocorrem com razoável frequência, enquanto os casos das extremidades, (0-10), (1-9), (9-1) e (10-0), são bem raros. Ou seja, jogando 10 moedas para cima podemos esperar diferenças entre o número de caras e coroas de cerca de 2 ou 4 (em valor absoluto). Essa é uma flutuação pequena em termos absolutos, mas grande em termos relativos. No entanto, se você jogar 1000 moedas para cima, a diferença esperada entre caras e coroas é da ordem de 30. Isso é grande em termos absolutos, embora seja pequeno relativamente (30 em 1000 corresponde a 3%, enquanto 3 em 10 é 30%).

É esse fenômeno (*pequenos números* \rightarrow *grandes flutuações relativas* e *grandes números* \rightarrow *pequenas flutuações relativas*) que torna, como um exemplo, socialmente interessantes associações de seguros e, como outro exemplo, a Termodinâmica uma ciência que permite descrever sistemas físicos muito complexos a partir de equações e leis simples.

O que você espera que aconteça com um histograma, quando fazemos um experimento com infinitos dados de uma experiência e usando intervalos de histogramação infinitesimais?

A resposta é que o histograma tenderia a uma função contínua. Essa função é a *função densidade de probabilidade* que governa a medida. O seu "centro", muitas vezes o seu ponto mais alto, está relacionado ao valor da grandeza medida. A sua forma está relacionada ao arranjo experimental e muitas vezes basta um único parâmetro - a largura - para defini-la.

e) A *f.d.p.* normal

Quando o número M de fontes de erro tende a infinito e cada uma delas dá um erro de $+e$ ou $-e$, com igual chance e tendendo a zero (mantendo Me^2 finito), a *f.d.p.* é uma gaussiana (ou *distribuição normal*),

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma_0} \exp\left(-\frac{(x-x_0)^2}{2\sigma_0^2}\right),$$

onde x_0 é o valor (verdadeiro) da grandeza medida e σ_0 , o desvio padrão verdadeiro associado ao processo de medida.

Essa *f.d.p.* é utilizada com muita frequência na Estatística e muitos fenômenos são, na prática, regidos por ela. Mas ela nem sempre é a melhor a ser utilizada em certos casos. É preciso refletir um pouco antes de usá-la! Há uma anedota histórica a respeito do uso indiscriminado da distribuição normal, atribuída ao astrônomo Lippman (século XVIII): "...*lei em que todos creem. Os experimentais pensam que é um teorema matemático e os matemáticos, que é um fato experimental.*" (citado em J.-P. Benzécri, *Histoire et Prehistoire de l'Analyse des Données*. Paris, Bordas, 1982.)

f) Parâmetros de uma *f.d.p.*

Os principais são média e variância.

$$x_0 = \text{média} = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\sigma_0^2 = \text{variância} = \int_{-\infty}^{+\infty} (x - x_0)^2 f(x) dx$$

Em geral, pode-se definir *momento central de ordem n* como

$$\mu_n^0 = \int_{-\infty}^{+\infty} (x - x_0)^n f(x) dx$$

e *momento de ordem n* como

$$\mu_n = \int_{-\infty}^{+\infty} x^n f(x) dx.$$

Assim, a média x_0 é $x_0 = \mu_1$ e a variância é $\sigma_0^2 = \mu_2^0 = \mu_2 - \mu_1^2$.

Para a gaussiana, só existem os momentos centrais pares (assim como para qualquer *f.d.p.* simétrica) e todos podem ser calculados a partir de σ_0^2 ,

$$\mu_{2n}^0 = \frac{\sigma_0^{2n}}{2^n} \frac{(2n)!}{n!}$$

Por esta propriedade percebe-se que a função de Gauss é *completamente* definida a partir de apenas dois parâmetros: a média e a variância.

Exercícios

1. O processo de medida é responsável por erros estatísticos. Imagine um experimento simples, no qual observa-se com um cronômetro manual o tempo de queda de um objeto solto também manualmente por um colega. Conhecendo-se a distância de queda, determina-se a aceleração da gravidade local (desprezando-se o atrito com o ar). No caso dessa experiência, escreva a expressão para o valor determinado de g em função da distância e do tempo de queda.

i) Faça uma lista de todos os efeitos que podem fazer com que o valor determinado de g varie.

ii) Suponha que essas fontes de erros sejam independentes umas das outras e que a chance de uma delas dar erro "para mais" seja igual à chance de dar erro "para menos". Neste caso, qual a chance de todas as fontes darem erros "para mais"? E "para menos"? E a de metade delas darem erros "para mais" e metade "para menos"?

iii) Como você esperaria que fosse o histograma dos dados obtidos nessa experiência simples?

2. Indique algumas possíveis causas de erros sistemáticos no caso da medida de g , descrita no Exercício 1 acima.

3. Este exercício é composto de uma sequência de itens encadeados. Resolva-os, portanto, na ordem.

i) Usando os dados da Tabela 2.1, determine o desvio padrão dos dados da experiência, dado por

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}},$$

onde x_i indicam os dados e $N = 20$.

TABELA 2.1 (reprodução da tabela 1.1 do roteiro anterior)

21,002	23,732	24,627	24,903	22,052
23,266	21,995	23,150	22,488	23,390
24,611	22,737	28,367	21,774	24,115
21,497	24,340	22,644	24,299	24,848

ii) Obtenha a expressão adequada para o cálculo de σ , usando as informações contidas no histograma, e determine o seu valor.

iii) A experiência que deu origem aos dados da Tabela 2.1 (Exercício 2 do roteiro anterior) foi repetida, tendo-se obtido os seguintes dados.

TABELA 2.2

23,027	24,408	19,215	21,681	21,918
23,241	21,252	20,053	22,654	22,786
23,384	24,532	20,951	22,375	24,757
23,343	20,429	18,919	22,630	23,841

Usando esses dados, determine sua média, seu desvio padrão e o desvio padrão da média

$$\sigma_m = \sigma / \sqrt{N}.$$

iv) Complete a tabela abaixo:

	\bar{x}	σ	σ_m
1ª experiência			
2ª experiência			

Você acredita que \bar{x} , σ e σ_m de cada experiência possam representar toda a informação contida nos dados originais? Ou seja, determinados esses parâmetros, podemos jogar fora os dados originais sem perder nenhuma informação relevante do experimento?

v) Comparando os resultados das duas experiências, tabelados no item iv), você pode dizer alguma coisa sobre a existência ou não de erros sistemáticos em alguma delas?

vi) Faça um histograma dos dados da 2ª experiência com a mesma escala que você usou ao resolver o Exercício 4 do roteiro 1. Compare os dois histogramas. O que eles têm em comum? Quais as diferenças significativas entre esses histogramas?

4. A probabilidade de um dado cair no intervalo A de um histograma é P_1 ; no intervalo B é P_2 . Os intervalos A e B não têm nenhum ponto em comum.

a) Qual a probabilidade de um dado cair no intervalo A ou no intervalo B ?

b) Qual a probabilidade de, numa medida com dois dados, se obter o primeiro dado em A e o segundo em B ? E o primeiro em B e o segundo em A ?

5. A definição da função de probabilidade poderia ser diferente da utilizada. Tendo em vista, porém, nossa "definição" de probabilidade (que é grosseira, mas muito aceitável! *Ou não?*), discuta por que as seguintes definições não foram preferidas.

$$a) P(x \in [x_1, x_2]) = F\left(\frac{x_1 + x_2}{2}\right)$$

$$b) P(x \in [x_1, x_2]) = F(x_1)$$

$$c) P(x \in [x_1, x_2]) = F(x_2)$$

6. Considere uma medida que esteja governada por uma *f.d.p.* $f(x)$,

$$f(x) = \begin{cases} N, & 0 \leq x \leq 1 \\ 0, & \text{outros valores de } x \end{cases}$$

onde N é uma constante

a) Determine N .

b) Calcule x_0 e σ_0^2 para $f(x)$.

c) Considere uma medida em que os seguintes dados foram obtidos:

TABELA 2.3

0,12	0,60	0,04	0,22	0,10	0,67	0,61	0,65
0,76	0,32	0,96	0,34	0,70	0,45	0,51	0,19
0,35	0,11	0,39	0,97	0,30	0,27	0,09	0,54

Faça um histograma desses dados, escolhendo adequadamente a largura dos canais. Faça na mesma figura um gráfico de $f(x)$, escolhendo adequadamente a escala.

d) Determine \bar{x} e σ a partir dos dados e compare-os com x_0 e σ_0 calculados acima.

e) Complete a tabela abaixo.

(x_1, x_2)	$P(x \in [x_1, x_2])$	fração de dados no intervalo
0,1; 0,2		
0,3; 1,0		
0,5; 1,0		

Por que a fração de dados no intervalo não é exatamente igual à $P(x \in [x_1, x_2])$?

7. Considere um experimento sujeito apenas a 4 fontes de erro independentes, cada uma delas podendo dar um erro $+e$ ou $-e$ com igual probabilidade.

a) Desenhe como você esperaria que fosse o histograma dos dados desse experimento.

b) Suponha agora que o número de fontes de erro fosse muito grande e que cada uma delas gerasse erros infinitesimais. Como seria, então, o histograma?

8. Apresente o resultado da integral

$$\int_{x_0 - \sigma_0}^{x_0 + \sigma_0} f(x) dx$$

para $f(x)$ normal. Considere agora um experimento com 100 dados, governado pela *f.d.p.* normal. Quantos dados você espera que caiam no intervalo $(x_0 - \sigma_0, x_0 + \sigma_0)$? Qual a probabilidade que $x_0 \in (x_i - \sigma_0, x_i + \sigma_0)$, onde x_i é um dado qualquer?

9. A distribuição de velocidades das moléculas de um gás é a conhecida *distribuição de Maxwell*,

$$f(v) = Av^2 e^{-\lambda v^2}$$

onde v é a velocidade das moléculas e $f(v) dv$, a fração de moléculas com velocidades entre v e $v+dv$. Determine \bar{v} para essa função e o valor mais provável de v .

3. E a medida?

a) Os dois caminhos. O valor verdadeiro. Estatística e Estimação

No início do curso, discutimos a questão dos dois caminhos: (a) a partir da experiência, inferimos algo sobre a grandeza medida (que é o que queremos) e (b) a partir da suposição de que conhecemos o valor verdadeiro da grandeza e a resposta do experimento (no caso, os parâmetros da função densidade da probabilidade), deduzimos o que ocorreria num experimento. Algumas vezes fizemos um caminho e outras, o outro. Vamos agora nos concentrar na indução do valor verdadeiro a partir de uma medida.

Suponha que tenha sido feita uma medida e que os dados sejam:

$$\{10,6; \quad 8,1; \quad 9,4; \quad 8,7; \quad 10,7\} \text{ unidades.}$$

Suponha que $f(x)$ que governa a medida seja uma gaussiana. O que podemos fazer nesta altura é estimar o valor de x_0 e de σ_0 pelas equações

$$\bar{x} = \frac{\sum x_i}{N}, \quad \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1} \quad (1)$$

Se σ acima for uma boa estimativa de σ_0 , podemos simplesmente sortear um dos dados, x_i , e dizer simplesmente que a probabilidade de que $x_i \pm \sigma$ cerque o valor verdadeiro é 68%. E o problema está resolvido.

De fato, o problema está resolvido. Mas, se temos (no caso) 5 dados, por que usar um deles, x_i ? Será que a média não é uma estimativa melhor de x_0 , de tal forma que podemos determinar um intervalo em torno dela que ou (a) seja igual a σ mas que tenha maior chance de cercar o valor verdadeiro ou (b) cerque o valor verdadeiro com 68% de chance mas seja menor que σ ?

Discuta essas duas alternativas, (a) e (b) e coleccione argumentos a favor de uma e de outra.

O usual é optarmos pela alternativa (b). A questão que fica pendente neste caso é "qual é o intervalo"? O intervalo é $\sigma_m = \sigma / \sqrt{N}$, onde N é o número de dados da medida. Ele, σ_m , é o desvio padrão da média dos dados e há 68% de chance que $\bar{x} \pm \sigma_m$ cerque o valor verdadeiro. Como σ_m é menor que σ , então a opção por $\bar{x} \pm \sigma_m$ é melhor do que a opção por $x_i \pm \sigma$.

Nós não vamos demonstrar aqui o seguinte teorema. Deixaremos para depois.

"Se os dados obedecem a uma gaussiana com média x_0 e desvio padrão σ_0 , então a média de N dados obedece uma gaussiana com a mesma média x_0 e desvio padrão σ_0 / \sqrt{N} ".

Ao invés de demonstrar o teorema, vamos convencê-lo por exaustão...

b) Estimativas da média, do desvio padrão e do desvio padrão da média

Uma certa grandeza tem o valor $x_0 = 10,00$ e foi medida com um arranjo experimental que corresponde a $\sigma_0 = 1,5$. A medida é governada por uma função densidade de probabilidade normal. Vários estudantes fizeram o experimento e os resultados estão a seguir (todos eles obtiveram 5 dados):

estudantes	dados					\bar{x}	σ	σ_m
1	11,2	10,9	9,8	10,9	9,7			
2	8,8	10,9	11,1	11,2	11,7			
3	9,8	8,5	8,6	13,0	10,6			
4	8,8	7,4	10,4	9,9	10,0			
5	11,1	8,8	13,4	9,7	11,6			
6	10,4	9,5	10,2	10,3	8,0			

- i) Complete a tabela calculando \bar{x} , σ e σ_m para cada estudante.
- ii) Nós estamos tentando lhe convencer das seguintes afirmações (**COMPLETE**):
- 1) A probabilidade que $x_i \pm$ _____ cerque o valor verdadeiro, é _____ %.
 - 2) A probabilidade que _____ $\pm \sigma_m$ cerque o valor verdadeiro é 68%.
 - 3) Se os dados obedecem a uma gaussiana com desvio padrão σ_0 e centrada em x_0 , então as médias obedecem a uma _____ centrada em _____ com desvio padrão _____.
- iii) Sabendo que $x_0 = 10,00$, que fração dos dados de todos os estudantes cerca o valor verdadeiro numa faixa de largura σ em torno deles? Essa fração é compatível com 68%?
- iv) Que fração dos resultados $\bar{x} \pm \sigma_m$ contém x_0 ?
- v) Os valores de σ obtidos pelos vários estudantes são boas estimativas de σ_0 ? Idem para σ_m , como estimativa de $\sigma_0/\sqrt{5}$.
- vi) Você que tem acesso a todos os dados de todos os estudantes, pode determinar o desvio padrão das 6 médias (note que um experimentador não pode fazer isso, pois ele tem conhecimento apenas de sua experiência, que é uma única da tabela). Faça isso e compare com os diversos σ_m obtidos e com $\sigma_0/\sqrt{5}$

c) A flutuação estatística num histograma.

O número de dados obtidos num canal de um histograma pode ser estimado a partir da probabilidade de obter-se um dado nesse canal. Essa probabilidade é

$$P_i = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} f(x) dx,$$

onde P_i é a probabilidade de se obter um dado no canal i , $x_i - \Delta x/2$ e $x_i + \Delta x/2$ são os limites do canal e $f(x)$ é a *f.d.p.* que governa a medida.

Estimamos o número de dados no canal i como

$$n_i \approx N.P_i,$$

onde N é o número total de dados no histograma.

No entanto, sabemos que n_i da fórmula acima é apenas uma indicação do valor que será obtido, sendo que podemos obter alguns (ou vários) dados a mais ou a menos nesse canal. A essa variação do número de dados num canal chamamos "*Flutuação Estatística*", expressão que se aplica também as outras situações onde o resultado obtido pode variar devido a razões estatísticas.

