

## Capítulo 5

# Inferência Estatística. Teste de Hipótese.

Neste capítulo, descrevemos o que se entende por *hipótese estatística* e por *teste de hipótese*. Teste de hipótese, junto com a estimativa de parâmetros, representa a maior parte do objeto da *inferência estatística*. A definição formal de hipótese estatística - afirmação acerca da f.d.p. de uma ou mais variáveis aleatórias - não ajuda muito a compreender este assunto, de maneira que iniciamos o capítulo diretamente com um exemplo de uma hipótese e o teste estatístico relacionado.

Ao longo deste capítulo, suporemos **sempre** que a f.d.p. dos dados seja *gaussiana* e que os dados obtidos sejam independentes. Completaremos aqui o estudo das f.d.p.s de estatísticas calculadas com dados gaussianos que podem ser obtidas em forma analítica fechada, deduzindo a f.d.p. da razão entre duas estimativas da variância de uma grandeza - a f.d.p. de  $F$  de Fisher, também chamada de  $F$  de Fisher-Snedecor. Relembrando quais outras estatísticas têm f.d.p.s em forma fechada, já vimos que a média tem f.d.p. gaussiana (seção 2.6), a variância tem f.d.p. de  $\chi^2$  (seção 2.9) e a razão entre a média e o desvio padrão da média tem a f.d.p. de  $t$  de Student (seção 3.4). Este conhecimento tão detalhado e em forma fechada das f.d.p.s das estatísticas utilizadas mais comumente só é possível com dados gaussianos e certamente contribui muito para nossa tendência em supor os dados gaussianos.

Descreveremos aqui um pouco da teoria geral do teste de hipótese mas, principalmente, apresentaremos os testes relacionados às variáveis  $t$ ,  $F$  e  $\chi^2$ . Repetindo um pouco o que já foi dito no parágrafo precedente, limitamo-nos ao modelo normal, porque conhecemos bem as f.d.p.s das estatísticas, o que

facilita a aplicação de um teste de hipótese.

## 5.1 O teste $t$

A média e o desvio padrão da média da intensidade de muons por segundo, no caso do exercício 3.4 do capítulo 3, são, respectivamente

$$\bar{x} = 9,30 \quad \text{e} \quad \sigma_m = 0,54.$$

Uma teoria bastante completa sobre a radiação cósmica e sobre o detector utilizado, prevê uma intensidade detetada de muons igual a 10,5 muons por segundo. Será que os dados suportam esta previsão teórica? Ou será que os dados a contradizem? A pergunta pode tanto ser feita para saber se o modelo da radiação é correto ou se o modelo do detector (eficiência) está correto. O teste de hipótese não pode distinguir entre os dois, vai apenas comparar o resultado medido com o esperado, que é uma composição dos dois modelos.

A proposta da Estatística para realizar esta inferência a partir dos dados do exercício 3.4 é formular uma hipótese estatística - *a intensidade média do fundo de muons por segundo no detector em questão é igual a 10,5 muons por segundo* - e testá-la.

Nas circunstâncias do experimento descrito no exercício 3.4, os 5 dados obtidos podem ser supostos gaussianos, de maneira que sabemos que a variável aleatória

$$\frac{\bar{x} - x_0}{\sigma_m} = t, \tag{5.1}$$

onde  $x_0$  é o valor verdadeiro, tem f.d.p. de  $t$  de Student com  $\nu = 4$  graus de liberdade, veja seção 3.4. A partir da f.d.p. de  $t$ , podemos montar um *teste de hipótese*, como proposto.

Definiremos formalmente *teste de hipótese estatística* como uma regra que, aplicada aos dados experimentais, leva ou à decisão de aceitar a hipótese em consideração ou à decisão de rejeitá-la. Exatamente pela decisão ser extrema – ou bem aceita-se a hipótese ou bem ela é rejeitada, há possibilidade da flutuação estatística conduzir à conclusão trocada, o que será discutido ao longo do resto deste capítulo.

A hipótese sob consideração neste exemplo pode ser expressa matematicamente por

$$\text{hipótese : } x_0 = 10,50.$$

Sabemos que, a cada medida efetuada, um valor diferente de  $t$  é obtido, em razão da flutuação estatística dos dados. Caso a hipótese seja verdadeira, os valores próximos de zero são os mais prováveis, sendo raro ocorrerem valores com módulos bastante maiores que dois<sup>1</sup>. Para ser preciso, a f.d.p. de  $t$  é *exatamente* dada pela expressão (3.22) e está representada na figura 5.1.

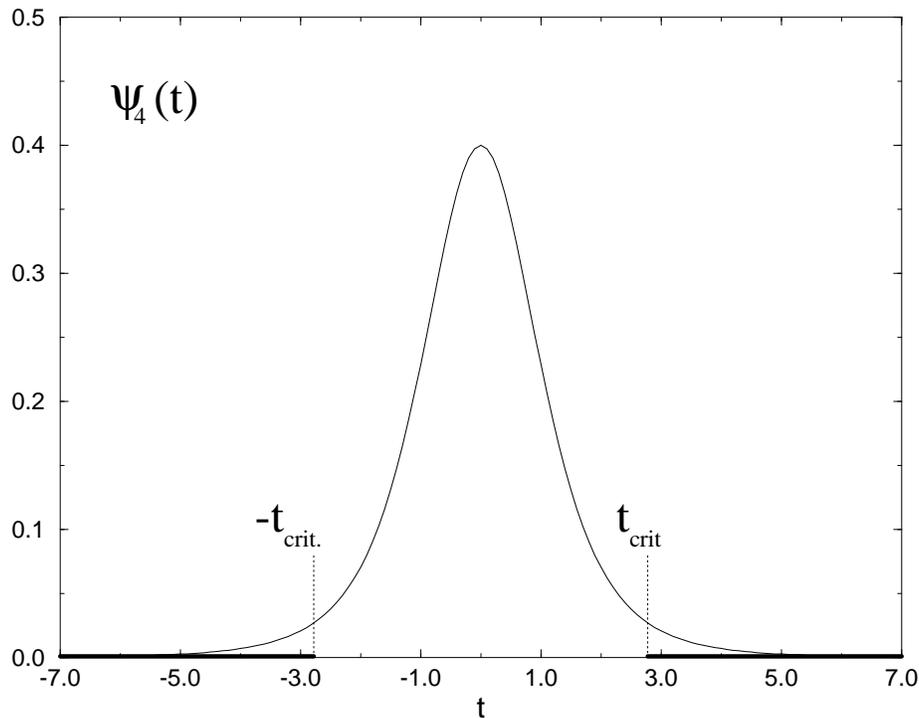


Figura 5.1: F.d.p. de  $t$  de Student, para  $\nu = 4$  graus de liberdade. O valor crítico de  $t$  está assinalado (aliás, os *valores críticos*, representados por  $t_{critico}$  e seu oposto) e a região crítica está representada como a parte do eixo das abscissas em linha grossa. A probabilidade de rejeição da hipótese, sendo ela verdadeira, é dada pela área sob a f.d.p. ao longo da região crítica.