Se não houvesse a flutuação estatística e sempre obtivéssemos exatamente n_i dados no canal i , de acordo com a fórmula acima, será que \bar{x} seria apenas uma estimativa de x_0 , ainda? Ou poderíamos determinar x_0 *exatamente* a partir do histograma?

Exercícios

1. Separe os dados da tabela 3.1 na página seguinte em grupos de dez dados cada.

- Conte quantos dados existem no intervalo $[8, 9[$ em cada um desses grupos. Faça um histograma de número de grupos em função do número de dados no intervalo. Discuta o significado desse histograma.
- Calcule a média do número de dados no intervalo $[8, 9[$. Represente-a no histograma do item (a).
- Calcule o desvio padrão do número de dados no intervalo $[8, 9[$. Represente-o no histograma.

2. Mostre que

$$\sigma_0^2 = \int_{-\infty}^{+\infty} dx f(x) (x - x_0)^2$$

é o valor da variância calculada a partir da função densidade de probabilidade, e pode ser escrita como

$$\sigma_0^2 = \langle x^2 \rangle_0 - (x_0)^2$$

onde

$$\langle x^2 \rangle_0 = \int_{-\infty}^{+\infty} dx f(x) x^2 .$$

Note que usamos uma notação *diferente* daquela do item f) do roteiro 2. Ambas as notações são utilizadas com frequência.

TABELA 3.1

14,37	10,02	15,37	10,19	13,84	11,71	9,74	7,41	10,73	8,11
12,16	11,40	11,18	11,34	12,11	10,05	10,90	10,91	11,62	9,38
10,36	6,21	7,45	9,75	9,93	10,18	10,08	9,41	11,90	8,70
10,40	8,68	10,61	10,45	9,75	11,24	11,33	8,47	9,05	8,72
9,76	8,91	8,78	11,39	9,86	9,82	10,21	10,15	8,99	10,71
9,93	10,24	10,11	9,35	11,12	8,55	9,24	9,81	9,00	10,56
8,31	8,35	9,82	12,94	10,80	12,34	9,25	9,94	10,26	12,96
9,84	7,05	10,32	12,90	12,47	12,93	9,09	8,44	10,15	8,56
10,26	9,65	10,44	11,76	10,14	9,54	9,37	12,40	9,48	8,47
6,99	9,78	9,67	9,32	9,76	7,03	7,51	11,26	10,39	9,83
9,77	10,38	8,42	11,25	10,23	8,82	11,09	12,46	11,61	11,69
7,14	11,53	10,62	11,21	13,26	9,64	6,65	8,97	9,73	11,23
10,25	7,95	10,41	6,91	8,56	9,40	9,12	7,09	13,26	11,07
11,02	9,37	9,52	10,17	12,23	11,74	10,24	9,83	8,77	10,94
12,62	11,75	8,71	9,08	10,65	11,84	12,26	10,57	11,11	8,75
8,39	11,40	12,27	7,20	10,50	11,29	11,99	8,61	10,01	9,71
8,81	10,34	12,03	9,14	11,56	9,64	8,74	9,85	9,90	10,93
10,77	7,42	9,99	9,74	11,44	9,62	11,66	8,13	10,42	9,81
9,13	5,89	9,19	12,96	10,46	11,34	11,20	8,49	7,44	10,93
9,60	9,63	11,30	8,39	8,49	8,81	10,64	8,79	11,97	11,19

4. A Função de Probabilidade Binomial

a) A Binomial para descrever a flutuação estatística num histograma.

Um exemplo da flutuação estatística de dados num histograma pôde ser verificado a partir do Exercício 7 do Roteiro 2. Tínhamos, naquele exercício, quatro fontes de erro que podiam ser $+\varepsilon$ ou $-\varepsilon$ com igual probabilidade (50% para cada sinal). Chamando as fontes de erro de F_i , $i=1, 2, 3$ ou 4 , e supondo x_0 o valor verdadeiro da grandeza que se procura medir, montamos a seguinte tabela:

F_1	F_2	F_3	F_4	valor observado	número de ocorrências
$+\varepsilon$	$+\varepsilon$	$+\varepsilon$	$+\varepsilon$	$x_0+4\varepsilon$	1
$+\varepsilon$	$+\varepsilon$	$+\varepsilon$	$-\varepsilon$	$x_0+2\varepsilon$	4
$+\varepsilon$	$+\varepsilon$	$-\varepsilon$	$+\varepsilon$		
$+\varepsilon$	$-\varepsilon$	$+\varepsilon$	$+\varepsilon$		
$-\varepsilon$	$+\varepsilon$	$+\varepsilon$	$+\varepsilon$		
$+\varepsilon$	$+\varepsilon$	$-\varepsilon$	$-\varepsilon$	x_0	6
$+\varepsilon$	$-\varepsilon$	$+\varepsilon$	$-\varepsilon$		
$-\varepsilon$	$+\varepsilon$	$+\varepsilon$	$-\varepsilon$		
$+\varepsilon$	$-\varepsilon$	$-\varepsilon$	$+\varepsilon$		
$-\varepsilon$	$+\varepsilon$	$-\varepsilon$	$+\varepsilon$		
$-\varepsilon$	$-\varepsilon$	$+\varepsilon$	$+\varepsilon$		
$+\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$x_0-2\varepsilon$	4
$-\varepsilon$	$+\varepsilon$	$-\varepsilon$	$-\varepsilon$		
$-\varepsilon$	$-\varepsilon$	$+\varepsilon$	$-\varepsilon$		
$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$+\varepsilon$		
$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$x_0-4\varepsilon$	1

Um erro total de $+2\varepsilon$ provém da combinação de 3 erros para mais com um erro para menos. Existem 4 combinações das 4 fontes F_i onde uma (e apenas uma) dá o erro para menos. Para erro total zero, 2 fontes dão erro para mais e 2 para menos, existindo 6 combinações das 4 fontes onde 2 (e apenas duas) dão o erro para menos.

A análise combinatória mostra que o número de combinações possíveis de n objetos dentro de um total de N é:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Veja que este resultado foi verificado na tabela que construímos.

Queremos aqui apresentar uma função de probabilidade para casos parecidos com os do exercício que discutimos acima, só que um pouco mais geral. Desejamos guardar o fato de que há apenas dois resultados possíveis da observação do evento, que chamaremos aqui de resultados “A” e “B”. No caso do exercício, “A” seria produzir um erro $+\varepsilon$ e “B” produzir um erro $-\varepsilon$. Outros exemplos seriam: “spin para cima” e “spin para baixo”, “cara” e “coroa”, “cair no intervalo $[x_1, x_2[$ do histograma” e “cair fora do intervalo $[x_1, x_2[$ do histograma”. É preciso explicitar que no caso aqui tratado, esses eventos devem ser mutuamente exclusivos. Não serve “cair no intervalo $] -\infty, x_1[$ do histograma” e “cair no intervalo $]x_2, +\infty[$ do histograma”, onde $x_1 > x_2$. Note que a probabilidade de “A” ($= p$) e a probabilidade de “B” ($= q$) não precisam ser iguais. A condição de que os dois eventos são mutuamente complementares leva ao vínculo,

$$p + q = 1.$$

Cada possível combinação com n eventos “A” e m eventos “B” tem a probabilidade

$$p^n q^m = p^n q^{N-n},$$

onde N é o número total de eventos, $N = n+m$. Como cada combinação de n eventos tipo A e $N-n$ eventos tipo B é independente das outras e igualmente provável, a probabilidade total de se observarem n eventos tipo A num total de N eventos é

$$P_{N,p}(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n},$$

que expressa a *função de probabilidade binomial*.

b) Parâmetros e propriedades da Função de Probabilidade Binomial

Para uma função de probabilidade de variável discreta, a média e a variância são calculadas através de somatórias e não de integrais.

$$\text{média} = n_0 = \sum_{n=0}^N n P_{N,p}(n),$$

$$\text{variância} = \sigma_0^2 = \sum_{n=0}^N (n - n_0)^2 P_{N,p}(n).$$

Substituindo a expressão da função de probabilidade binomial, obtém-se:

$$n_0 = Np,$$

$$\sigma_0^2 = Npq = Np(1-p).$$

Veja a coerência do resultado obtido para a média com a definição da probabilidade p . Observe também o resultado para a *flutuação estatística relativa*,

$$\frac{\sigma_0}{n_0} = \sqrt{\frac{1-p}{Np}} \propto \frac{1}{\sqrt{N}}$$

c) Condições de validade da função de probabilidade binomial.

São três, essas condições:

- 1) O número total de eventos considerados é conhecido e fixo ($= N$).
- 2) Só há dois tipos de eventos possíveis e eles são mutuamente exclusivos e exaustivos.
- 3) A probabilidade de cada tipo de evento é fixa e igual para todas as observações. Essas probabilidades valem p e q , onde $q = 1-p$ por causa da condição 2.

Note que o número de dados em cada canal dos histogramas que vimos construindo nas aulas obedecem a *todas* as 3 condições. 1) o número de dados com que construímos um histograma é conhecido; 2) para um canal específico, cada dado ou cai dentro ou cai fora do intervalo; e 3) a probabilidade do dado cair no intervalo é definida pela *f.d.p.* que governa a medida, que é sempre a mesma para cada dado, desde que o arranjo experimental não se altere com a obtenção dos dados:

$$p = \int_{x_1}^{x_2} f(x) dx,$$

$$q = \int_{-\infty}^{x_1} f(x) dx + \int_{x_2}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} f(x) dx - \int_{x_1}^{x_2} f(x) dx = 1 - p.$$

Podemos deduzir um resultado interessante para os histogramas quando o número de canais é suficientemente grande para que a probabilidade de cair num canal seja pequena. Nesse caso,

$$\sigma_0 = \sqrt{Npq} \cong \sqrt{Np}, \text{ pois } q \cong 1.$$

Como $n_0 = Np$, vemos que $\sigma_0 = \sqrt{n_0}$, ou seja, o desvio padrão é a raiz quadrada do número de dados do canal.

d) Quando há muitos dados...

A *função de probabilidade binomial* $P_{N,p}(n)$ - que dá a probabilidade de se observar n eventos tipo "A" num total de N eventos observados, onde a probabilidade do evento ser do tipo "A" é p - pode ser expressa de maneira mais simples e em função de um único parâmetro, quando

$$N \rightarrow \infty \text{ e } p \rightarrow 0, \text{ com } Np \rightarrow a,$$

sendo a uma constante finita. Com estas condições,

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \approx \frac{N^n}{n!} \quad \text{e} \quad (1-p)^N = \left(1 - \frac{a}{N}\right)^N \rightarrow e^{-a}.$$

E introduzindo estas aproximações na expressão da binomial obtém-se

$$P(n) \rightarrow \frac{N^n p^n}{n!} e^{-a} = \frac{a^n e^{-a}}{n!} = P_a(n),$$

que é uma função de um único parâmetro a .

Essa função de probabilidade $P_a(n)$ é conhecida como *função de Poisson* e é observada em fenômenos tais como: o decaimento radiativo, o número de erros de datilografia por página de um livro, o número de dados em cada canal de um histograma (às vezes apenas aproximadamente) e muitos outros.

Antes de usá-la é preciso verificar as condições de validade, número total de “fontes” *muito grande* e constante, cada uma tendo a mesma probabilidade de produzir o evento detetado. Discutiremos, em seguida, um exemplo ilustrativo dessa dificuldade, como é apresentado no livro de W.T. Eadie *et al.*, *Statistical Methods in Experimental Physics* (North Holland, 1971).

Num experimento de procura de *quarks*, C. McCusker e I. Cairns [Phys. Rev. Lett. **23**, 658 (1969)] mediram o número de bolhas por unidade de comprimento (uc) produzidas por partículas carregadas e rápidas que atravessaram a câmara. Esse número, no caso do experimento, depende do quadrado da carga. Uma partícula que deixasse um traço com muito menos bolhas por uc poderia ser associada ao *quark* que tem carga $(2/3)e$. Eles supuseram que o número de bolhas era distribuído como Poisson - afinal, o número de moléculas atravessadas pela partícula era muito grande e muito maior que o número médio de bolhas formadas (229 bolhas/uc, em média) e a probabilidade de formar uma bolha em cada colisão não depende da molécula específica, elas são idênticas.

Ao todo, foram observados 55 000 traços, um deles com apenas 110 bolhas/uc. Como a probabilidade de obter n bolhas num traço é

$$P_{229}(n) = \frac{229^n e^{-229}}{n!},$$

a probabilidade de obter 110 ou menos traços é

$$P(n \leq 110) = \sum_{n=0}^{110} \frac{229^n e^{-229}}{n!} \approx 1,6 \times 10^{-18},$$

o que implicaria numa chance desprezível de se observar um traço de apenas 110 bolhas/uc em 55 000 dados.

Entretanto, R. Adair e H. Kasha [Phys. Rev. Lett. **23**, 1355 (1969)] lembraram que cada colisão originava 4 bolhas (em média) e que o número de *colisões* é que era distribuído como Poisson. Assim os cálculos deveriam ser baseados no número de bolhas dividido por quatro. Em média, havia portanto $229/4 \cong 57$ colisões e o evento com poucas bolhas tinha provindo de $110/4 \cong 28$ colisões,

$$P(m \leq 28) = \sum_{m=0}^{28} \frac{57^m e^{-57}}{m!} \approx 6,7 \times 10^{-6},$$

mostrando que o traço supostamente “raro” é bastante provável em 55 000 traços... Um cálculo mais exato, levando em conta que essas “4 bolhas/colisão” também têm flutuação estatística, mostra que $P(n \leq 110) \approx 5 \times 10^{-5}$...

e) Parâmetros e propriedades da Poisson

É fácil mostrar que a *f.p.* de Poisson é normalizada,

$$\sum_{n=0}^{\infty} P_a(n) = 1;$$

que o valor médio é a ,

$$n_0 = \sum_{n=0}^{\infty} n P_a(n) = a;$$

e que a variância é

$$\sigma_0^2 = \sum_{n=0}^{\infty} (n-a)^2 P_a(n) = a.$$

Esta última identidade depende do seguinte resultado:

$$\sum_{n=0}^{\infty} n^2 P_a(n) = a \sum_{n=1}^{\infty} \frac{n a^{n-1}}{(n-1)!} e^{-a} = a \sum_{m=0}^{\infty} (m+1) \frac{a^m}{m!} e^{-a} =$$

$$a \sum_{m=0}^{\infty} m \frac{a^m}{m!} e^{-a} + a \sum_{m=0}^{\infty} \frac{a^m}{m!} e^{-a} = a^2 + a.$$

f) Condições de validade da Poisson

São apenas duas e simples. O número de fontes N deve ser muito grande, e esse muito grande pode ser estabelecido em comparação com o número de eventos observados n , ou seja, $N \gg n$. A probabilidade de ocorrer o evento, p , deve ser rigorosamente constante, o que obriga os eventos a serem completamente independentes uns dos outros.

Exercícios

1. Comparar o histograma do item *a*) do Exercício 1 do Roteiro 3 com o de Função de probabilidade binomial com $N = 10$ e p estimado a partir do item *b*). Compare σ do item *c*) com o esperado para essa função de probabilidade binomial.
2. Os dados da Tabela 3.1 do Exercício 1 do Roteiro 3 foram gerados a partir de uma *f.d.p.* normal com média $x_0 = 10,0$ e desvio padrão $\sigma_0 = 1,5$.
 - a) Qual a probabilidade que num grupo de 10 dados, todos estejam na faixa [8,5; 11,5]?
 - b) Qual a probabilidade que *ao menos um* de 20 grupos tenha todos os dados na faixa [8,5; 11,5]?
 - c) Repita os itens *a*) e *b*) para o intervalo [7, 13], sabendo que a probabilidade de um único dado pertencer a esse intervalo é de 95,4%.
3. Um exame consta de 25 questões e para ser aprovado um estudante deverá responder corretamente pelo menos 17 dessas questões. Um determinado estudante sabe 60% do material sobre o qual serão formuladas as questões. Qual a probabilidade desse estudante ser aprovado?

4. Para montar um determinado equipamento são necessários 12 parafusos. Sabe-se que aproximadamente 10% dos conjuntos completos de peças para montagem desse equipamento contém um parafuso defeituoso. Uma loja vendeu 25 conjuntos desse tipo. Qual é a probabilidade de que a loja:

- a) não receba nenhuma reclamação?
- b) receba no máximo duas reclamações?

Quais as suposições que devem ser feitas para poder responder a essas questões?

5. Uma fábrica de automóveis anuncia que um tipo de carro por ela produzido tem apresentado um defeito nos freios com uma frequência aproximadamente igual a 30%. Como se trata de um defeito grave, a fábrica solicita que os proprietários de carros desse tipo levem seus carros às revendedoras para que seja realizada uma inspeção e, caso necessário, a correção do defeito. Uma particular revendedora recebeu da fábrica 10 conjuntos de peças para consertar o defeito e 25 carros para serem inspecionados. Qual é a probabilidade que ela possa consertar todos os carros que apresentarem o defeito?

6. Duas pessoas que sentam ao redor de uma mesa quadrada poderão sentar-se em posições opostas (uma em frente à outra) ou em posições vizinhas (uma ao lado da outra).

- a) Mostre que, se a escolha for feita ao acaso, a probabilidade de que elas se sentem em posições opostas vale $1/3$.
- b) Se 10 casais sentam-se em 10 mesas, qual é a probabilidade de que pelo menos 8 deles sentem-se em posições vizinhas? (Admitindo-se que a escolha é feita ao acaso.)

7. Os passageiros de um avião podem sair pela porta da frente ou pela porta traseira. Se n passageiros vão sair, você acredita que a distribuição binomial para n ensaios e probabilidade de sucesso igual a $1/2$ é adequada para calcularmos as probabilidades relativas ao número de passageiros que sairão pela porta traseira? Por quê?

8. a) Mostre que a função de probabilidade de Poisson apresenta a propriedade:

$$P_a(n+1) = \frac{a}{n+1} P_a(n);$$

- b) mostre que o valor mais provável da distribuição de Poisson, v , nunca é maior que o valor médio, a ;
- c) mostre que a razão $(a - v)/a$ diminui quando a média aumenta.

9. Pode-se definir *assimetria* de uma função de probabilidade (ou de uma f.d.p.) pelo parâmetro,

$$\beta = \frac{\mu_3^0}{(\sigma^2)^{3/2}}, \quad [\mu_3^0 \text{ está definido no item f) do Roteiro 2}]. \text{ Calcule } \beta \text{ para a Poisson.}$$

Mostre que $\beta \xrightarrow{a \rightarrow \infty} 0$, o que é compatível com $P_a(n) \rightarrow$ Gaussiana.

10. Uma fonte radioativa apresenta 2,5 desintegrações por segundo (dps), em média.

- a) Monte uma tabela com a probabilidade de se observar n desintegrações em 1 s.
- b) Monte uma tabela com a probabilidade de se observar n desintegrações em 2 s.

c) Interprete o fato de *duas* desintegrações em 2 segundos serem menos prováveis do que duas desintegrações em 1 segundo.

d) Compreendendo que um intervalo de 2 s é idêntico a 2 intervalos distintos e consecutivos de 1 s, é fácil ver que

$$P_5(0) = [P_{2,5}(0)]^2;$$

$$P_5(1) = [P_{2,5}(0)] \cdot [P_{2,5}(1)] + [P_{2,5}(1)] \cdot [P_{2,5}(0)];$$

$$P_5(2) = \text{soma de três termos, etc.}$$

Generalizando o raciocínio acima, de que n eventos em 2 s correspondem a n_1 eventos no primeiro intervalo de 1 s e $n_2 = n - n_1$ no segundo intervalo, escreva uma expressão para $P_5(n)$ em função de $P_{2,5}(n_1)$ e $P_{2,5}(n_2)$.

e) Qual a probabilidade de em 2 segundos observar-se um número de desintegrações entre $[n_0 - \sigma_0; n_0 + \sigma_0]$?

5. A f.d.p. normal e mudanças de variáveis.

a) A normal como limite da Poisson.

No livro *Tratamento Estatístico de Dados Experimentais*, p. ex., pode ser encontrada uma dedução da distribuição normal como um limite da distribuição binomial. É uma dedução mais adequada pois mantém, desde o início, a independência dos parâmetros *média* e *variância*.

Aqui será apresentada uma “dedução” muito particular da distribuição normal, que serve como ilustração do fato que muitas distribuições têm como limite a gaussiana. Serão feitas algumas aproximações drásticas, pois uma função de variável *discreta* será transformada em função de variável *contínua*. Normalmente isso só pode ser feito com valores inteiros muito grandes, uma vez que diferenças entre valores próximos podem ser consideradas muito pequenas.

A razão da variação da probabilidade $P_a(n)$ relativamente à variação do valor de n pode ser escrita, para valores consecutivos, como:

$$\frac{\Delta P}{\Delta n} = \frac{\Delta P}{(n+1) - n} = P_a(n+1) - P_a(n) = \left(\frac{a^{n+1}}{(n+1)!} - \frac{a^n}{n!} \right) e^{-a} = \left(\frac{a}{n+1} - 1 \right) P_a(n).$$

Transformar n em x implica em que $n \gg 1$. Na distribuição de Poisson, $P_a(n)$ tem valores não desprezíveis na vizinhança de a . Por isso, a aproximação que será desenvolvida baseia-se na suposição que $a \gg 1$. Dessa forma,

$$\frac{\Delta P}{\Delta x} = \frac{a - x - 1}{x + 1} P_a(x).$$

Como $x = n \gg 1$, $x+1 \approx x$, e ainda como os valores de x correspondentes a $P_a(n)$ não nula são muito próximos de a , identificamos x com a no denominador acima. Em conseqüência, já considerando $P(x)$ como uma função de variável contínua,

$$\frac{dP}{dx} = \frac{a - x}{a} P(x).$$

$$\frac{dP}{P} = -\frac{x - a}{a} dx \quad \Leftrightarrow \quad \ln P = -\frac{(x - a)^2}{2a}.$$

$$P = A \exp\left[-\frac{(x - a)^2}{2a}\right].$$

Conclusão: $P_a(n)$ tende a uma Gaussiana com média a e desvio padrão \sqrt{a} , quando a tende a infinito.

b) Por que a normal? Teorema Central do Limite.

A função de probabilidade binomial,

$$P_{N,p}(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n},$$

como sugerido no início deste roteiro, quando N for grande (Na verdade, há mais condições. Veja o livro acima citado.), tende à gaussiana

$$P(n) = \frac{1}{\sqrt{2\pi} \sigma_0} \exp\left[-\frac{(n-n_0)^2}{2\sigma_0^2}\right],$$

onde $n_0 = Np$ e $\sigma_0 = \sqrt{n_0(1-p)}$.

Assim, se houver N fontes de erros infinitesimais, cada uma dando um erro $+\varepsilon$ ou $-\varepsilon$ com igual probabilidade ($p = q = 1/2$), a *f.d.p.* do erro total (em média, zero) será uma Gaussiana.

O *Teorema Central do Limite* generaliza este resultado. A soma de variáveis aleatórias x_i , cada uma obedecendo à *f.d.p.* $f_i(x)$ com variância σ_i^2 e média μ_i ,

$$S = \sum_{i=1}^N x_i,$$

tende a uma normal com média,

$$s_0 = \sum_{i=1}^N \mu_i,$$

e variância,

$$\sigma_0^2 = \sum_{i=1}^N \sigma_i^2,$$

quando $N \rightarrow \infty$. Note que cada $f_i(x)$ pode ser diferente e diferente da gaussiana. Basta que $f_i(x)$ exista, com σ_i finito, e que σ_0^2 também seja finito.

Conclusão: superposição de muitos erros infinitesimais independentes originam um erro total *gaussiano*. Nem sempre, porém, é possível provar esse resultado numa medida específica, e aí ...

c) Mudanças de variáveis.

Se x obedece a *f.d.p.* $f(x)$ e $g(x)$ é uma função de x tal que, para cada $g(x')$ existe um e um só x' que origina esse valor de g , a *f.d.p.* de g , $h(g)$, é

$$h(g) = \frac{f(x)}{\left| \frac{dg}{dx} \right|}.$$

Observe que esta expressão indica a invariância da probabilidade de se encontrar eventos no intervalo $(x, x+dx)$ ou no intervalo $(g, g+dg)$, com $g = g(x)$ e $g+dg = g(x+dx)$, além do fato de *f.d.p.*'s e as probabilidades associadas a intervalos serem sempre positivas. De fato, $h(g) \cdot |dg| = f(x) \cdot dx$.

Exercícios

1. No Exercício 10 do Roteiro 4, você calculou a função de probabilidade de Poisson com média 5, $P_5(n)$. Compare $P_5(n)$ com a aproximação gaussiana adequada. O acordo é bom? **Obs.: Para médias acima de 30, a aproximação gaussiana é excelente.**

2. Usando o Teorema Central do Limite, mostre que $\sigma_m = \sigma/\sqrt{N}$.

3. Em um experimento de medida do equivalente mecânico do calor, os 9 grupos de uma turma relataram os seguintes resultados (em J/cal):

3,90(10); 4,34(15); 4,25(20); 4,31(5); 4,52(20); 4,15(30); 3,87(20); 4,09(5); 4,16(5).

a) Ignore as estimativas dos erros nas grandezas fornecidas pelos alunos. Supondo todas as medidas igualmente precisas, qual é o “melhor” valor que se pode atribuir ao equivalente mecânico do calor? Qual o resultado da medida da classe *toda*?

b) Supondo que todos tenham adotado o mesmo procedimento na medida e devessem ter obtido a mesma estimativa de erro, qual (ou quais) grupos estimaram corretamente o erro?

4. A variável x é distribuída normalmente, em torno de x_0 , com desvio padrão σ_0 . Mostre que a distribuição de $y = x^2$ é

$$G(y) = \frac{\exp[-(\sqrt{y} - \sqrt{y_0})^2 / 2\sigma_0^2]}{\sqrt{2\pi} \, 2\sqrt{y_0} \, \sigma_0},$$

onde $y_0 = x_0^2$.

a) Mostre que, se $\sigma_0 \ll x_0$ (de forma que a *f.d.p.* de x seja desprezível para $x < 0$), podemos aproximar

$$G(y) = \frac{\exp[-(y - y_0)^2 / 8y_0\sigma_0^2]}{\sqrt{2\pi} \, 2\sqrt{y_0} \, \sigma_0}.$$

(Sugestão: expanda o expoente em torno de $y = y_0$.)

b) Mostre que o desvio padrão em y , nesta última equação, é exatamente o valor dado pela expressão da propagação de erro. (Faça esse item após o Roteiro 6.)

5. Para determinar a densidade de um material, foram medidos a massa e o volume de uma amostra, obtendo-se os valores

$$m = (8,5 \pm 1,7) \text{ g} \quad \text{e} \quad v = (6,3 \pm 0,9) \text{ cm}^3.$$

Os valores indicados são a média e o desvio padrão da média. Qual a densidade e seu desvio padrão? O valor verdadeiro da densidade desse material pode ser $1,7 \text{ g/cm}^3$? E $2,0 \text{ g/cm}^3$? Que bases quantitativas tem-se para discutir essas duas últimas questões?

6. Deseja-se determinar o valor da grandeza z , com desvio padrão da média da ordem de σ_m . A grandeza z é determinada a partir de medidas de x e y , sendo $z = x + y$. Por outras medidas feitas, sabe-se que o desvio padrão da série para medidas de x é da ordem de $10 \cdot \sigma_m$ e de y , da ordem de $5 \cdot \sigma_m$.

a) Se forem obtidos 50 dados y_i , quantos dados x_i serão necessários para atingir a precisão desejada?

b) Qual a menor quantidade de dados, incluindo x_i e y_i , necessários para atingir o objetivo almejado?

6. Estimativa da incerteza numa função

a) Propagação de incertezas

Suponha uma função f de várias variáveis,

$$f(a_1, a_2, \dots, a_n),$$

cada variável a_i com desvio padrão σ_i tal que

$$\sigma_i \ll a_i$$

Podemos demonstrar que

$$\overline{f(a_1, a_2, \dots, a_n)} \cong f(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_n)$$

e

$$\sigma_f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial a_i} \right)^2 \sigma_i^2 + 2 \sum_{i>j} \frac{\partial f}{\partial \bar{a}_j} \text{cov}(a_i, a_j)$$

onde

$$\frac{\partial f}{\partial \bar{a}_i} = \frac{\partial f}{\partial a_i} \quad \begin{array}{l} a_1 = \bar{a}_1 \\ a_2 = \bar{a}_2 \\ a_3 = \bar{a}_3 \\ \cdot \\ \cdot \\ \cdot \\ a_n = \bar{a}_n \end{array}$$

Quando σ_i precisa ser menor a_i ? A resposta depende de quanta precisão é necessária nas estimativas de \bar{f} e σ_f . Se o problema não é muito especial e f varia suavemente com a_i (nada como $\exp(a_i)$, por exemplo), então uma regra genérica é

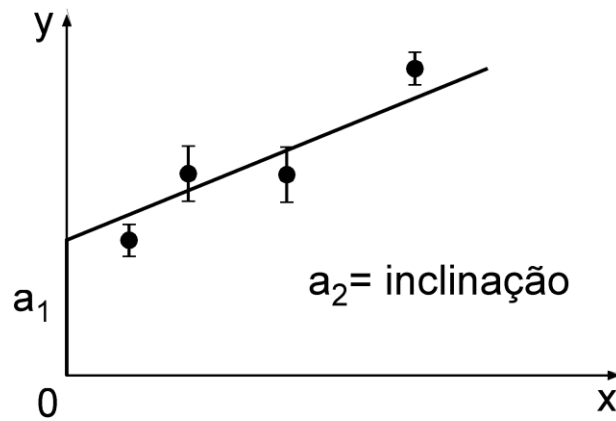
$$\sigma_i < \frac{\bar{a}_i}{4}$$

b) Covariância

Na fórmula de propagação de incertezas apareceu a covariância entre a_i e a_j . No momento, não trabalhamos com variáveis dependentes entre si, mas o faremos. O caso mais corriqueiro é o ajuste de funções, onde os parâmetros ajustados são, quase sempre, covariantes. Por exemplo, no ajuste de uma reta por pontos (x_i, y_i) , os parâmetros de reta

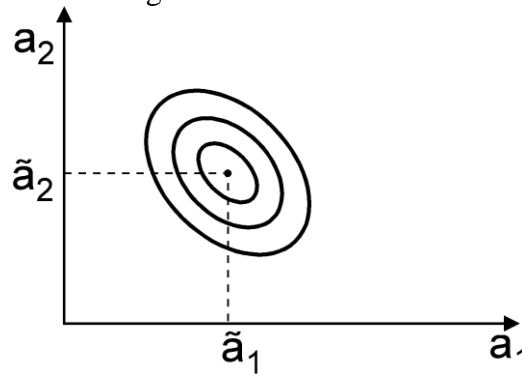
$$y = a_1 + a_2 x$$

são correlacionados.