A regra a ser aplicada e que constitui, portanto, o *teste estatístico*, consiste em aceitar a hipótese como verdadeira se  $|t|$  é pequeno e rejeitá-la, se  $|t|$  é grande. Quanto grande  $|t|$  precisa ser para que rejeitemos a hipótese? Habitualmente, escolhe-se *a priori* um valor crítico de  $|t|$ ,  $t_{critico}$ , a partir do qual consideramos falsa a hipótese. Este valor  $t_{critico}$  define uma região crítica, como ilustrada na figura 5.1, que compreende os intervalos  $] -\infty, -t_{critico}]$

<sup>1</sup>Quando o número de graus de liberdade é pequeno, podem ocorrer valores significativamente maiores que 2 com boa probabilidade, veja discussão no capítulo 3.

e  $[t_{critico}, \infty[$ . A figura 5.1 mostra também que a região crítica define uma probabilidade, que identificamos como *probabilidade de erro Tipo I*,

$$\alpha = P(|t| > t_{critico}) = \int_{-\infty}^{-t_{critico}} \psi_{\nu}(t) dt + \int_{t_{critico}}^{\infty} \psi_{\nu}(t) dt = 2 \int_{t_{critico}}^{\infty} \psi_{\nu}(t) dt, \quad (5.2)$$

onde utilizamos a simetria de  $\psi_{\nu}(t)$  para estabelecer a última igualdade em (5.2). Esta probabilidade é chamada de *nível de significância* ou também *tamanho do teste*. Ao adotarmos a postura de rejeitar a hipótese sempre que  $t$  excede  $t_{critico}$ , incorremos num *erro* numa fração  $\alpha$  das vezes em que o teste é realizado quando a hipótese é verdadeira, porque  $|t|$  supera  $t_{critico}$  com probabilidade  $\alpha$ . Este tipo de engano é chamado de *erro Tipo I*. Reflita e perceba que não pode existir um teste que nunca erre! O inconveniente de aumentar  $t_{critico}$ , de maneira que  $\alpha$  seja praticamente nulo, está relacionado a outro tipo de engano (o *erro Tipo II*) e será discutido na seção seguinte. Agora, finalizaremos o exemplo.

A escolha de  $t_{critico} = 2,78$  representada na figura 5.1 corresponde, com 4 graus de liberdade, a um nível de significância

$$\alpha = 0,05,$$

sendo comum também escolher valores críticos de maneira que  $\alpha$  seja igual a 0,01 ou a 0,001. Os valores de  $t$  correspondentes a estes valores mais habituais de  $\alpha$ , para diversos valores de  $\nu$ , estão apresentados na tabela 5.1.

Efetando finalmente os cálculos para este exemplo, levando em conta a hipótese :  $x_0 = 10,50$ , temos

$$t = \frac{\bar{x} - x_0}{\sigma_m} = \frac{9,30 - 10,50}{0,54} = -2,2,$$

o que, por ter módulo menor que 2,78, leva-nos a *aceitar* a hipótese. Dizemos que a hipótese  $x_0 = 10,50$  é aceitável ao *nível de significância de 5%*.

Note que calculamos  $t$  utilizando na representação decimal da média e do desvio padrão um algarismo significativo além dos que utilizaríamos para representar o resultado final. Como a f.d.p. de  $t$  já considera *toda* a flutuação estatística das grandezas estimadas, podemos em princípio utilizar os resultados dos cálculos com todos os dígitos que dispusermos. Por outro lado, a

---

<sup>2</sup>O número de graus de liberdade,  $\nu$ , corresponde ao número de dados,  $N$ , subtraindo-se 1 correspondente ao vínculo representado pela média,  $\nu = N - 1$  (Eq. (3.21)), veja seção 3.4.

$\alpha =$	0,05	0,01	0,001
$\nu$			
1	12,7	63,7	636
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	2,85	3,85
30	2,04	2,75	3,65
$\infty$	1,96	2,58	3,29

Tabela 5.1: Valores de  $t_\alpha$  para os quais o nível de significância,  $\alpha = 2 \int_{t_\alpha}^{\infty} \psi_\nu(t) dt$  onde  $\psi_\nu(t)$  é a f.d.p. de  $t$  de Student para  $\nu$  graus de liberdade<sup>2</sup>, tem os valores apresentados na primeira linha da tabela e que correspondem aos valores mais comuns utilizados no teste t. A primeira coluna identifica o número de graus de liberdade.

probabilidade de um certo  $t_{critico}$  ser excedido varia muito rapidamente com  $t_{critico}$ , de maneira que *necessitamos* de um pouco mais de precisão no cálculo de t. Por isso, seguimos a prática habitual, utilizando *nestes* cálculos um algarismo significativo a mais que os utilizados na apresentação do resultado final para a intensidade média experimental.

**Q5.1** *Um experimentador comprou uma pequena folha de ouro cuja pureza é garantida, sendo especificado que mais do que 99,99% do material é Au. Para verificar o grau de pureza do material, ele realiza uma medida para determinar a densidade, obtendo os seguintes dados: {19,06; 18,94; 19,08; 19,28; 19,02; 18,90} , todos em  $g/cm^3$  e obtidos a  $20^\circ C$  de temperatura. Sabendo que a densidade do Au é  $19,32(1)g/cm^3$  a  $20^\circ C$  e que verificou-se a inexistência de bolhas internas no material por meio de uma chapa de Raio-X, você concordaria que o material adquirido tem o grau de pureza especificado? O teste aplicado é sensível a pequenas misturas de outros materiais? Para fixar idéias, caso o material misturado fosse Ag ( $d_{Ag}=10,50g/cm^3$  a  $20^\circ C$ ), a partir de que proporção o teste realizado seria capaz de detetar sua mistura com Au?*

## 5.2 Erro Tipo I e erro Tipo II.

Chamemos de  $H_0$  a hipótese testada no exemplo da seção anterior, ou seja

$$H_0 \equiv \text{hipótese : } x_0 = 10,50.$$

Suponha que a natureza permitisse *apenas uma outra* possibilidade para  $x_0$ , digamos  $x_0 = 7,00$  muons/segundo. Teríamos, então, uma hipótese alternativa a considerar, que chamaremos  $H_1$ ,

$$H_1 \equiv \text{hipótese : } x_0 = 7,00.$$

Chama-se de *erro Tipo II* a aceitação *errônea* da hipótese quando ela é falsa. A probabilidade de um erro Tipo II depende da hipótese alternativa considerada, no caso  $H_1$ . A probabilidade de aceitação de  $H_0$  quando  $H_1$  é verdadeira é dada por  $\beta$ ,

$$\beta = P(|t| \leq t_{\text{critico}} \text{ quando } H_1 \text{ é verdadeira})$$

de maneira que a probabilidade complementar corresponde à probabilidade de rejeição *correta* de  $H_0$  quando  $H_1$  é verdadeira. Assim, à grandeza

$$1 - \beta = P(|t| \geq t_{\text{critico}} \text{ quando } H_1 \text{ é verdadeira}),$$

dá-se o nome de *poder do teste*.