Note que se a_2 aumenta, a_1 diminui. Assim $\langle (a_1 - \bar{a}_1)(a_2 - \bar{a}_2) \rangle$ é menor que zero

O desenho da função de probabilidade de a_1 e a_2 , $g(a_1, a_2)$, é tridimensional. Desenharemos curvas de nível de g . No desenho, \tilde{a}_1 e \tilde{a}_2 correspondem ao ponto de máximo de g , e as elipses a curvas de nível de g .



c) O desvio padrão da média

Pode ser deduzido como está no livro, na página 41, supondo muitas medidas da mesma grandeza e calculando σ_m como desvio padrão das médias obtidas em cada medida.

Ou então usar

$$\bar{x} = f(x_1, x_2, \dots, x_n) = \frac{1}{N} (x_1 + x_2 + \dots + x_n),$$

onde $\sigma_i^2 = \sigma^2$ e x_i é independente de x_j para $i \neq j$. Notando que

$$\frac{\partial \bar{x}}{\partial x_i} = \frac{1}{N} \quad \text{e} \quad \text{cov}(x_i, x_j) = 0,$$

calcula-se

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^N \left(\frac{1}{N}\right)^2 \sigma^2 = N \frac{\sigma^2}{N^2} = \frac{\sigma^2}{N} = \sigma_m^2,$$

ou $\sigma_m = \frac{\sigma}{\sqrt{N}}$

d) Uso da tabela de integral da normal

Suponha que uma máquina fabrica tubos de ferro com diâmetro interno 20,0 (5) mm. Como podemos estimar a fração de tubos com diâmetro entre 20,7 e 21,1 mm?

Supondo a fdp do diâmetro gaussiana, podemos dizer que

$$f(r) = \frac{1}{\sqrt{2\pi} \cdot 0,5} e^{-\frac{(r-20,0)^2}{2 \cdot 0,5^2}},$$

e daí deduzir que a fração pedida é

$$P(20,7 \leq r \leq 21,2) = \int_{20,7}^{21,2} f(r) dr,$$

A integral definida da gaussiana só pode ser efetuada numericamente. Assim, ou programamos a calculadora ou recorremos à tabela.

As tabelas contêm a integral

$$I(y) = \int_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy,$$

correspondente à integral da gaussiana de média 0 e desvio padrão 1, entre 0 e y. Isso basta, porque a função é simétrica em torno de 0 e a integral num intervalo $[y_1, y_2]$ com $y_2 > y_1 > 0$ pode ser escrita

$$\int_{y_1}^{y_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = I(y_2) - I(y_1)$$

É necessário ainda normalizar a variável, através da transformação

$$y = \frac{x - \text{média}}{\text{desviopadrão}}$$

No caso descrito, calculamos então

$$y_2 = \frac{21,2 - 20,0}{0,5} = 2,4,$$

$$y_1 = \frac{20,7 - 20,0}{0,5} = 1,4.$$

Então, recorrendo à Tabela 40 do livro, $P(20,7 \leq r \leq 21,2) = I(2,4) - I(1,4) = 0,4918 - 0,4192 = 0,0726$, ou seja, cerca de 7 % dos tubos tem diâmetro entre 20,7 e 21,2 cm.

Exercícios

1. Uma máquina produz capacitores de 47 pF que obedecem a uma fdp gaussiana de média 47,0 pF e desvio-padrão $\sigma = 2,0$ pF.

a) Qual a fração de capacitores com capacidade entre 43,0 e 51,0 pF?

b) O fabricante tem uma ponte de altíssima precisão para a medida de capacidade ($\sigma_{\text{ponte}} \ll 0,1$ pF).

De um lote de 1000 capacitores, quantos ele selecionará (em média) com capacidade entre 46,5 e 47,5 pF?

c) Qual deveria ser o desvio-padrão para que 80 % dos capacitores tenham capacidade entre 45 e 49 pF?

2. Deseja-se determinar a espessura de um disco com precisão de 0,01mm. A espessura é calculada pela diferença de duas distâncias,

$$z = y - x.$$

A precisão do instrumento que mede y é $\sigma_y = 0,1$ mm e, do que mede x é $\sigma_x = 0,05$ mm.

a) Se foram obtidos 50 dados x_i , quantos dados y_i serão necessários para obter a espessura do disco com a precisão desejada?

b) Qual a menor quantidade de dados, incluindo x_i e y_i , necessários para obter a espessura do disco com a precisão desejada?

3. É possível medir a dimensão de um objeto com 2 instrumentos que têm precisões σ_A e σ_B diferentes, com $\sigma_B = 2\sigma_A$. No entanto, demora-se 3 minutos para efetuar uma observação com o instrumento A e 1 minuto para efetuar uma observação com o instrumento B. Se você tem 60 minutos para efetuar a medida, qual instrumento dará melhor precisão?

4. Um espectrômetro detecta partículas com uma eficiência ε variável com a energia de acordo com a fórmula $\varepsilon = \alpha/E$. O parâmetro α é conhecido e vale $1,00 (4) 10^{-2}$. A tabela 6.1 abaixo dá o número de partículas de cada energia detectado. As energias têm erros desprezíveis.

TABELA 6.1

E (MEV)	ε	# partículas	I = #/ ε
2,500		370 (25)	
3,333		870 (35)	
4,000		1025 (40)	
5,000		210 (20)	
7,500		480 (28)	
10,000		800 (30)	

a) Calcule o número de partículas de cada energia emitidas.

b) Calcule a intensidade relativa às partículas de energia de 4,000 MeV. A intensidade relativa é obtida dividindo-se o número de partículas da energia desejada pelo número de partículas de 4,000 MeV.

5. Para medir a carga de um capacitor, um estudante fez uma medida de capacidade C e uma medida de voltagem V com o seguinte resultado:

$$C = \{ 0,95; 0,93; 1,24; 0,95; 0,85 \} \text{ em } \mu\text{F}$$

$$V = \{ 288; 303; 320; 291; 316; 296; 287; 298; 299; 309 \} \text{ em Volt}$$

a) Qual a carga do capacitor?

b) Seria possível obter a mesma precisão na carga em menos tempo? Discuta.

c) Seria possível obter uma melhor precisão na carga tomando-se os mesmos 15 dados no total? Se sim, sugira uma forma possível de fazê-lo.

7. O método da Máxima Verossimilhança

a) O princípio da Máxima Verossimilhança

Uma emissora que deseja saber sua audiência "sorteia" um determinado número de residências, digamos 1000, e consulta em um mesmo horário todas as residências. Suponhamos que 172 residências estivessem assistindo àquela emissora. A partir desse resultado estima-se que dos 3 milhões de televisores da cidade, cerca de 516 mil estão ligados na emissora em questão.

Note que a conclusão não é trivial. Pode ser que os únicos a assistirem a emissora fossem os 172 moradores das residências visitadas que declararam assisti-la. Pode ainda ser que no resto da população a assistência fosse muito maior, quem sabe 2 milhões. É impossível dar o número exato, ele restará desconhecido até o fim. Porém, o mais provável (se a amostragem foi adequada) é que 17,2 % de todas as TV's estivessem ligadas na emissora, o que nos levou ao número 516 mil.

O princípio da Máxima Verossimilhança propõe exatamente isso: "estime a realidade como o evento mais provável".

Formalmente, para um conjunto de dados $\{x_i\}$ que obedecem a uma fdp $f(x | \vec{a})$, onde \vec{a} representa o conjunto de parâmetros de fdp (por exemplo, $a_1 = x_0$ e $a_2 = \sigma_0$ para a normal), escrevemos a função verossimilhança $L(\{x_i\} | \vec{a})$ como

$$L(\{x_i\} | \vec{a}) = f(x_1 | \vec{a}) f(x_2 | \vec{a}) \dots f(x_n | \vec{a}).$$

O símbolo L vem de Likelihood, verossimilhança em inglês.

b) Ainda o desvio-padrão...

E esta não é a última vez que falaremos nessa grandeza!

Vamos aqui achar a estimativa de máxima verossimilhança do desvio-padrão e concluir que não a usaremos.

Para um dado x_i que obedece à fdp gaussiana, temos

$$L(x_i/x_0, \sigma_0) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(x_i - x_0)^2}{2\sigma_0^2}}$$

Então

$$L(\{x_i\}/x_0, \sigma_0) = \left(\frac{1}{\sigma_0 \sqrt{2\pi}}\right)^N e^{-\frac{\sum (x_i - x_0)^2}{2\sigma_0^2}}$$

Os valores mais prováveis de x_0 e σ_0 , que chamaremos \tilde{x} e $\tilde{\sigma}$, são aqueles que maximizam a função L . É mais fácil, entretanto, trabalhar com $\ln L$. Como L e $\ln L$ são máximos no mesmo ponto, tanto faz do ponto de vista teórico. Assim, calculamos

$$\ln L = -N \ln \sqrt{2\pi} - N \ln \sigma_0 - \frac{\sum_{i=1}^N (x_i - x_0)^2}{2\sigma_0^2}$$

Devemos encarar $\ln L$ como função de x_0 , σ_0 e achar \tilde{x} e $\tilde{\sigma}$ que maximizam $\ln L$. Assim, calculamos

$$\frac{\partial \ln L}{\partial x_0} = \frac{\sum (x_i - x_0)}{\sigma_0^2}$$

$$\frac{\partial \ln L}{\partial \sigma_0} = -\frac{N}{\sigma_0} + \frac{\sum_{i=1}^N (x_i - x_0)^2}{\sigma_0^3}$$

No ponto \tilde{x} , $\tilde{\sigma}$ ambas as derivadas devem se anular para que $\ln L$ (e, portanto, L) seja máximo, o que dá as equações

$$\frac{\sum (x_i - \tilde{x})}{\tilde{\sigma}^2} = 0 \quad ,$$

$$-\frac{N}{\tilde{\sigma}} + \frac{\sum_{i=1}^N (x_i - \tilde{x})^2}{\tilde{\sigma}^3} = 0 \quad ,$$

O que resulta em

$$\tilde{x} = \frac{\sum x_i}{N} \quad ,$$

$$\tilde{\sigma}^2 = \frac{\sum (x_i - \tilde{x})^2}{N}$$

Note que $\tilde{x} = \bar{x}$ que vínhamos utilizando, mas que $\tilde{\sigma}$ não está calculado da maneira habitual.

Embora $\tilde{\sigma}^2$ seja o valor mais provável da variância verdadeira acontece que em média ele não é igual a σ_0^2 ! Essa tendenciosidade é relativamente grave do ponto de vista da estatística e convém corrigi-la, mesmo que para isso percamos o atributo "mais provável", considerado menos importante que o da "não-tendenciosidade". Para corrigir, precisamos calcular o valor médio de $\tilde{\sigma}^2$,

$$\langle \tilde{\sigma}^2 \rangle = \frac{1}{N} \langle \sum (x_i - \bar{x})^2 \rangle \quad .$$

Escreveremos $\bar{x} = x_0 + \alpha$ onde α é a diferença (desconhecida!) entre a média e o valor verdadeiro. Assim,

$$(x_i - \bar{x})^2 = (x_i - x_0 - \alpha)^2 = (x_i - x_0)^2 - 2\alpha(x_i - x_0) + \alpha^2 \quad .$$

Recolocando essa expressão na de $\langle \tilde{\sigma}^2 \rangle$,

$$\langle \tilde{\sigma}^2 \rangle = \frac{1}{N} \langle \sum (x_i - x_0)^2 - 2\alpha \sum (x_i - x_0) + N\alpha^2 \rangle.$$

Notando que $\sum (x_i - x_0) = N(\bar{x} - x_0) = N\alpha$, chegamos a

$$\langle \tilde{\sigma}^2 \rangle = \frac{1}{N} \langle \sum (x_i - x_0)^2 \rangle - \langle \alpha^2 \rangle.$$

Em primeiro lugar, $\langle \sum (x_i - x_0)^2 \rangle = N\sigma_0^2$,

e podemos calcular

$$\langle \alpha^2 \rangle = \langle (\bar{x} - x_0)^2 \rangle = \sigma_m^2 = \frac{\sigma_0^2}{N}.$$

Assim,

$$\langle \sigma^2 \rangle = \sigma_0^2 - \frac{\sigma_0^2}{N} = \frac{N-1}{N} \sigma_0^2,$$

mostrando que o $\tilde{\sigma}^2$ subestima em média a variância verdadeira.

Para corrigir a subestimação, basta estimar a variância por

$$\sigma^2 = \frac{N}{N-1} \tilde{\sigma}^2,$$

O que dá, de imediato, $\langle \sigma^2 \rangle = \langle \sigma_0^2 \rangle$. Assim a estimativa

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - x_0)^2,$$

nada mais é que a estimativa de máxima verossimilhança corrigida da sua tendenciosidade.

Uma observação: o método de máxima verossimilhança dá quase sempre estimativas tendenciosas quando o número de dados é pequeno, exceção feita a funções lineares como veremos mais tarde.

c) O método do Mínimo Quadrado

Frequentemente desejamos conhecer parâmetros a_1, a_2, \dots, a_m de uma função

$$g(\bar{a}; x),$$

que depende de uma variável x . Para isso, obtemos valores experimentais de f para diversos x_i . Chamaremos de y_i o valor experimental de $g(\bar{a}; x_i)$. Para que a notação seja compacta, escreveremos \bar{a} para o conjunto dos parâmetros a_1, a_2, \dots, a_m .

O erro da medida faz com que y_i não seja idêntico a $g(\bar{a}; x_i)$ mas sim

$$y_i = g(\bar{a}; x_i) + \varepsilon_i,$$

onde ε_i é o erro. Supondo que o erro ε seja gaussiano, podemos escrever a fdp de y_i como

$$f(y_i/\bar{a}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(y_i - g(\bar{a}; x_i))^2}{2\sigma_i^2}\right\}.$$

Para a medida completa $\{ (x_i, y_i, \sigma_i) \}$, temos que a função verossimilhança será

$$L(\{x_i, y_i, \sigma_i\}/\bar{a}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(y_i - g(\bar{a}, x_i))^2}{2\sigma_i^2}\right\} = \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}}\right] \exp\left\{-\sum_{i=1}^N \frac{(y_i - g(\bar{a}, x_i))^2}{2\sigma_i^2}\right\}$$

o que dá, tomando o logaritmo,

$$\ln L = -\sum_{i=1}^N (\ln \sqrt{2\pi} \sigma_i) - \frac{1}{2} \sum_{i=1}^N \frac{(y_i - g(\bar{a}, x_i))^2}{\sigma_i^2}$$

Assim, $\ln L$ será máximo quando a soma

$$Q = \sum_{i=1}^N \frac{(y_i - g(\bar{a}, x_i))^2}{\sigma_i^2}$$

for mínima.

Os parâmetros a_1, a_2, \dots, a_n de máxima verossimilhança serão aqueles que minimizarem a soma dos quadrados ponderados pelas variâncias, Q . Esse é o método dos mínimos quadrados.

Uma observação: Toda a vez que a grandeza experimental for gaussiana, os métodos da máxima verossimilhança e dos mínimos quadrados recaem nas mesmas equações.

Outra observação: Se a fdp de y_i não for gaussiana, o método da máxima verossimilhança costuma ser mais adequado se $g(\bar{a}; x_i)$ não for linear nos parâmetros a_i . Se $g(\bar{a}; x_i)$ é linear em a_i , então o método mais adequado é o dos mínimos quadrados.