No exemplo acima temos, supondo  $H_1$  verdadeira,  $\alpha = 0,05$  e  $t_{\text{critico}} = 2,78$ ,

$$\beta \cong 0,007 \quad (7\%)$$

no caso de um experimento com 5 dados e  $\sigma_m = 0,54$ .

Gostaríamos que o poder do teste utilizado fosse próximo a 1, assim quase sempre rejeitaríamos  $H_0$  quando  $H_1$  fosse correta. Entretanto, nota-se que o aumento do poder implica necessariamente na redução de  $t_{\text{critico}}$ , o que significa que a probabilidade de erro Tipo I aumenta. Esta relação de dependência inversa das probabilidades dos erros Tipo I e Tipo II pode ser compreendida também pelo outro lado: reduzir a probabilidade de erro Tipo I significa aumentar  $t_{\text{critico}}$ , o que aumenta a probabilidade de erro Tipo II,  $\beta$ .

A escolha de um nível de significância pequeno corresponde, então, ao experimentador tomar partido a respeito da hipótese que ele testa. Como normalmente, porém, não há outra maneira de proceder, tudo o que se pode

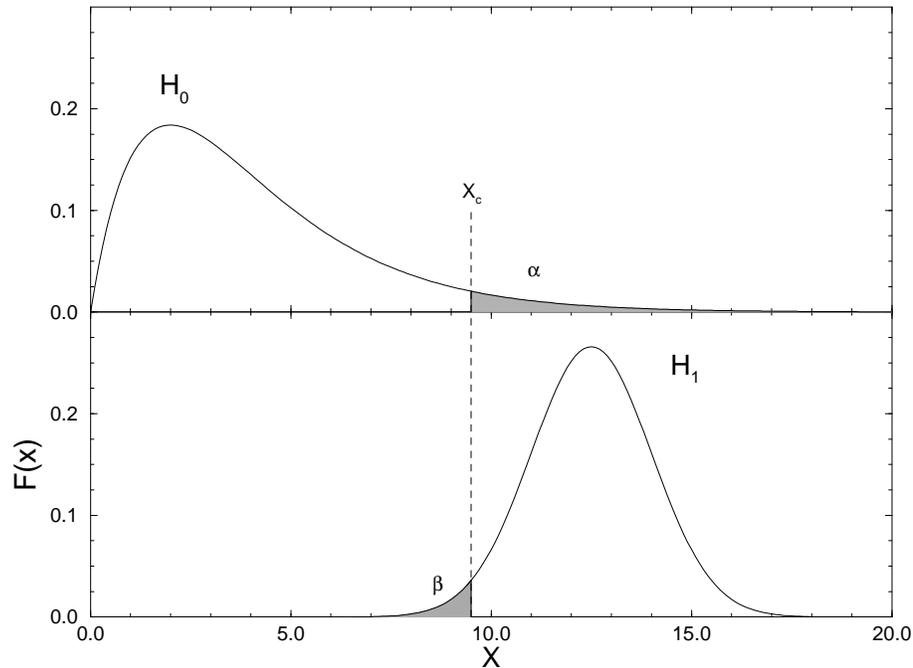


Figura 5.2: Curvas mostrando duas hipóteses em teste:  $H_0$  vs  $H_1$ .  $X_c$  é o valor crítico,  $\alpha$  a probabilidade de cometer um erro tipo I e  $\beta$  um erro tipo II

fazer é enunciar claramente a hipótese sendo testada e o nível de significância utilizado no teste.

A probabilidade de erro Tipo II está vinculado à hipótese alternativa que, na grande maioria das vezes, não pode ser formulada de maneira explícita, o que habitualmente impede o cálculo dessa probabilidade. Porém, quando a probabilidade de erro Tipo I cresce (decrece), a de erro Tipo II decresce (cresce), sendo válida esta propriedade de dependência inversa mesmo que não saibamos calcular essas probabilidades. A figura 5.2 mostra bem esta relação.  $X_c$  é o valor crítico para a hipótese  $H_0$ .  $X_c$  é escolhido tal que a probabilidade de rejeitar  $H_0$  (se  $X > X_c$ ), ou seja, de cometer um erro tipo I, vale  $\alpha$ . Este mesmo valor  $X_c$  define uma probabilidade  $\beta$  de que  $H_1$  sendo verdadeira,  $X$  seja menor que  $X_c$ , ou seja  $H_0$  seja aceita sendo falsa (erro tipo II). Baixar  $\alpha$  aumenta  $\beta$  e vice versa.

**Q5.2** *É a gravidade da consequência de um erro tipo I comparada à gravidade de um erro tipo II que define se o nível de significância a utilizar*

no teste é grande ou pequeno. Nas situações abaixo, identifique quais delas correspondem a situações onde deve-se procurar reduzir o nível de significância (ou seja, evitar erros tipo I) e quais as situações onde deve-se utilizar um nível de significância grande (ou seja, evitar erros tipo II).

- (a). Testar a hipótese segundo a qual a aceleração da gravidade é  $g_0 = 9,78m/s^2$ , em um experimento de Laboratório de Física I, efetuada por uma dentre muitas equipes de calouros.
- (b). Certificar a espessura de tábuas de construção.
- (c). Testar a resistência da estrutura de um prédio de 25 andares.
- (d). Testar a contaminação da vacina contra a pólio.
- (e). Testar a possibilidade de falha do rádio de um automóvel.
- (f). Testar a possibilidade de falha do sistema de freios de um automóvel.
- (g). Testar o ajuste de uma curva de calibração de eficiência de um sistema de detecção, recalibrado diariamente.

### 5.3 Comparação de duas médias. Ainda o teste $t$ .

Na seção 5.1 discutimos o teste  $t$  na comparação entre um resultado experimental e um possível valor para a grandeza medida. Nesta seção discutiremos a aplicação do teste  $t$  na comparação de duas medidas.

Em uma medida da capacidade de um capacitor, com um número  $n$  de dados supostamente gaussianos, o experimentador determinou o valor médio,  $\hat{x}$ , e o desvio padrão do conjunto de dados,  $\hat{\sigma}$  (o desvio padrão da média é, portanto,  $\hat{\sigma}/\sqrt{n}$ ). Um segundo experimentador mediu a capacidade desse mesmo capacitor, tomando um número  $m$  de dados supostamente gaussianos, obtendo para a média e o desvio padrão os valores  $\hat{x}'$  e  $\hat{\sigma}'$ , respectivamente.

Como os dois experimentadores utilizaram o mesmo capacitor, os resultados obtidos devem ser os mesmos, embora os números  $\hat{x}$  e  $\hat{x}'$  **certamente** sejam diferentes devido à flutuação estatística. Esperamos, porém, que eles não difiram de mais do que uma ou duas "barras de incerteza", ou ainda que eles sejam compatíveis "dentro da flutuação estatística" ou "dentro da incerteza

experimental”. Como transformar estas idéias qualitativas em uma idéia objetiva? Podemos entender o interesse desta questão, neste caso específico, como o interesse dos dois experimentadores em verificarem a existência de algum erro sistemático, o que resultaria em medidas diferentes da capacidade.

Para simplificar o teste, consideraremos que ambos os experimentadores utilizaram-se de métodos de igual precisão, de maneira que faremos a hipótese  $\sigma_0 = \sigma'_0$ . Para formalizar mais facilmente a hipótese, chamaremos de  $\mu$  e  $\mu'$  os valores verdadeiros da capacidade nas medidas efetuadas pelo primeiro e pelo segundo experimentador, respectivamente. Assim, testaremos a

$$\text{Hipótese : } \mu = \mu' \quad \text{e} \quad \sigma_0 = \sigma'_0,$$

onde  $\mu$  e  $\sigma_0$  não são especificados.

Podemos transformar esta hipótese numa outra, equivalente, e que inclui o fato de  $\mu$  não ser determinado,

$$\text{Hipótese : } \mu - \mu' = 0 \quad \text{e} \quad \sigma_0 = \sigma'_0$$

A estimativa de  $\mu - \mu'$  é  $\hat{x} - \hat{x}'$ , com variância verdadeira igual a

$$\text{var}(\hat{x} - \hat{x}') = \text{var}(\hat{x}) + \text{var}(\hat{x}') = \frac{\sigma_0^2}{n} + \frac{\sigma_0^2}{m} = \sigma_0^2 \left\{ \frac{1}{n} + \frac{1}{m} \right\} . \quad (5.3)$$

Para estimar a variância verdadeira,  $\sigma_0^2$ , utilizamos a hipótese das medidas terem o mesmo desvio padrão. Chamando de  $\theta^2$  esta estimativa global da variância, temos

$$\begin{aligned} \theta^2 &= \frac{\sum_{i=1}^n (x_i - \hat{x})^2 + \sum_{j=1}^m (x'_j - \hat{x}')^2}{n - 1 + m - 1} \\ \theta^2 &= \frac{(n - 1)\sigma^2 + (m - 1)\sigma'^2}{n - 1 + m - 1} . \end{aligned} \quad (5.4)$$

A variável aleatória  $t$  fica, então,

$$t = \frac{(\hat{x} - \hat{x}') - 0}{\theta \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\hat{x} - \hat{x}'}{\theta} \sqrt{\frac{nm}{n + m}}, \quad (5.5)$$

com f.d.p. de  $t$  de Student com  $n + m - 2$  graus de liberdade.

Assim, o procedimento para comparar as duas medidas consiste em calcular  $\theta$  pela fórmula (5.4) e, lembrando que o número de graus de liberdade a

considerar é  $n + m - 2$ , comparar  $t$  da fórmula (5.5) com o valor crítico escolhido conforme discussão das seções anteriores na tabela 5.1. Uma tabela mais completa está no apêndice.

O exercício 3 apresenta um exemplo da utilização prática deste procedimento, que normalmente requer o teste prévio da hipótese  $\sigma_0 = \sigma'_0$ , para o qual utiliza-se o teste  $F$ , que será discutido na seção 5.5 adiante. O exercício 3, portanto, só poderá ser resolvido após o estudo da seção 5.5. Já na situação onde as variâncias são diferentes, mas conhecidas, ajusta-se a média pelo método dos mínimos quadrados e utiliza-se o teste de  $\chi^2$ , como discutido neste capítulo a partir da seção 5.7. Enfim, se as variâncias são distintas e suas estimativas têm pouca precisão, o problema é muito difícil e não trataremos dele neste curso.

## 5.4 A f.d.p. de $F$ de Fisher

Deduzimos aqui a f.d.p. da razão de duas estimativas da variância, obtidas em medidas independentes.

Suponha uma medida de uma grandeza  $x$  de valor verdadeiro  $x_0$ , com  $n$  dados cuja f.d.p. é gaussiana<sup>3</sup> de variância  $\sigma^2$ . Para essa medida, a partir dos dados estima-se a variância da maneira habitual,

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \hat{x})^2, \text{ onde } \hat{x} = \frac{1}{n} \sum_{k=1}^n x_k. \quad (5.6)$$

Efetua-se uma segunda medida da mesma grandeza, com  $n'$  dados e variância  $\sigma'^2$ , que pode ser diferente de  $\sigma^2$ . Também para esta medida estima-se a variância da maneira habitual, sendo que denominamos esta estimativa por  $\hat{\sigma}'^2$ . Note que não estamos utilizando o sub-índice zero para identificar as variâncias verdadeiras e denotamos as estimativas utilizando um acento circunflexo sobre os símbolos, o que objetiva facilitar a notação no que segue. (Consistência notacional completa é muito difícil de obter e não garante clareza.)

Nesta seção, determinaremos a f.d.p. da razão  $\hat{\sigma}^2/\hat{\sigma}'^2$ , a partir das f.d.p.s de  $\hat{\sigma}^2$  e de  $\hat{\sigma}'^2$ , por meio de uma transformação muito parecida com a utilizada para calcular a f.d.p. de  $t$  de Student (seção 3.4). Analogamente ao procedimento

---

<sup>3</sup>Este detalhe foi repetido apenas por redundância pois todos os dados mencionados neste capítulo são gaussianos.

adotado no cálculo da f.d.p. de  $t$ , definimos

$$s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \quad \text{e} \quad s'^2 = \frac{1}{n'} \sum_{j=1}^{n'} (x_j - \bar{x})^2, \quad (5.7)$$

estatísticas muito parecidas com as estimativas das variâncias e mais fáceis de trabalhar. As f.d.p.s de  $s^2$  e  $s'^2$  são (usando a discussão da seção 2.9 e a fórmula (2.43))

$$f(s^2) = C \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} s^{n-3}, \quad (5.8)$$

com

$$C = \left( \frac{1}{2\sigma^2} \right)^{\frac{n-3}{2}} \frac{1}{2 \cdot \Gamma \left( \frac{n-1}{2} \right)} \quad (5.9)$$

$$g(s'^2) = C' \exp \left\{ -\frac{n's'^2}{2\sigma'^2} \right\} s'^{m'-3}, \quad (5.10)$$

com

$$C' = \left( \frac{1}{2\sigma'^2} \right)^{\frac{m'-3}{2}} \frac{1}{2\Gamma \left( \frac{m'-1}{2} \right)} \quad (5.11)$$

onde  $C$  e  $C'$  são constantes. Efetuando a transformação de variáveis

$$\left. \begin{array}{l} s^2 \\ s'^2 \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \phi = \frac{s^2}{s'^2} \\ \omega = s'^2 \end{array} \right. \quad (5.12)$$

podemos obter a f.d.p. de  $\phi$  transformando a f.d.p. conjunta de  $s$  e  $s'$  na f.d.p. conjunta de  $\phi$  e  $\omega$  e integrando esta última f.d.p. para todo  $\omega$ .