Em resumo, para casos "comuns", usa-se:

fdp de y gaussiana \rightarrow mínimos quadrados = máxima verossimilhança

g linear \rightarrow mínimos quadrados

fdp de y não gaussiana

g não linear \rightarrow máxima probabilidade

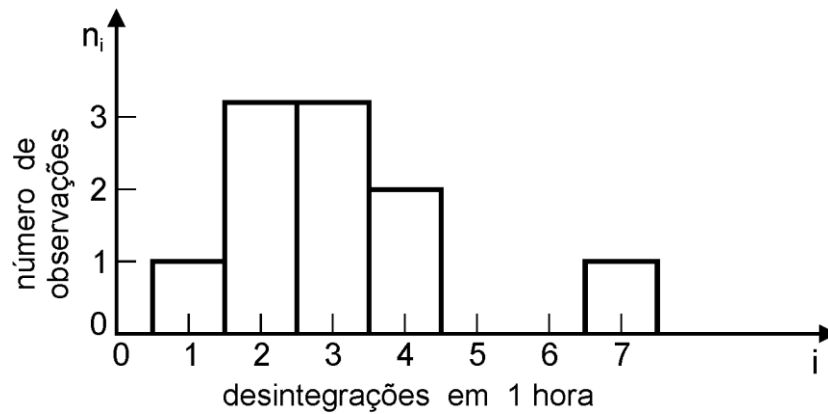
Exercícios

1. Um experimentador observa o decaimento de uma fonte radioativa de meia-vida muito longa, digamos 10^6 anos.

a) Num primeiro experimento, ele obtém um único dado, observando 6 desintegrações em 1 hora. Estime o valor médio do número de desintegrações por hora (dph) dessa fonte, usando o método da máxima verossimilhança.

b) Num segundo experimento, ele obtém dois dados: {3, 6} desintegrações em 1 hora. Novamente, estime o número de dph dessa fonte usando o método da máxima verossimilhança.

c) Finalmente, ele decide-se a levantar um histograma:



Seja a o número verdadeiro de desintegrações em 1 hora, prove que a função verossimilhança é

$$L(\{n_i\}/a) = \prod_{i=1}^{\infty} \left[\frac{a^i e^{-a}}{i!} \right]^{n_i}$$

onde n_i é o número de vezes que ele observou i contagens em uma hora. Calcule a estimativa de máxima verossimilhança de a , \hat{a} , a partir do máximo de $\ln L$.

d) Estime o desvio-padrão de \hat{a} a partir da fórmula que você obteve no item c notando que o número total de contagens nas 10 observações obedece a uma função de Poisson e use o resultado do item a. (Obs.: perceba, então que você poderia se passar sem a expressão complicada do item c, neste caso particular. Este exercício foi o melhor que achei para ilustrar o método nesta altura do curso).

2. Efetuam-se duas medidas independentes de uma mesma grandeza x , obtendo-se os resultados x_1, σ_1 e x_2, σ_2 , sendo σ_1 e σ_2 os desvios-padrão das médias. Os resultados das duas medidas podem ser reunidos de forma a obtermos o melhor valor da grandeza. Mostre, usando o método da máxima verossimilhança, que

$$x = \frac{x_1 / \sigma_1^2 + x_2 / \sigma_2^2}{1 / \sigma_1^2 + 1 / \sigma_2^2}$$

é o valor mais provável da grandeza, considerando-se os resultados das duas medidas.

8) Ajuste de funções lineares nos parâmetros

a) Aplicação do método do mínimo quadrado - forma geral para ajuste de uma função linear nos parâmetros

Suponha uma medida de uma grandeza y que é um polinômio em x . Deseja-se estimar os coeficientes do polinômio

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_i x^i + \dots + a_m x^m.$$

A variável x é a variável independente, medida sem erro, e y é a grandeza medida (variável dependente) para a qual mede-se também o desvio padrão. Assim, a medida é representada pelo conjunto dos termos de valores, $\{ (x_i, y_i, \sigma_i) \}$, $i = 1, 2, \dots, N$.

O método do mínimo quadrado estima os coeficientes a_k do polinômio como o conjunto dos \tilde{a}_k que minimizam a função

$$Q = \sum_{i=1}^N \frac{\left[y_i - \sum_{k=0}^m a_k x_i^k \right]^2}{\sigma_i^2}$$

O mínimo de Q está no ponto $\tilde{\mathbf{a}}$ onde todas as derivadas parciais de Q em relação aos a_k são nulas. Calculando

$$\frac{\partial Q}{\partial a_\ell} = -2 \sum_{i=1}^N \frac{\left[y_i - \sum_{k=0}^m a_k x_i^k \right] x_i^\ell}{\sigma_i^2},$$

que igualadas a zero no ponto $\tilde{\mathbf{a}}$ dá as equações

$$\sum_{k=0}^m \tilde{a}_k \sum_{i=1}^N \frac{x_i^{\ell+k}}{\sigma_i^2} = \sum_{i=1}^N \frac{y_i x_i^\ell}{\sigma_i^2} \quad \ell = 0, 1, 2, \dots, m.$$

Definindo a matriz M pela relação

$$M_{\ell k} = \sum_{i=1}^N \frac{x_i^{\ell+k}}{\sigma_i^2},$$

o vetor $\tilde{\mathbf{D}}$ por

$$\tilde{D}_\ell = \sum_{i=1}^N \frac{y_i x_i^\ell}{\sigma_i^2}$$

o conjunto das equações pode ser escrita em forma matricial como

$$\tilde{\mathbf{D}} = \mathbf{M} \tilde{\mathbf{a}},$$

o que permite calcular $\tilde{\mathbf{a}}$ de imediato como

$$\tilde{\mathbf{a}} = \mathbf{M}^{-1} \tilde{\mathbf{D}} = \mathbf{V} \tilde{\mathbf{D}}.$$

Pode-se mostrar que a matriz $\mathbf{V} = \mathbf{M}^{-1}$ é a matriz das covariâncias

$$V_{ij} = \text{cov}(\tilde{a}_i, \tilde{a}_j)$$

e portanto,

$$\sigma_{a_i} = \sqrt{V_{ii}} .$$

Note que V_{ii} é independente de y , ou seja os erros em \tilde{a} não dependem dos valores experimentais dos y_i , apenas dos seus desvios-padrão σ_i

Estas fórmulas podem ser generalizadas para qualquer função linear f . Escrevendo f como

$$f(x, \vec{a}) = \sum_{k=1}^m a_k g_k(x) ,$$

onde g_k é uma função apenas de x , a soma dos quadrados dos resíduos fica

$$Q = \sum_{i=1}^N \frac{\left[y_i - \sum_{k=1}^m a_k g_k(x_i) \right]^2}{\sigma_i^2}$$

Dando as derivadas

$$\frac{\partial Q}{\partial a_\ell} = -2 \sum_{i=1}^N \frac{\left[y_i - \sum_{k=1}^m a_k g_k(x_i) \right] g_\ell(x_i)}{\sigma_i^2}$$

que, igualadas a zero no ponto \tilde{a} ficam

$$\sum_{i=1}^N \frac{y_i g_\ell(x_i)}{\sigma_i^2} = \sum_{k=0}^m \tilde{a}_k \sum_{i=1}^N \frac{g_k(x_i) g_\ell(x_i)}{\sigma_i^2} .$$

Novamente, definindo o vetor \vec{D} e a matriz M pelas relações

$$D_\ell = \sum_{i=1}^N \frac{y_i g_\ell(x_i)}{\sigma_i^2} \quad \text{e} \quad M_{k\ell} = \sum_{i=1}^N \frac{g_k(x_i) g_\ell(x_i)}{\sigma_i^2} ,$$

o sistema de equações pode ser reescrito como

$$\vec{D} = M \vec{\tilde{a}} ,$$

com a solução

$$\vec{\tilde{a}} = M^{-1} \vec{D} ,$$

e a matriz M^{-1} é a matriz das covariâncias de \tilde{a} .

Note que a facilidade de montar um algoritmo para a construção da matriz M a do vetor \vec{D} - basta, num único "loop" sobre o número de dados, calcular um vetor $\vec{g}(x_i)$ (com m componentes) e acumular em D_ℓ e $M_{k\ell}$ os somatórios.

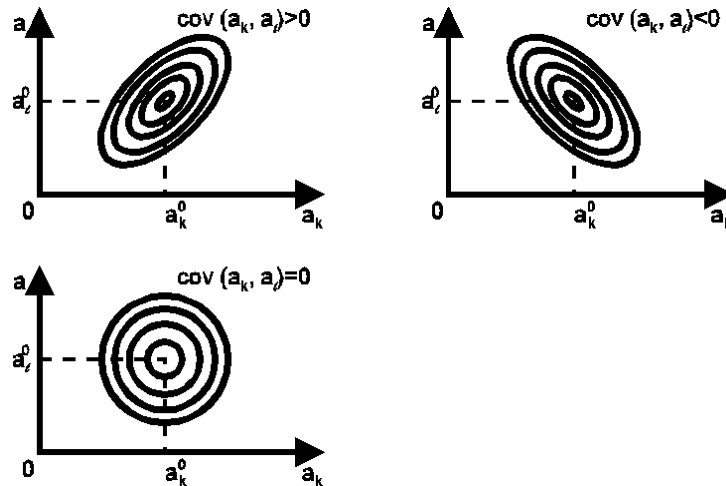
Note também que $M_{k\ell}$ é simétrica, podendo-se calcular apenas cerca da metade da matriz, simetrizando-a depois.

b) As covariâncias: como interpretá-las.

A expressão

$$\vec{a}_\ell = \sum_{i=1}^m V_{i\ell} D_\ell ,$$

mostra que todos \tilde{a}_ℓ dependem dos mesmos valores experimentais y_i embutidos no vetor \vec{D} . Como eles são dependentes, as covariâncias entre eles não são, geralmente, nulas. Se a covariância entre dois parâmetros a_k e a_l não é nula, isso significa que a fdp de a_k e a_l favorece certas combinações desses parâmetros. As curvas de nível da fdp de a_k , a_l , nos 3 casos: covariância nula, positiva, e negativa, estão esboçadas abaixo.



A covariância ser positiva não implica, obrigatoriamente, que ambas a_k e a_l são maiores ou menores, simultaneamente, que a_k^0 e a_l^0 . Apenas diz que é mais provável que isso aconteça. Já covariância nula diz que não há nenhuma relação de dependência estatística entre as variáveis (embora possa haver uma dependência física). Esta discussão já foi exemplificada no roteiro 7, quando discutimos propagação de incertezas.

c) Incerteza do valor interpolado de uma função.

Suponha que você faça o ajuste de uma reta $y = a_0 + a_1x$ com dados de uma medida $\{(x_i, y_i, \sigma_i)\}$.

Você obtém $\vec{\tilde{a}} = V \vec{D}$ e as covariâncias como

$$\text{cov}(a_i, a_j) = V_{ij}.$$

Se você desejar calcular o valor de y para um certo x , é imediato que

$$y = \tilde{a}_0 + \tilde{a}_1 x$$

e a variância de y

$$\sigma_y^2 = \sigma_{a_0}^2 + x^2 \sigma_{a_1}^2 + 2x \text{cov}(\tilde{a}_0, \tilde{a}_1)$$

que, expressando as covariâncias pela matriz V dá

$$\sigma_y^2 = V_{00} + x^2 V_{11} + 2x V_{10}$$

Exercícios

1. Realizaram-se 5 observações da posição de um objeto que se move com velocidade constante. Os dados (t_i, x_i, σ_i) , onde σ_i é o desvio-padrão da posição x_i observada no instante t_i , são:

$\{(1; 1,06; 0,15); (2; 2,19; 0,15); (3; 3,12; 0,15); (4; 4,12; 0,15); (5; 5,32; 0,15)\}$

onde t está calculado em segundo, x e σ em centímetro.

- Calcule a posição inicial ($t = 0$) e a velocidade.
- Determine a posição (e o desvio-padrão) no instante $t = 3,5$ s.
- Estime a probabilidade da posição inicial ser nula, sabendo que devido a montagem utilizada, a posição inicial é positiva ou nula, jamais negativa.

2. Utilizando os dados do problema 1, ajuste uma função

$$x = a_1 + a_2 (\alpha + t),$$

de maneira que os parâmetros \tilde{a}_1 e \tilde{a}_2 obtidos sejam independentes estatisticamente.

3. Deseja-se ajustar uma reta

$$y = ax + b$$

por pontos experimentais (x_i, y_i, σ_i) onde os desvio-padrão σ_i são independentes. Sabendo que você pode observar no máximo 6 pontos entre $-10 \leq x \leq 10$, (e nenhum fora deste intervalo) e que você deve observar pontos separados de pelo menos $\Delta x = 2$:

- Como devem ser distribuídas as observações para que \underline{a} e \underline{b} sejam independentes?
- Qual o conjunto de pontos que torna σ_a^2 mínimo? (dado que a e b devem ser independentes)
- Qual o conjunto de pontos que torna σ_b^2 mínimo, com a e b independentes?

4. Uma superfície plana está inclinada em relação à vertical. Sendo definido um sistema de eixos triortogonal de maneira que z seja o eixo vertical, e corte o plano $(0,0,0)$, a superfície do plano é definida por

$$z = ax + by$$

Para determinar a e b realizaram-se 4 observações de $(x_i, y_i, z_i, \sigma_i)$ de pontos do plano, onde σ_i é o desvio-padrão de z_i e as variáveis x_i, y_i são observadas com erro desprezível em relação a σ_i . Os dados obtidos foram

x_i	y_i	z_i
1	1	3,03(10)
1	2	5,17(15)
2	1	4,01(15)
2	2	5,93(20)

- Determine \underline{a} e \underline{b} com suas respectivas incertezas.
- Determine um intervalo com 68 % de probabilidade de conter o valor verdadeiro de \underline{a} .
- Mesmo que em ii, para o parâmetro b .
- Determine z do plano para o ponto $x = 1,5$ e $y = 1,5$ e determine um intervalo com cerca de 68 % de probabilidade de conter o valor verdadeiro desse z .
- Seria possível, observando-se apenas quatro pontos do plano, determinar \underline{a} e \underline{b} com maior precisão? Discuta essa questão e dê uma diretriz genérica sobre como selecionar os pontos (x, y) onde deveria ser observado z para aumentar a precisão de \underline{a} e de \underline{b} .

9. Testes estatísticos

a) Ainda a flutuação estatística

Um ajuste de função a pontos experimentais $\{(x_i, y_i, \sigma_i)\}$ pode ser avaliado *qualitativamente* verificando-se se o gráfico da função passa próximo ou não aos pontos experimentais, considerando-se as barras de erro, $y_i \pm \sigma_i$. Em média, o gráfico deve cortar pouco mais que 68% das barras de erro; passar a menos de 2 barras de erro em pouco mais de 95% dos pontos; passar a menos que 3 barras de erro na quase totalidade dos pontos.

b) Teste de hipótese

Para chegarmos a uma decisão sobre se uma hipótese particular é confirmada por um conjunto de dados, devemos dispor de um procedimento objetivo que nos permita rejeitar ou aceitar a hipótese. Este procedimento envolve algumas etapas:

- i) Definir a chamada *hipótese de nulidade* (H_0).
- ii) Escolher uma prova estatística para provar H_0 .
- iii) Especificar um nível de significância (α) e um tamanho de amostra (N).
- iv) Determinar (ou supor determinada) a distribuição amostral da prova estatística sob a hipótese de nulidade.
- v) Com base em (ii), (iii) e (iv), definir a região de rejeição.
- vi) Calcular o valor da prova estatística, utilizando os dados obtidos da(s) amostra(s). Se tal valor estiver na região de rejeição, a decisão será rejeitar H_0 ; se o valor estiver fora da região de rejeição, a decisão será que a hipótese H_0 não pode ser rejeitada ao nível de significância escolhido.

A *hipótese de nulidade* (H_0) é uma hipótese de que não haja diferenças. Em muitos casos é formulada com o exposto propósito de ser rejeitada. No entanto, no caso de ajustes de função, como o acima indicado, é formulada para ser aceita. Se for rejeitada, pode-se aceitar a *hipótese alternativa* (H_1).