A f.d.p. conjunta de  $s^2$  e  $s'^2$  é

$$h(s^2, s'^2) = CC' \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} s^{n-3} \exp \left\{ -\frac{n's'^2}{2\sigma'^2} \right\} s'^{m'-3}.$$

O Jacobiano da transformação é

$$\frac{\partial(\phi, \omega)}{\partial(s^2, s'^2)} = \frac{1}{s'^2}.$$

Calculamos então a f.d.p.  $\eta(\phi, \omega)$  como

$$\eta(\phi, \omega) = CC' \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} s^{n-3} \exp \left\{ -\frac{n's'^2}{2\sigma'^2} \right\} s'^{m'-1},$$

onde devemos ainda trocar  $s^2 = \phi \cdot \omega$  e  $s'^2 = \omega$ . Obtemos, finalmente,

$$\eta(\phi, \omega) = CC' \exp \left\{ -\frac{n\phi\omega}{2\sigma^2} \right\} (\phi \cdot \omega)^{\frac{n-3}{2}} \exp \left\{ -\frac{n'\omega}{2\sigma'^2} \right\} (\omega)^{\frac{n'-1}{2}}$$

$$\eta(\phi, \omega) = CC' \exp \left\{ -\frac{\omega}{2} \left( \frac{n\phi}{\sigma^2} + \frac{n'}{\sigma'^2} \right) \right\} (\omega)^{\frac{n+n'-4}{2}} (\phi)^{\frac{n-3}{2}}.$$

A partir da expressão acima, determinamos a f.d.p. de  $\phi$  integrando para todo  $\omega$ ,

$$g(\phi) = CC' (\phi)^{\frac{n-3}{2}} \int_0^\infty \exp \left\{ -\frac{\omega}{2} \left( \frac{n\phi}{\sigma^2} + \frac{n'}{\sigma'^2} \right) \right\} (\omega)^{\frac{n+n'-4}{2}} d\omega.$$

Chamando de  $a$  o fator de  $\omega$  na exponencial,

$$a = \frac{1}{2} \left( \frac{n\phi}{\sigma^2} + \frac{n'}{\sigma'^2} \right)$$

independente de  $\omega$ , e de  $m$  o expoente de  $\omega$ ,

$$m = (n + n' - 4)/2 \quad ,$$

podemos efetuar uma transformação de variáveis

$$\Omega = a\omega$$

e obter

$$g(\phi) = CC' (\phi)^{\frac{n-3}{2}} a^{-m-1} \int_0^\infty \exp(-\Omega)\Omega^m d\Omega.$$

A integral da expressão acima é, simplesmente,  $\Gamma(m + 1)$ . Então,

$$g(\phi) = CC'\Gamma(m + 1) \frac{(\phi)^{\frac{n-3}{2}}}{\left( \frac{n\phi}{\sigma^2} + \frac{n'}{\sigma'^2} \right)^{\frac{n+n'-2}{2}}}$$

Precisamos voltar a utilizar como argumento da f.d.p. as estimativas não tendenciosas das variâncias e não  $s^2$  e  $s'^2$  da fórmula (5.7). Definindo

$$\nu = n - 1 \quad \text{e} \quad \nu' = n' - 1 \quad , \quad (5.13)$$

podemos escrever  $\hat{\sigma}^2$  e  $\hat{\sigma}'^2$  em função de  $s^2$  e  $s'^2$  como

$$\hat{\sigma}^2 = \frac{\nu + 1}{\nu} s^2 \quad \text{e} \quad \hat{\sigma}'^2 = \frac{\nu' + 1}{\nu'} s'^2.$$

A variável  $F$  de Fisher é definida como

$$F = \frac{\hat{\sigma}^2 \sigma'^2}{\sigma^2 \hat{\sigma}'^2} \quad , \quad (5.14)$$

que relaciona-se com  $\phi$  como

$$F = \frac{\hat{\sigma}^2 \sigma'^2}{\sigma^2 \hat{\sigma}'^2} = \frac{(\nu + 1)\nu' s^2 \sigma'^2}{(\nu' + 1)\nu s'^2 \sigma^2} = \frac{n\nu' \sigma'^2}{n'\nu \sigma^2} \phi.$$

Após alguma álgebra para ajuntar todas as constantes, mais as necessárias para transformar  $g(\phi)$  em  $f(F)$ ,

$$f(F) = \frac{g(\phi)}{\left| \frac{dF}{d\phi} \right|}$$

obtemos finalmente

$$f(F) = \frac{\nu^{\nu/2} \nu'^{\nu'/2} \Gamma\left(\frac{\nu+\nu'}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu'}{2}\right)} \frac{F^{(\nu-2)/2}}{(\nu' + \nu F)^{(\nu+\nu')/2}} \quad (5.15)$$

O valor médio de  $F$  e a variância de  $F$  são, respectivamente,

$$\langle F \rangle = \frac{\nu'}{\nu' - 2} \quad \text{para } \nu' > 2 \quad \text{e} \quad (5.16)$$

$$\text{var}(F) = \frac{2\nu'^2(\nu' + \nu' - 2)}{\nu(\nu' - 2)^2(\nu' - 4)} \quad \text{para } \nu' > 4 \quad . \quad (5.17)$$

As últimas expressões são facilmente calculadas utilizando a definição da função Beta (veja, por exemplo, [Arfken, cap. 10.4]).

$$B(m + 1, n + 1) = \int_0^{\infty} \frac{u^m}{(1 + u)^{n+m+2}} du \quad , \quad (5.18)$$

$$\text{com} \quad B(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p + q)} \quad . \quad (5.19)$$

Na seção seguinte apresentamos um exemplo de aplicação do teste F.

## 5.5 Comparação de duas variâncias estimadas. O teste F.

Para ilustrar a aplicação prática do teste F vamos usar uma situação real ocorrida por ocasião da compra de dois conjuntos de multímetros por uma oficina eletrônica e um laboratório didático.

Um fabricante especifica para seus multímetros digitais que a precisão na medida de resistências na faixa de  $32k\Omega$  é igual a 1,0% do valor medido.

Uma oficina eletrônica comprou 5 desses multímetros e decidiu verificar se a precisão dos aparelhos estava de acordo com a precisão especificada. Esta oficina dispunha de um resistor padrão de  $25,000(5)k\Omega$ , conhecido, portanto, com precisão muito superior a dos multímetros adquiridos, servindo para o teste proposto. Os dados obtidos, com os 5 multímetros, foram

$$\{25,06; 25,04; 24,97; 25,26; 25,18\}, \text{ em } k\Omega$$

(um dado para cada um dos multímetros). Destes dados, deduz-se

$$\bar{r} = 25,10 \quad \text{e} \quad \hat{\sigma} = 0,116.$$

A primeira razão de contentamento dos técnicos do laboratório foi verificar que o intervalo  $[\bar{r} - t_{II}\sigma_m; \bar{r} + t_{II}\sigma_m] = [24,95; 25,25]$  contém o valor verdadeiro, sendo que concordamos com seu contentamento (utilizamos  $t_{II}$  da tabela 3.3). Já a segunda razão de alegria com a compra, foi verificar que seu lote de multímetros apresentou uma precisão de 0,4%, melhor que a especificada pelo fabricante, *com o que não concordamos*. O procedimento correto para verificar se a precisão é melhor que a especificada consiste em formular a hipótese

$$H : \sigma = 0,25$$

e testá-la. O teste consiste em estabelecer um valor crítico para  $\hat{\sigma}$  e verificar, utilizando a f.d.p. de  $\hat{\sigma}$ , se é possível aceitar H. A f.d.p. da variância é a mesma de  $\chi^2$ , e dedicaremos as demais seções a estudar esse teste, de maneira que completaremos esta parte do exemplo adiante, adiantando que a hipótese H é aceita com um nível de significância grande.

A polêmica ocorrida e que pretendemos resolver nesta seção precisa ainda ser descrita. Um laboratório didático adquiriu 10 desses aparelhos e quis verificar se também seu lote tinha a *mesma* precisão que o lote adquirido pela oficina eletrônica. Usando o mesmo resistor padrão obteve-se os seguintes dados,

$$\{25,22; 24,81; 24,56; 24,98; 24,73; 24,93; 24,79; 25,39; 25,34; 25,07\}(k\Omega).$$

### 5.5. COMPARAÇÃO DE DUAS VARIÂNCIAS ESTIMADAS. O TESTE F.131

De novo, a partir de  $\bar{r}' = 24,98$  e  $\hat{\sigma}' = 0,273$ , determinados a partir dos 10 dados obtidos, nota-se que o valor verdadeiro está contido no intervalo em torno da média com largura igual a um desvio padrão da média,  $\sigma_m = 0,086$ . A precisão obtida também é evidentemente compatível com a especificada pelo fabricante. Porém, comparando seus resultados com os da oficina eletrônica, os técnicos do laboratório ficaram descontentes com seu lote de aparelhos, afirmando que *tinham pior precisão que os da oficina*.

Utilizando os métodos da inferência estatística, estabelecemos, então, a hipótese:

$$\mathcal{H} : \sigma = \sigma' , \text{ com } \sigma \text{ desconhecido}$$

Como a disputa envolve apenas a diferença de precisão obtida, independente da especificação do fabricante (o que foi testado por cada grupo), incluímos na hipótese que  $\sigma$  é desconhecido, o que facilita o teste. A variável aleatória

$\nu =$	1	2	3	4	9	19
$\nu'$						
1	161	200	225	230	240	248
2	18,5	19,0	19,2	19,3	19,4	19,4
3	10,1	9,55	9,28	9,12	8,8	8,67
4	7,71	6,94	6,59	6,39	6,00	5,81
9	5,12	4,26	3,86	3,63	3,18	2,95
19	4,38	3,52	3,13	2,90	2,42	2,17

Tabela 5.2: Valores de  $F_\alpha$  para os quais  $\int_{F_\alpha}^{\infty} f(F)dF = 0,05$ , para alguns valores de  $\nu$  e  $\nu'$ , números de graus de liberdade para o numerador e o denominador, respectivamente, da razão  $F$  de duas variâncias estimadas a partir de dados gaussianos, com numerador maior que o denominador.

$$F = \frac{\hat{\sigma}'^2 \sigma^2}{\sigma'^2 \hat{\sigma}^2} = \frac{\hat{\sigma}'^2}{\hat{\sigma}^2}, \quad (5.20)$$

onde utilizamos a hipótese, é distribuída como  $F$  de Fisher, sendo que o número de graus de liberdade do numerador é  $\nu = 9$  e do denominador,  $\nu' = 4$ . Conhecendo a f.d.p. de  $F$ , fórmula (5.15), baseamos o teste nessa f.d.p., sendo o nível de significância dado por  $\alpha$ ,

$$\alpha = \int_{F_{criti}}^{\infty} f(F)dF \quad .$$

A tabela 5.2 apresenta os valores críticos de  $F$ , para alguns valores de  $n$  e  $m$  ( $\nu$  e  $\nu'$ ), quando o tamanho do teste é 5%. A tabela foi construída de maneira a utilizar como numerador sempre a maior estimativa. Da tabela 5.2, obtemos  $F_{critico} = 6,00$  e utilizando os resultados experimentais na fórmula (5.20) obtemos

$$F = 5,54 \quad ,$$

o que significa que *aceitamos*  $\mathcal{H}$  ao nível de significância de 5%. Interpretamos, então, a diferença nas estimativas como resultado exclusivo da flutuação estatística.

## 5.6 Um teste qualitativo do ajuste de uma função.

No capítulo precedente, ajustamos uma função  $h(\mathbf{x}; \mathbf{p})$  a dados experimentais  $\{(x_i, y_i, \sigma_i), i = 1, \dots, N\}$ . O vetor  $\mathbf{p}$  representa os parâmetros a serem determinados e  $\mathbf{x}$  a variável independente. Os dados  $y_i$  são tais que

$$y_i = h(\mathbf{x}, \mathbf{p}_0) + \epsilon_i,$$

onde  $\mathbf{p}_0$  representa os valores verdadeiros dos parâmetros e  $\epsilon_i$  os erros experimentais, com a propriedade

$$\langle \epsilon_i^2 \rangle = \sigma_i^2.$$

Veja as seções 4.5 a 4.8 para os detalhes do modelo.

Pretendemos avaliar se o ajuste da função aos pontos experimentais foi bom ou ruim. Uma idéia mais precisa e objetiva exige a utilização dos conceitos discutidos nas seções precedentes deste capítulo. Entretanto, pode-se verificar o ajuste por meio de um critério qualitativo baseado na distância do gráfico da função ajustada aos pontos experimentais, considerando-se os desvios padrões  $\sigma_i$ .