A prova estatística depende do problema estudado e nesta aula veremos duas delas: o *Teste de χ^2* e o *Teste t*.

O *nível de significância* (α) é a probabilidade de se rejeitar H_0 quando H_0 é verdadeira. Isto é, α é a probabilidade de se cometer o erro *Tipo I* (rejeitar impropriamente H_0). A escolha da significância depende muito do risco, que se admite correr, de estar errado ao se tomar uma decisão estatística de aceitar ou rejeitar uma hipótese de nulidade. Além do erro Tipo I, podemos cometer o erro *Tipo II* (aceitar H_0 quando ela é falsa). A Figura 1 mostra estes dois tipos de erros em relação às duas decisões e às duas possibilidades para H_0 (verdadeira ou falsa).

A probabilidade de se cometer um erro Tipo I é simplesmente o nível de probabilidade que os pesquisadores escolhem para rejeitar H_0 . Por que, então, não adotar um alfa extremamente pequeno e praticamente nunca cometer este tipo de erro? O problema é que quando diminuimos α , automaticamente aumentamos a probabilidade de cometermos um erro Tipo II. A probabilidade de se cometer um erro Tipo II é usualmente denotada por β . Alfa e beta são inversamente relacionados: quando um aumenta, o outro diminui. Alfa está sempre sob o controle do experimentador; beta está apenas indiretamente sob nosso controle, depende de α e do tamanho da amostra, para uma dada prova estatística.

Figura 1.

		DECISÃO	
		rejeitar H_0	aceitar H_0
H_0 verdadeiro	erro Tipo I $p = \alpha$	correto	
H_0 falso	correto	erro Tipo II $p = \beta$	

c) A variável aleatória χ^2

O teste que avalia quantitativamente a qualidade do ajuste citado no item (a) é o *Teste de χ^2* .

O mínimo da função Q ,

$$Q_{\text{mínimo}} = \sum_{i=1}^N \frac{[y_i - g(x_i; \tilde{\mathbf{a}})]^2}{\sigma_i^2},$$

onde $\tilde{\mathbf{a}}$ é o vetor dos m parâmetros que minimizam Q , é uma variável aleatória χ^2_L , cuja fdp é conhecida (estamos supondo que a fdp dos erros $y_i - g(x_i; \mathbf{a}_0)$ é gaussiana). Aqui, $L = N - m$ é o número de graus de liberdade. No livro de O. Helene e V. Vanin, *Tratamento Estatístico de Dados em Física Experimental*, a fdp da variável χ^2 está deduzida e apresentada algumas de suas propriedades.

A fdp de χ^2 tem sua expressão

$$F_L(\chi^2) = \frac{1}{2\Gamma\left(\frac{L}{2}\right)} \left(\frac{\chi^2}{2}\right)^{\frac{L}{2}-1} e^{-\frac{\chi^2}{2}},$$

cujo domínio é de zero a infinito e prova-se que a média do χ^2 , $\chi_0^2 = L$. Se todos os dados y_i estiverem muito próximos de dos valores verdadeiros, aqui estimados por $g(x_i; \tilde{\mathbf{a}})$, então χ^2 poderá ser bem menor que L , mas isto é pouco provável. O mesmo acontece com χ^2 bem maior do que L .

É mais conveniente trabalhar com a função cumulativa da probabilidade, que é a integral da fdp, pois se extrai diretamente o valor da probabilidade do χ^2 estar em certo intervalo através de uma diferença de probabilidades acumuladas. Assim

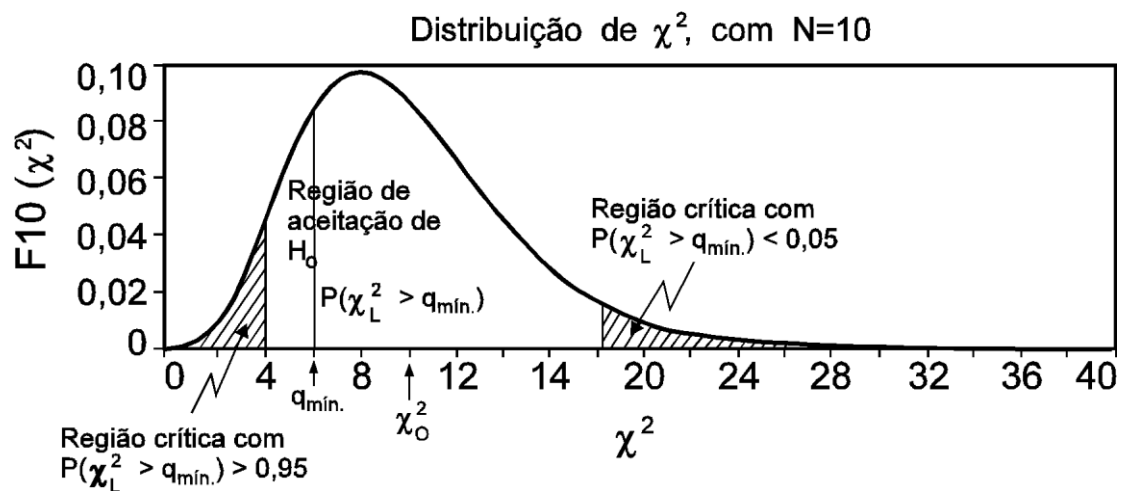
$$P(\chi_L^2 > Q_{\text{mínimo}}) = P_L(Q_{\text{mínimo}}) = \frac{1}{2\Gamma\left(\frac{L}{2}\right)} \int_{Q_{\text{mínimo}}}^{\infty} \left(\frac{\chi^2}{2}\right)^{\frac{L}{2}-1} e^{-\frac{\chi^2}{2}} d\chi^2$$

é a probabilidade χ^2 exceder o valor $Q_{\text{mínimo}}$, ou seja, é a probabilidade da flutuação estatística gerar uma medida ao menos tão "distante" dos pontos experimentais quanto a medida em questão.

A integral da fdp de χ^2 não tem forma analítica, por isso recorreremos a tabelas. Quando o número de graus de liberdade superar 30, pode-se aproximar a fdp de χ^2 pela distribuição normal de média L e desvio padrão $\sqrt{2L}$, o que nos permite usar a tabela da distribuição normal.

d) A escolha do nível de significância

Quando temos apenas um ajuste a tratar, é habitual tomar o nível de significância em $\alpha = 0,10$. A hipótese de nulidade (o ajuste é adequado, isto é, os parâmetros obtidos no ajuste diferem dos valores verdadeiros apenas por flutuação estatística, sendo portanto boas estimativas) é considerada verificada se o valor de $Q_{\text{mínimo}}$ estiver próximo de L .



A região crítica (região de rejeição) é considerada nos extremos (veja figura acima). Se $P(\chi_L^2 > Q_{\text{mínimo}}) > 0,95$, isto é, for muito grande, provavelmente as variâncias dos y_i estão superestimadas. É um engano comum propagar incertezas de calibração, que afetam igualmente todo o conjunto dos y_i , nas variâncias σ_i . Se esses erros são iguais para todos os pontos, são erros sistemáticos e não se deve aumentar as variâncias individuais para levá-los em conta. Se $P(\chi_L^2 > Q_{\text{mínimo}}) < 0,05$, isto é, for muito pequena, ao excluirmos a "má sorte", restam-nos duas possibilidades:

- i) as variâncias estão subestimadas; ou
- ii) a função g é inadequada.

e) Erros Tipo I e Tipo II

Se tivermos muitos ajustes, os valores críticos de χ^2 devem ser alterados. Vamos imaginar uma situação onde realizamos $M = 150$ ajustes de funções. Mantendo-se o limite

crítico ao nível de probabilidade $p = 5\%$, o número n de casos que cairão nessa faixa de χ^2 é dado pela binomial

$$P_{M,p}(n) = \binom{150}{n} \cdot 0,05^n \cdot 0,95^{150-n}.$$

O número de ajustes nessa faixa é $\bar{n} = M \cdot p = 7,5$, de forma que teremos algo em torno de 8 ajustes nessa faixa sem que nem σ_i nem a função sejam inadequadas. Esses ajustes seriam, portanto, rejeitados equivocadamente. A rejeição equivocada de um ajuste é aqui um erro Tipo I, onde "erro" está empregado no sentido de engano de avaliação. O erro Tipo II corresponderia à aceitação de um ajuste inadequado, mas que passou pelo teste porque a flutuação estatística deixou os dados experimentais de maneira a assemelharem-se à função g .

EXERCÍCIOS

1. Um experimentador obteve a grandeza A a partir da relação

$$A = x - y,$$

tendo obtido para x e y os resultados

$$x = 13,4 \text{ (6)}$$

$$y = 10,4 \text{ (3)}$$

$$\text{cov}(x,y) = 0,1$$

a) Qual o valor da grandeza A e de sua incerteza?

b) Um segundo experimentador mediu diretamente a grandeza A obtendo-a como

$$A = 4,0 \text{ (10)}$$

Estime o melhor valor A (e a respectiva incerteza) que pode ser obtido a partir da reunião das duas medidas.

c) Para verificar se os dois experimentadores obtiveram valores compatíveis, calcule χ^2 e a probabilidade desse χ^2 obtido ser excedido (consulte tabela na página 84 do livro de Helene e Vanin). Os valores obtidos pelos dois experimentadores são compatíveis? Justifique.

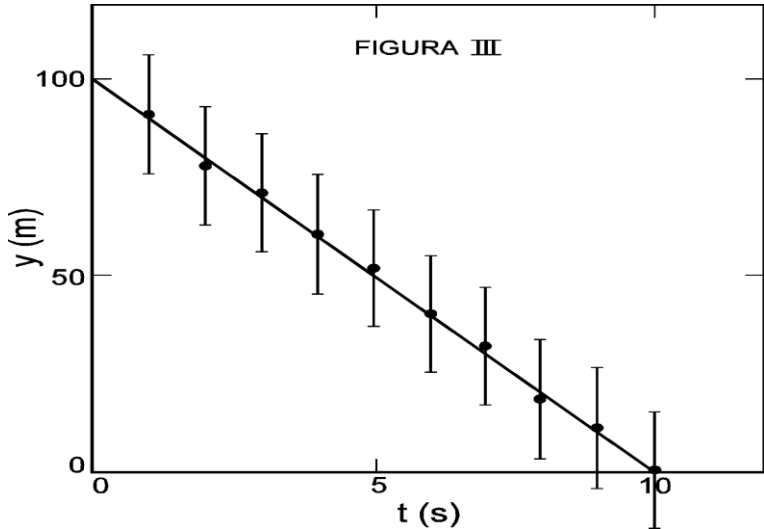
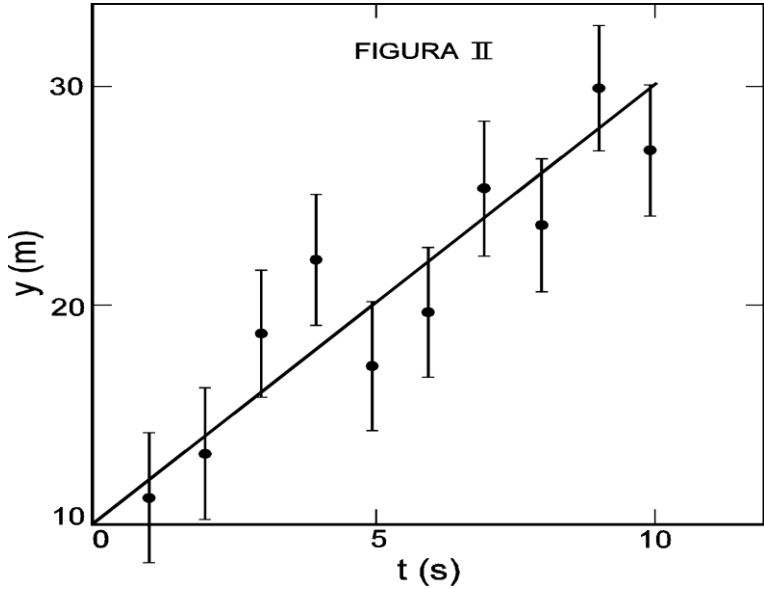
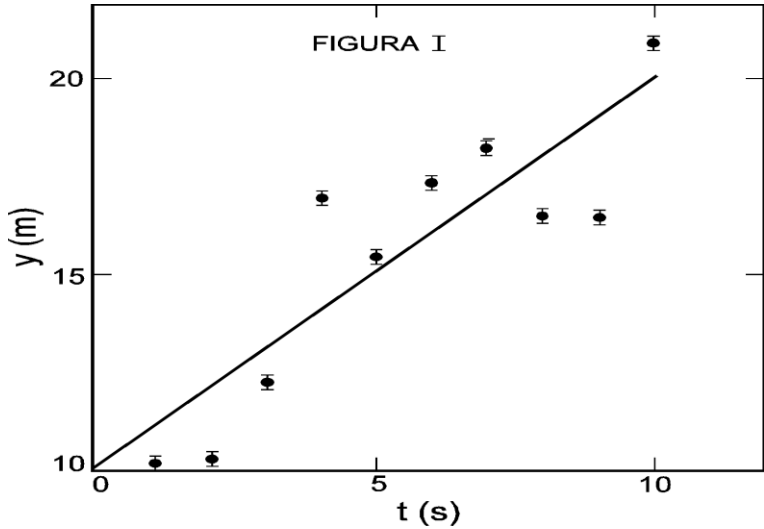
2. As figuras I, II e III mostram os dados experimentais, com suas respectivas barras de erro σ_i , obtidos em medidas da posição de um corpo em Movimento Retilíneo e Uniforme,

$$y = y_0 + v_0 t$$

Em cada gráfico, está representada a reta

$$y = \tilde{y}_0 + \tilde{v}_0 t$$

obtida no ajuste dos dados pelo método dos mínimos quadrados. Pergunta-se, em cada caso, se o ajuste obtido é adequado ou não. Justifique suas respostas e aponte eventuais problemas com os dados experimentais (x_i, y_i, σ_i).



3. Um experimentador ajustou uma reta $y = a + bx$ aos pontos experimentais (x_i, y_i, σ_i) :

$\{ (1; 4,12; 0,15), (2; 5,32; 0,15), (3; 6,31; 0,15) \}$, obtendo-se o resultado

$$a = 3,06 \text{ (23)}$$

$$b = 1,10 \text{ (11)}$$

$$V = \begin{pmatrix} 0,0525 & -0,0225 \\ -0,0225 & 0,0113 \end{pmatrix}$$

onde V é a matriz de covariâncias entre a e b .

a) Calcule χ^2 e a probabilidade desse χ^2 ser excedido (consulte tabela 9.1).

b) Calcule y para $x = 1,5$, com sua respectiva incerteza.

TABELA 9.1 – Probabilidde do χ^2 ser excedido em termos do número de graus de liberdade L .

L	P (%)								
	99	95	90	75	50	25	10	5	1
1	0.00016	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	6.63
2	0.0201	0.103	0.211	0.575	1.39	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.21	2.57	4.11	6.25	7.81	11.3

10. O Teste t . Intervalos de Confiança.

a) Largura do histograma em função do número de dados

A largura de um histograma é medida pelo desvio padrão, a raiz quadrada da variância. Esta é estimada de maneira não tendenciosa pela expressão:

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

A fdp de σ^2 pode ser calculada notando que

$$(N-1) \frac{\sigma^2}{\sigma_0^2} = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma_0^2} = \chi_{N-1}^2,$$

onde σ_0^2 é o valor verdadeiro da variância. Assim $\langle \sigma^2 / \sigma_0^2 \rangle = 1$, pois $\langle \chi_{N-1}^2 \rangle = N-1$, e a fdp de σ^2 é a de χ^2 , convenientemente normalizada. O número de graus de liberdade é $L = N-1$.

Dada a assimetria da fdp de χ_L^2 , para um conjunto de N dados, a chance de subestimar σ é maior do que a de superestimar: ~70% contra ~30% para $L=1$; ~65% contra ~35% para $L=2$; ~55% contra ~45% para $L=10$; tendendo a 50%-50% para $L>30$, onde a fdp já é praticamente normal.

Conclusão: embora a fdp que governa a medida x_i não dependa do número de dados (isto é, existem x_0 e σ_0 , valores verdadeiros da grandeza e da incerteza de medida), a largura **provável** do histograma **medido aumenta** com o número de dados. Vale a pena, entretanto, ressaltar que **em média** a largura do histograma **não muda**, pois $\langle \sigma^2 / \sigma_0^2 \rangle = 1$. Escolhe-se medir a largura do histograma corretamente em média, mas poderíamos ter escolhido a largura dos histogramas pelo valor mais provável de ser a correta. Não é possível ter ambas as propriedades: “média correta” e “provavelmente correta” simultaneamente!

b) O desvio padrão efetivo

Em função da provável subestimação do desvio padrão e conseqüentemente do desvio padrão da média, em geral, a probabilidade

$$P(\bar{x} - \sigma_m \leq x_0 \leq \bar{x} + \sigma_m) < 68,3\%,$$

tendendo ao valor 68,3% à medida que N aumenta.

Chama-se desvio padrão efetivo ao valor $t_I \sigma_m$ tal que

$$P(\bar{x} - t_I \sigma_m \leq x_0 \leq \bar{x} + t_I \sigma_m) = 68,3\%.$$

Usando-se a definição da variável aleatória t de Student,

$$t = \frac{\bar{x} - x_0}{\sigma_m},$$

a equação acima pode ser escrita como

$$P(-t_I \leq t \leq +t_I) = 68,3\%,$$

o que nos permite calcular t_I resolvendo a equação

$$\int_{-t_I}^{+t_I} f_L(t) dt = 68,3\%,$$

onde $f_L(t)$ é a fdp de t para $L=N-1$ graus de liberdade. A expressão de $f_L(t)$ é

$$f_L(t) = \left(\frac{1}{\pi L}\right)^{1/2} \cdot \frac{\Gamma\left(\frac{L+1}{2}\right)}{\Gamma\left(\frac{L}{2}\right)} \cdot \left(\frac{1}{1+t^2/L}\right)^{\frac{L+1}{2}}$$

A variável t de Student tem média nula para $L>1$ e variância $\sigma^2 = L/(L-2)$, para $L>2$. Podemos mostrar que $f_L(t)$ tende à distribuição normal com média nula e desvio padrão unitário, quando L tende a infinito.

O intervalo que contém o valor verdadeiro com 95,4% de probabilidade é

$$\left[\bar{x} - t_{II} \sigma_m, \bar{x} + t_{II} \sigma_m \right]$$

e podemos calcular t_{II} a partir da equação

$$\int_{-t_{II}}^{+t_{II}} f_L(t) dt = 95,4\%,$$

Os valores de t_I e t_{II} estão tabelados abaixo. Note que $t_{II} > 2 t_I$, tendendo para $t_{II} = 2 t_I$ quando L tende a infinito.

Tabela 10.1. Valores de t_I e t_{II} , que definem os intervalos de confiança de 68,3% e 95,5%, para alguns valores do número de dados N na medida.

N	t_I	t_{II}
	$\alpha=68,3\%$	$\alpha=95,5\%$
2	1,84	14,0
3	1,32	4,53
4	1,20	3,31
5	1,14	2,87
10	1,06	2,32
20	1,03	2,14
∞	1	2

c) O Teste t

Suponha que seja realizada certa medida com N dados, cujos erros obedecem uma fdp normal, tendo-se obtido, para a grandeza, o valor médio, \bar{x} , e o desvio padrão da média, σ_m . Suponha ainda que esta grandeza tenha um valor, x_0 , previsto por determinada teoria.

Como comparar o valor medido com o valor previsto? Os dados confirmam ou contradizem a previsão teórica?

Nestas condições, a variável aleatória

$$t = \frac{\bar{x} - x_0}{\sigma_m},$$

correspondente à hipótese de que \bar{x} difere de x_0 só por flutuação estatística, tem uma fdp t de Student com $L=N-1$ graus de liberdade, sobre a qual pode-se montar um *teste de hipótese*, conforme apresentado a seguir.

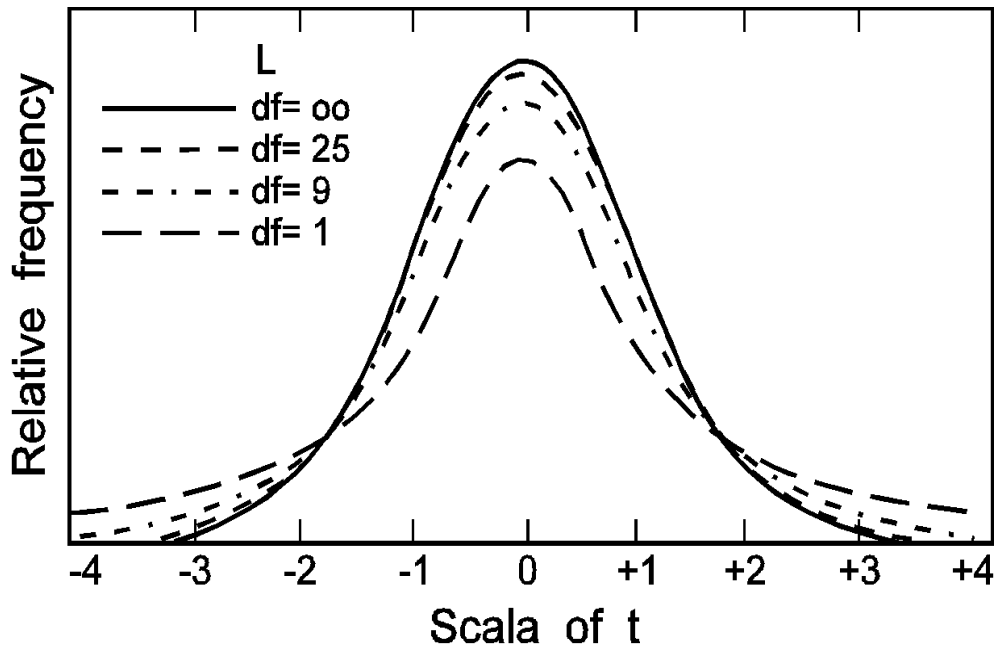


Figura 1. Fdp de t de Student, para alguns graus de liberdade L . As regiões críticas, para dado nível de significância, dependem de L .

Aceita-se a hipótese como verdadeira se $|t|$ for pequeno e rejeita-se quando $|t|$ for grande. Quanto grande? Habitualmente escolhe-se *a priori* um valor $t_{\text{crítico}}$ de $|t|$, associado ao nível de significância α , a partir do qual, $|t| > t_{\text{crítico}}$, considera-se falsa a hipótese.

$$\alpha = P(|t| > t_{\text{crítico}}) = 2 \int_{t_{\text{crítico}}}^{\infty} f_L(t) dt,$$

utilizando-se a simetria de $f_L(t)$. É comum escolher valores críticos de maneira que α seja igual a 0,05, a 0,01 ou a 0,001. Os valores de $t_{\text{crítico}}$, para diversos graus de liberdade L , correspondentes aos valores de α mencionados, estão apresentados na Tabela 2.

Tabela 10.2. Valores de $t_{\text{crítico}}$, associados aos níveis de significância α , para os graus de liberdade L .

L	$t(\alpha)$		
	$\alpha=0,05$	$\alpha=0,01$	$\alpha=0,001$
1	12,7	63,7	640
2	4,30	6,97	31,6
3	3,18	5,84	12,9
4	2,78	4,60	8,61
5	2,57	4,03	6,87
10	2,23	3,17	4,59
20	2,09	2,87	3,85
30	2,04	2,76	3,66
∞	1,96	2,58	3,29

No caso de se comparar duas medidas, com números de dados, médias e desvios padrão da média, respectivamente $(N_1, \bar{x}_1, \sigma_{1m})$ e $(N_2, \bar{x}_2, \sigma_{2m})$, trabalhar com a variável aleatória

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_{1m}^2 + \sigma_{2m}^2}},$$

com $L=(N_1-1)+(N_2-1)$ graus de liberdade.

EXERCÍCIOS

1. Uma fonte radioativa emite radiações das seguintes energias: 127, 196, 238, 307 e 507 keV. Mediu-se a intensidade relativa à transição de 307 keV em cinco espectros. A tabela abaixo resume os resultados em medidas consecutivas:

E_γ	Espectro A	Espectro B	Espectro C	Espectro D	Espectro E	I	σ	χ^2
127	1,22(11)	1,33(11)	1,46(11)	1,25(11)	1,42(11)			
196	0,53(6)	0,57(6)	0,67(6)	0,78(6)	0,80(6)			
238	0,73(9)	0,53(9)	0,73(9)	0,96(9)	0,81(9)			
507	71(8)	71(7)	58(7)	53(6)	69(6)			

Tendo em vista o resultado do teste χ^2 , a hipótese de que estas radiações provenham de um único nuclídeo é razoável?

2. Retome a tabela do exercício 1 do roteiro 3 E a medida? Escolha 6 linhas da tabela, separe os 60 dados em 20 grupos de 3, determinando para cada grupo \bar{x}_i , σ_i , $\sigma_{m_i} = \frac{\sigma_i}{\sqrt{3}}$.

a) Em quantos casos o valor verdadeiro $x_0 = 10,0$ está no intervalo $[\bar{x}_i - \sigma_{m_i}; \bar{x}_i + \sigma_{m_i}]$? Compare com o esperado pela tabela da integral da fdp da estatística “t” de Student’s no roteiro anterior.

b) Calcule o desvio-padrão efetivo para os grupos de dados, utilizando-se da tabela do livro. Determine a fração de grupos em que x_0 está no intervalo $[\bar{x}_i - \sigma_{m_i}; \bar{x}_i + \sigma_{m_i}]$

3. Com os mesmos grupos de dados do exercício anterior, determine a fração deles para os quais $\bar{x}_i - 2\sigma_{m_i} \leq x_0 \leq \bar{x}_i + 2\sigma_{m_i}$. Se usarmos $4,5\sigma_{m_i}$ ($= t_{II}\sigma_{m_i}$), o que ocorreria? Verifique e explique!

4. Mostre que o desvio-padrão do desvio-padrão é

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2(N-1)}},$$

que pode ser aproximado, para N grande, por

$$\sigma_{\sigma} \approx 0,7 \sigma_m,$$

onde σ_m é o desvio-padrão da média.

Algarismos Significativos

5. Dado o conjunto de observações da intensidade de muons por segundo num grande detector de raios cósmicos: { 8,8; 7,4; 10,4; 9,9; 10,0 } em s^{-1} .

a) calcule \bar{X} , σ , σ_m e σ_{σ} (use o resultado do exercício 4 para σ_{σ}).

b) Mostre, em função dos valores obtidos, que você deve considerar apenas uma casa depois da vírgula na representação decimal de \bar{X} .

c) Calcule a probabilidade do valor verdadeiro estar no intervalo $[\bar{X}_i - \sigma_{m_i}; \bar{X}_i + \sigma_{m_i}]$.

d) Calcule os intervalos de confiança de 68,3 % e 95,5 % de probabilidade de conterem o valor verdadeiro da grandeza.

6. Os dados da tabela do exercício 1 do roteiro 3 correspondem a uma gaussiana de média 10 e desvio-padrão 1,5.

a) Se você toma 3 dados, com quantas casas você deve representar \bar{X} ? E com 20 dados? E com 200 dados?

b) Quantos dados precisariam ser tomados para que houvessem 3 algarismos significativos após a vírgula na representação decimal de \bar{X} ?

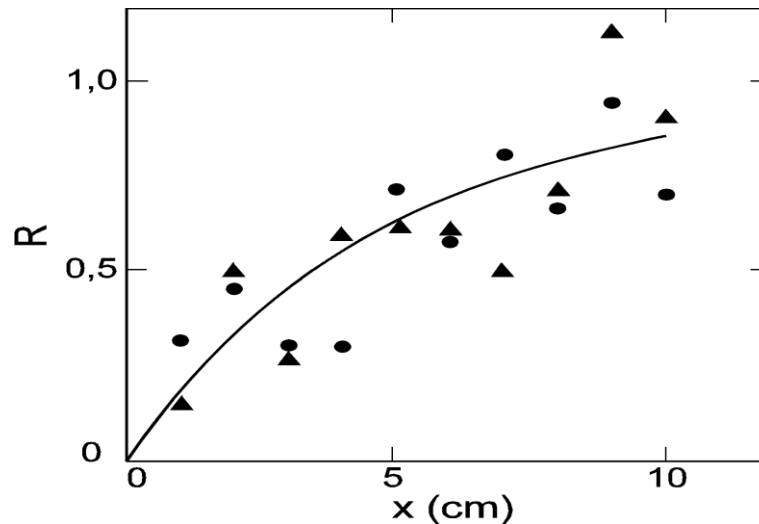
7. Numa medida de absorção de radiação pela matéria, um experimentador observou a razão entre a intensidade da radiação absorvida por uma placa de Pb de espessura x e a intensidade incidente, que segue a lei

$$R = 1 - e^{-\mu x}, \text{ com } \mu = \text{constante.}$$

gráfico representa os dados experimentais obtidos em duas medidas independentes mas realizadas sob as mesmas condições, a primeira delas representada pelo símbolo x e a segunda pelo símbolo o . A tabela dá os valores obtidos em cada uma dessas medidas. A variável independente x tem desvio-padrão desprezível.

Tabela 10.3 – Absorção (R) vs espessura (x)

x (cm)	1ª medida	2ª medida
1	0,15	0,32
2	0,50	0,45
3	0,27	0,30
4	0,59	0,29
5	0,61	0,72
6	0,61	0,58
7	0,50	0,82
8	0,72	0,68
9	1,13	0,95
10	0,91	0,71



a) Mostre que, com dois dados,

$$\sigma^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2 = \frac{1}{2} (y_1 - y_2)^2$$

b) Considerando os dois dados independentes de R para um mesmo valor de x_i , estime o desvio-padrão dos 2 dados obtidos para cada x_i (não é pedido o desvio-padrão da média!)

Considere agora que o instrumento havia sido ajustado de maneira que $\sigma_R = 0,15$ para qualquer x.

c) Os valores obtidos para σ_R no item b ignorando o fato acima são compatíveis com $\sigma_R = 0,15$ para qualquer x? Justifique e comente qualquer resultado sistemático de subestimação ou superestimação de σ_R no item b.

d) O experimentador obteve, através de um ajuste pelo método dos mínimos quadrados, a curva representada no gráfico. Em vista do resultado do ajuste e do resultado do item b, teve vontade de descartar o resultado da 2ª medida para $x = 4$ e o da 1ª medida para $x = 7$. Você acha que ele deve descartar algum desses dados, ou ambos? Se você acha que sim, explique porque, se acha que não explique porque o (s) dado (s) é (são) estatisticamente plausível (plausíveis)?