Se a f.d.p. do erros  $\epsilon_i$  é a gaussiana, podemos calcular a probabilidade de qualquer intervalo conter o valor verdadeiro, em particular, calculamos as probabilidades :

$$\left. \begin{aligned} P(y_i - \sigma_i \leq h(\mathbf{x}_i, \mathbf{P}_0) \leq y_i + \sigma_i) &= 68,3\% \\ P(y_i - 2\sigma_i \leq h(\mathbf{x}_i, \mathbf{P}_0) \leq y_i + 2\sigma_i) &= 95,5\% \\ P(y_i - 3\sigma_i \leq h(\mathbf{x}_i, \mathbf{P}_0) \leq y_i + 3\sigma_i) &= 99,7\% \end{aligned} \right\} \quad (5.21)$$

Isto significa que, **em média**, o gráfico da **função verdadeira** deve: cortar cerca de 68,3% das barras de incerteza; passar a menos de 2 barras de incerteza em cerca de 95,5% dos pontos; passar a menos de 3 barras de incerteza

na quase totalidade dos pontos. Lembre-se que as incertezas são representadas convencionalmente por uma barra que estende-se 1 desvio padrão acima do valor experimental e 1 desvio padrão abaixo do valor experimental. É importante notar que estes números (68% dos pontos, 95% dos pontos, etc) são números médios, o valor real observado num ajuste particular é afetado pela flutuação estatística, de maneira análoga ao que discutimos na seção 2.1.

Infelizmente, não temos o gráfico da função verdadeira, apenas o gráfico da função ajustada,

$$h(\mathbf{x}, \hat{\mathbf{p}}) \sim h(\mathbf{x}, \mathbf{p}_0).$$

O resultado acima deixa de valer exatamente e corresponde a uma *subestimação* das frações médias reais dos pontos cujas barras de incerteza serão cortadas, para os quais a função ajustada passará a menos de 2 barras de incerteza, etc. Isto porque a curva ajustada segue os pontos experimentais melhor que a verdadeira justamente por ter sido ajustada a eles.

Como aproximação, porém, as probabilidades dadas em (5.21) valem, de maneira que as utilizamos substituindo  $h(\mathbf{x}_i; \mathbf{p}_0)$  por  $h(\mathbf{x}_i, \hat{\mathbf{p}})$  e trocando as igualdades por aproximações.

O fato do resultado (refeq:intervalosnormais) ser aproximado, aliado à flutuação estatística das frações observadas, atrapalham a aplicação deste método simples, de contagem de pontos a uma, duas, e três barras de incerteza da curva ajustada, de forma conclusiva. Como método qualitativo, porém, é muito útil. Em particular, quando o número de graus de liberdade é muito maior que o número de parâmetros ajustados, a aproximação é excelente e este método pode fornecer respostas claras, no sentido de revelar se o ajuste é adequado ou não e se os desvios padrões foram subestimados.

## 5.7 Teste quantitativo da qualidade de um ajuste de função. O teste de $\chi^2$ .

O teste que avalia quantitativamente o critério da seção precedente é o teste de  $\chi^2$ . Utilizemos a mesma notação e consideremos que há um número  $\mu$  de parâmetros representados pelo vetor  $\mathbf{p}$ ,

$$\mathbf{p} = (p_1, p_2, \dots, p_\mu) \quad .$$

Após efetuarmos o ajuste da função, sendo  $\hat{\mathbf{p}}$  o vetor dos parâmetros ajustados, podemos calcular a variável aleatória

$$Q_{\text{minimo}} = \sum_{i=1}^n \frac{[y_i - h(\mathbf{x}_i; \hat{\mathbf{p}})]^2}{\sigma_i^2} \quad (5.22)$$

Esta variável tem a f.d.p. de  $\chi^2$  com

$$\ell = N - \mu \quad (5.23)$$

graus de liberdade,  $F_{N-\mu}(\chi^2)$ , da fórmula (2.43). Estamos supondo que a f.d.p. do erro

$$\epsilon_i = y_i - h(\mathbf{x}_i; \mathbf{p}_0) \quad (5.24)$$

seja gaussiana de desvio padrão  $\sigma_i$ . O erro  $\epsilon_i$  tem f.d.p. gaussiana porque  $y_i$  é gaussiano, como aliás são todos os dados mencionados neste capítulo. Descontar o número de parâmetros,  $\mu$ , do número de pontos para obter o número de graus de liberdade da f.d.p. de  $\chi^2$ , é a correção necessária pela utilização da função ajustada,  $h(\mathbf{x}; \hat{\mathbf{p}})$ , no lugar da função verdadeira,  $h(\mathbf{x}, \mathbf{p}_0)$ , como demonstraremos no capítulo 9.

Como a f.d.p. de  $\chi^2$  concentra-se em torno do valor médio, ao menos quando há vários graus de liberdade, esperamos obter  $Q_{\text{minimo}} \sim \ell$ , sendo pouco prováveis valores próximos de 0 ou valores muito maiores que  $\ell$ . Não podemos usar, porém, o valor de  $F_{\ell}(\chi^2)$  no ponto  $\chi^2 = Q_{\text{minimo}}$  como medida de aderência da função ajustada aos pontos experimentais porque um valor determinado de  $\chi^2$  tem probabilidade estritamente nula. Para contornar esta dificuldade, utilizamos a teoria de teste de hipótese desenvolvida nas seções anteriores. Relembrando, a sequência esquemática de um teste de hipótese é:

- (i). Formula-se a hipótese estatística;
- (ii). Escolhe-se uma estatística<sup>4</sup> com f.d.p. conhecida, considerando-se verdadeira a hipótese, e
- (iii). Escolhe-se uma região crítica para a rejeição da hipótese.

A hipótese que formularemos é

---

<sup>4</sup>Estatística, aqui, tem o significado de função cujas únicas variáveis aleatórias são os dados

H : os erros  $\epsilon_i$  tem f.d.p. gaussiana de média 0 e desvio padrão  $\sigma_i$  (5.25)

Adotando a hipótese, o valor de  $Q_{\text{minimo}}$  tem f.d.p. de  $\chi^2$  com  $\ell = N - \mu$  graus de liberdade, conforme a discussão do início desta seção. Há duas possibilidades distintas para a escolha da região crítica, que correspondem a possibilidades diferentes de violação da hipótese. Embora ambas precisem ser testadas, costumamos testá-las isoladamente.

Note que  $Q_{\text{minimo}}$  é um somatório de razões. Assim, um valor elevado de  $Q_{\text{minimo}}$  pode decorrer tanto de numeradores excessivamente grandes quanto de denominadores pequenos. Já um valor muito pequeno de  $Q_{\text{minimo}}$  deve decorrer de denominadores excessivamente grandes, não havendo razão para os numeradores serem responsáveis pela pequenez de  $Q_{\text{minimo}}$ , uma vez que os numeradores dependem basicamente da adequação da função ajustada.

Se tentamos ajustar uma função de tipo diferente da função verdadeira, estamos fazendo com que o numerador seja sistematicamente aumentado – ele é a superposição de um erro estatístico à diferença entre o valor de duas funções – o que resulta em aumento de  $Q_{\text{minimo}}$ . Quando escolhemos uma região crítica do tipo

$$[\chi_{\text{critico}}^2, +\infty[,$$

verificamos principalmente se a função ajustada é adequada.

É claro da discussão acima que há uma outra maneira da hipótese ser falsa que resulta em  $Q_{\text{minimo}}$  elevado, que corresponde à *subestimação das variâncias*.