11. Ajuste de funções não lineares nos parâmetros

a) Introdução

Suponha uma medida $\{(x_i, y_i, \sigma_i)\}$, com $i=1, 2, \dots, N$, onde y_i é o valor da variável dependente observada com desvio padrão σ_i e x_i é o valor da variável independente. A relação entre y e x é dada pela função $g(x; \mathbf{a})$, onde \mathbf{a} representa o conjunto dos m parâmetros desconhecidos, de maneira que

$$y_i = g(x_i; \mathbf{a}_0) + \varepsilon_i$$

onde \mathbf{a}_0 é o valor verdadeiro de \mathbf{a} (desconhecido!) e ε_i é o erro estatístico na observação de y_i .

O método dos mínimos quadrados estima $\tilde{\mathbf{a}}$ como o vetor que minimiza a função

$$Q = \sum_{i=1}^N \frac{[y_i - g(x_i; \mathbf{a})]^2}{\sigma_i^2}.$$

Se a fdp de y for gaussiana, este procedimento corresponde a maximizar a probabilidade de que os $g(x; \mathbf{a})$ descrevam os y_i .

Se $g(x; \mathbf{a})$ não for linear em \mathbf{a} , as equações

$$\left. \frac{\partial Q}{\partial a_k} \right|_{\tilde{\mathbf{a}}} = 0,$$

não conduzem a um sistema linear de equações em a_k . Pelo contrário, estas equações são difíceis de resolver, embora possam dar soluções relativamente simples em casos particulares (veja o exemplo do Ex. 2).

Um procedimento possível consiste em calcular numericamente $Q = Q(\mathbf{a})$ para diversos \mathbf{a} e ir “cercando” o mínimo. É o equivalente a traçar superfícies (“curvas”) de nível de Q no hiperespaço de m dimensões, para identificar o(s) mínimo(s) de $Q(\mathbf{a})$. Este método é trabalhoso e mostra-se inviável com muita frequência.

Há dois procedimentos iterativos, mais diretos, para se abreviar a localização do mínimo de Q .

i) O primeiro corresponde a escolher, a partir de uma estimativa inicial \mathbf{a}' de $\tilde{\mathbf{a}}$, a trajetória de maior variação (neste caso, interessa o decréscimo mais íngreme) de Q no hiper-espaço dos a_k . Calcula-se o gradiente de Q em \mathbf{a}' , passa-se à nova estimativa

$$\mathbf{a}'' = \mathbf{a}' - \frac{\nabla Q}{|\nabla Q|} h,$$

onde o sinal indica o sentido da diminuição de Q e $h = |\Delta \mathbf{a}|$ é o módulo da variação de \mathbf{a} . Esta expressão, escrita para cada parâmetro, torna-se

$$a''_k = a'_k - \frac{\frac{\partial Q}{\partial a_k} \Delta a_k}{\sqrt{\sum_{l=1}^m \left(\frac{\partial Q}{\partial a_l} \Delta a_l \right)^2}} h,$$

na qual está evidente que o incremento em a'_k só depende da componente k do vetor unitário na direção e sentido do gradiente. Em aplicações sucessivas, pode-se chegar a um mínimo. No entanto, a principal limitação deste método é que na proximidade do mínimo a convergência é dificultada, uma vez que esta "aproximação linearizada" não admite mínimo.

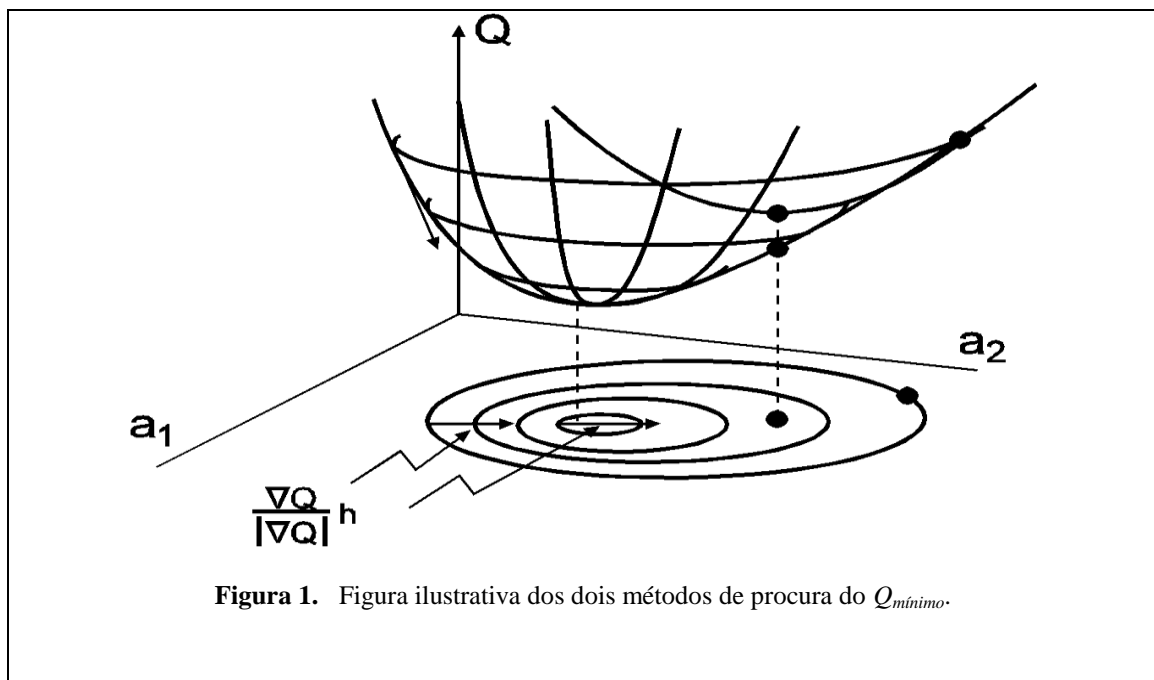


Figura 1. Figura ilustrativa dos dois métodos de procura do $Q_{\text{mínimo}}$.

ii) O segundo procedimento envolve uma expansão de segunda ordem na hiper-superfície de Q . É como se $Q(\mathbf{a})$ fosse aproximada por um "parabolóide",

$$Q(\mathbf{a}) = Q(\mathbf{a}') + \sum_{k=1}^m \left(\frac{\partial Q}{\partial a_k} \Delta a_k \right) + \frac{1}{2} \sum_{k=1}^m \sum_{l=1}^m \left(\frac{\partial^2 Q}{\partial a_k \partial a_l} \Delta a_k \Delta a_l \right),$$

com

$$\Delta a_k = a_k - a'_k,$$

do qual se localizasse o mínimo. Este método tem formalmente a estrutura do cálculo dos mínimos quadrados para funções lineares nos parâmetros $\Delta a_k = a_k - a'_k$, de modo que, para esta aproximação, há uma solução analítica. Deve-se também repetir sucessivamente este cálculo até se obter convergência. A principal desvantagem aparece quando a estimativa inicial \mathbf{a}' estiver suficientemente longe do mínimo e a curvatura da hypersuperfície $Q(\mathbf{a})$ for negativa, de maneira que se encontrará \mathbf{a}'' que maximiza Q ao

invés de minimizar. Não obstante, se se partir de uma razoável estimativa de \mathbf{a} , este método é rápido e também preciso.

b) Método iterativo de Gauss

Este método é praticamente equivalente ao de aproximar $Q(\mathbf{a})$ por um "parabolóide". Trata-se de aproximar a função g por sua expansão em primeira ordem, em torno de uma estimativa \mathbf{a}' de $\tilde{\mathbf{a}}$,

$$g(x; \mathbf{a}) \cong g(x; \mathbf{a}') + \sum_{k=1}^m \frac{\partial g(x; \mathbf{a})}{\partial a'_k} (a_k - a'_k),$$

onde

$$\left. \frac{\partial g(x; \mathbf{a})}{\partial a'_k} = \frac{\partial g(x; \mathbf{a})}{\partial a_k} \right|_{\mathbf{a}'},$$

sendo a aproximação tanto melhor quanto mais \mathbf{a}' estiver vizinho de $\tilde{\mathbf{a}}$.

Encontrar o mínimo de Q , em função da diferença $\Delta a_k = a_k - a'_k$, recai na solução de um sistema linear em Δa_k , pois

$$\frac{\partial Q}{\partial (\Delta a_k)} = \frac{\partial Q}{\partial a_k}.$$

O mínimo de Q é calculado resolvendo-se a equação

$$\mathbf{D} = \mathbf{M} \cdot \Delta \mathbf{a}$$

que tem como solução

$$\Delta \mathbf{a} = \mathbf{M}^{-1} \cdot \mathbf{D},$$

onde o vetor \mathbf{D} é dado por

$$D_k = \sum_{i=1}^N \frac{[y_i - g(x_i; \mathbf{a}')] \frac{\partial g(x_i; \mathbf{a})}{\partial a'_k}}{\sigma_i^2},$$

e a matriz \mathbf{M} , por

$$M_{kl} = \sum_{i=1}^N \frac{\frac{\partial g(x_i; \mathbf{a})}{\partial a'_k} \frac{\partial g(x_i; \mathbf{a})}{\partial a'_l}}{\sigma_i^2},$$

Esta não é, necessariamente, a solução do problema porque desprezamos os termos de ordem mais alta na série de Taylor de g . Dá, provavelmente, melhor aproximação dos parâmetros, isto é, os novos parâmetros, $\mathbf{a}'' = \mathbf{a}' + \Delta \mathbf{a}$, devem estar mais próximos de $\tilde{\mathbf{a}}$ que \mathbf{a}' .

Repetindo-se o processo, substituindo-se \mathbf{a}' por \mathbf{a}'' , chega-se mais próximo de $\tilde{\mathbf{a}}$; e a nossa expectativa é de que, após diversas tentativas, possa se obter o vetor $\Delta \mathbf{a}$ com todas as componentes nulas, mostrando a identificação de um mínimo.

A estimativa $\tilde{\mathbf{a}}$ corresponde aos valores da última iteração e pode-se mostrar que a matriz de covariância dos parâmetros é a matriz \mathbf{M}^{-1} , como no caso da regressão linear.

c) O método de Gauss-Marquardt

Este método reúne as qualidades dos métodos do gradiente e iterativo de Gauss. Proceder-se como no método de Gauss, mas os elementos diagonais de \mathbf{M} são multiplicados por um fator $f = (1 + \lambda)$ com λ positivo, ou seja,

$$M_{kl}^{\lambda} = (1 + \lambda \delta_{kl}) \cdot M_{kl}.$$

Com este fator λ controla-se o caráter da aproximação. De fato, se λ for grande, a matriz \mathbf{M}^{λ} é praticamente diagonal, forçando a procura do mínimo de Q numa direção próxima à do gradiente, e se λ for zero, recai-se no método de Gauss, mais adequado já na vizinhança do mínimo procurado.

Começa-se calculando

$$\Delta \mathbf{a} = (\mathbf{M}^{\lambda})^{-1} \cdot \mathbf{D},$$

com λ pequeno, digamos, 10^{-5} . Se $\Delta \mathbf{a}$ calculado com \mathbf{M}^{λ} leva a um aumento de Q , aumenta-se λ tipicamente de um fator 10 e recalcula-se $\Delta \mathbf{a}$. Enquanto Q calculado com a nova estimativa de \mathbf{a} não diminuir, continua-se aumentando λ . Uma vez obtida uma estimativa de \mathbf{a} que reduza Q , a matriz \mathbf{M}^{λ} e o vetor \mathbf{D} são recalculados considerando essa nova estimativa e um λ menor (tipicamente por um fator 10) e o processo todo é repetido até que fique claro que se chegou ao mínimo de Q . O valor de \mathbf{a} da última iteração é a estimativa de $\tilde{\mathbf{a}}$ e a matriz de covariância dos \tilde{a}_k é $(\mathbf{M}^0)^{-1}$, calculada com $\lambda = 0$.

Este método é bastante tolerante em relação às más estimativas, porém se nas primeiras iterações chega-se a $\lambda \geq 10$, as chances de alcançar-se um mínimo de Q são pequenas e o processo iterativo provavelmente divergirá.

d) Ainda o Teste de χ^2

Como podemos testar o ajuste efetuado particularmente se

- i) houver duas funções $g(x;\mathbf{a})$ e $h(x;\mathbf{b})$ possíveis;
- ii) Q tem mais de um mínimo, o que quase sempre acontece quando $g(x;\mathbf{a})$ não é linear em \mathbf{a} .

Procede-se ao Teste de χ^2 com um nível de significância adequado e, se for o caso, descarta-se a escolha rejeitada. Se ambas forem aceitas estatisticamente, o critério deverá ser com argumentos físicos.

EXERCÍCIOS

1. a) Ajuste os parâmetros A e λ da expressão

$$y = A e^{-\lambda t}$$

considerando os dados da tabela 11.1 abaixo, usando o artifício de linearizar a expressão.

TABELA 11.1

t (s)	y
2	89 (9)
4	65 (8)
6	63 (8)
8	44 (7)
10	38 (6)
12	28 (5)
14	30 (5)

b) Ajuste a expressão sem linearizar a equação, usando o método de Gauss.

c) Calcule χ^2 . Quantos graus de liberdade tem este ajuste? É razoável o valor obtido para χ^2 ? (Discuta qualitativamente).

2. Se a função $g(t, A, \omega)$ é linear em A mas não em ω , é possível achar-se uma equação da forma $F(\tilde{\omega}) = 0$ (independente de \tilde{A}) onde $\tilde{\omega}$ é a estimativa do método dos mínimos quadrados. Considerando a função

$$g(t, A, \omega) = A \sin \omega t$$

e que mede-se $\{ (t_i, y_i, \sigma_i) \}$ com y_i estimando $g(t_i)$:

a) Calcule

$$\frac{\partial Q}{\partial A} \text{ e } \frac{\partial Q}{\partial \omega} \quad \text{onde } Q \text{ é a soma dos quadrados dos resíduos ponderados pelas variâncias.}$$

b) Igualando essas derivadas a zero no ponto $\tilde{A}, \tilde{\omega}$, elimine \tilde{A} e obtenha uma expressão do tipo $F(\tilde{\omega}) = 0$.

Obs.: Esse método pode ser interessante por não ser iterativo, embora quase certamente a solução de $F(\tilde{\omega}) = 0$ deva ser obtida numericamente de forma iterativa.

Porém, além de haverem bons programas para resolver $F(\tilde{\omega}) = 0$, os problemas de convergência encontrados na solução dessa equação costumam ser menores. Além disso é possível fazer um gráfico de $F(\omega)$ em toda a região de interesse físico e localizar todas as soluções, o que é sempre duvidoso no método de Gauss.

3. Suponha que mediu-se a variável dependente y_i para diversos valores da coordenada independente x_i , onde a relação entre y e x é

$$y = g(\mathbf{x}; \mathbf{a});$$

\vec{a} simbolizando o conjunto dos m parâmetros a_k ,

$$g(\mathbf{x}; \vec{a}) = a_1 e^{a_2 t} \cos(a_3 t + a_4).$$

Supondo que os dados experimentais obtidos são afetados de um erro experimental ε_i ,

$$y_i = g(\mathbf{x}_i; \vec{a}^0) + \varepsilon_i,$$

onde \vec{a}^0 corresponde aos valores verdadeiros dos parâmetros, e a fdp (ε_i) é a gaussiana de média 0 e desvio-padrão σ_i , responda:

a) Qual é a função a minimizar através do método dos quadrados mínimos para achar os “melhores” valores \vec{a} dos parâmetros?

b) Por que é necessário um método iterativo para achar \vec{a} ?

c) Você poderá calcular as variâncias das estimativas \tilde{a}_k ? Como? As estimativas \tilde{a}_k serão independentes estatisticamente?

d) Supondo que você consiga determinar \vec{a} , os “melhores” valores dos parâmetros, você pode testar o resultado obtido? Como?