Já ao testarmos uma região crítica do tipo

$$[0, \chi_{\text{critico}}^2],$$

estamos verificando principalmente se há *superestimação das variâncias*.

Discutiremos separadamente a utilização de cada uma das regiões críticas. A figura 5.3 mostra as duas regiões de exclusão para uma curva de  $\chi^2$  com 10 graus de liberdade.

## 5.8 Teste para $\chi^2$ alto

Testar a hipótese (5.25) na interpretação do título desta seção corresponde a escolher como região crítica o intervalo

$$[\chi_{\text{critico}}^2, +\infty[,$$

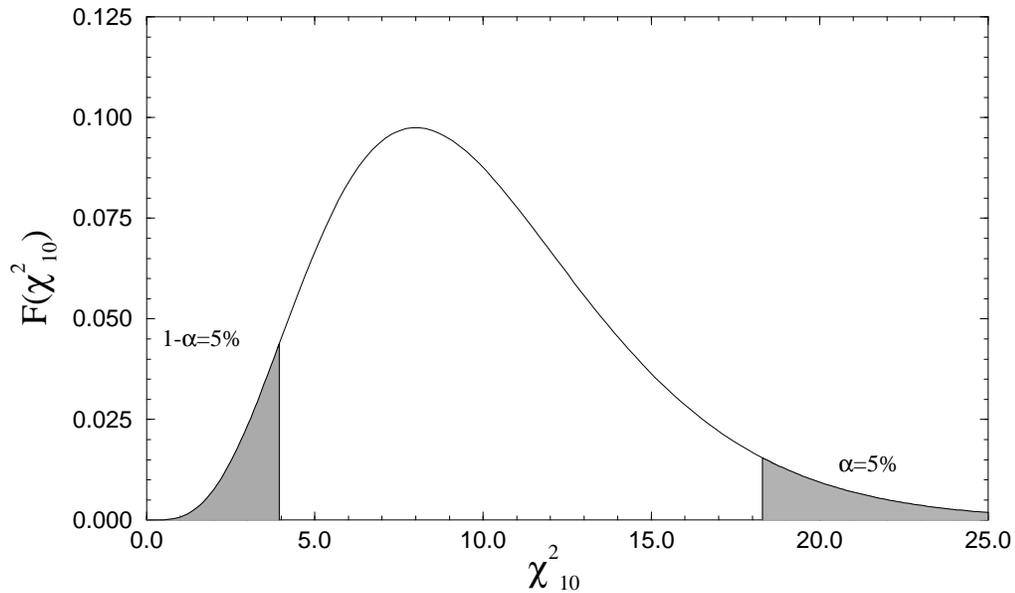


Figura 5.3: Curva de  $\chi^2_{10}$  com regiões de exclusão de 5%. A região à esquerda exclui um ajuste por ter um  $Q_{\text{mínimo}}$  baixo (variâncias superestimadas), caso discutido na seção 5.9. A região à direita exclui um ajuste por ter um  $Q_{\text{mínimo}}$  alto (variâncias subestimadas ou função inadequada), caso discutido na seção 5.8.

O nível de significância do teste é

$$\alpha = P(\chi^2 \geq \chi^2_{\text{crítico}} \text{ com } \ell \text{ g.l.}) = \int_{\chi^2_{\text{crítico}}}^{\infty} F_{\ell}(\chi^2) d\chi^2. \quad (5.26)$$

A integral da f.d.p. de  $\chi^2$ , analogamente às de  $t$  e  $F$ , não tem forma analítica, por isso recorreremos a tabelas. A tabela 5.3 apresenta alguns valores de  $\chi^2_{\text{crítico}}$  para os níveis de significância utilizados mais comumente e para alguns números de graus de liberdade. (Uma tabela mais completa aparece no apêndice.) Para um número grande de graus de liberdade, pode-se usar a aproximação dada no fim do capítulo 2, fórmula (2.54). Adiante discutiremos a escolha do nível de significância.

Quando realizamos um único ajuste de função, é muito comum utilizar-se o teste de  $\chi^2$  sem escolher-se explicitamente um nível de significância. Pode-se seguir o procedimento descrito no segundo tópico à parte da seção 5.2, calculando qual seria o maior nível de significância com o qual  $H$  poderia ser rejeitada, que corresponde ao valor acumulado da probabilidade de  $\chi^2$  numa

$\alpha$	0,05	0,01	0,001	$\alpha$	0,05	0,01	0,001
$\ell$				$\ell$			
1	3,84	6,63	10,8	8	15,5	20,1	26,1
2	5,99	9,21	13,8	9	16,9	21,7	27,9
3	7,81	11,3	16,3	10	18,3	23,2	29,6
4	9,49	13,3	18,5	15	25,0	30,6	37,7
5	11,1	15,1	20,5	20	31,4	37,6	45,3
6	12,6	16,8	22,5	25	37,7	44,3	52,6
7	14,1	18,5	24,3	30	43,8	50,9	59,7

Tabela 5.3: Valores críticos de  $\chi^2$  para alguns níveis de significância,  $\alpha$ , e alguns valores do número de graus de liberdade,  $\ell$ .

região crítica definida pelo valor  $Q_{\text{minimo}}$ ,

$$\alpha_{\text{maximo}} = P(\chi^2 \geq Q_{\text{minimo}} \text{ com } \ell \text{ g.l.}) = \int_{Q_{\text{minimo}}}^{\infty} F_{\ell}(\chi^2) d\chi^2. \quad (5.27)$$

Interpretamos este valor como a probabilidade da flutuação estatística originar um conjunto de dados experimentais *ao menos* tão distante da função  $h(\mathbf{x}, \mathbf{p})$  quanto a medida particular para a qual o ajuste foi efetuado. Aceitamos, então, o ajuste para o qual  $\alpha_{\text{maximo}}$  é grande, digamos 5% ou mesmo 1%.

Imaginemos que o teste efetuado conduza à rejeição da hipótese, seja porque escolhemos um valor crítico que foi ultrapassado por  $Q_{\text{minimo}}$ , seja porque  $\alpha_{\text{maximo}}$  é muito pequeno. Claro que esta rejeição pode corresponder a um erro tipo I, mas, se excluirmos esta "má sorte" só nos restam suas possibilidades:

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

Em princípio, o engano correspondente ao caso (i) deve ser investigado, voltando-se à etapa de obtenção dos dados experimentais e conferindo os desvios-padrões, seja efetuando várias observações e estimando os desvios padrões, seja conferindo se todas as fontes de erros foram incluídas no cálculo de  $\sigma_i$ . Além disso, o caso (i) pode ter solução, dentro da hipótese da função ajustada ser adequada e os desvios padrões dos pontos experimentais serem bem conhecidos a menos de um fator comum, estando errados apenas os valores absolutos das variâncias. Trataremos deste caso no capítulo 9 mas adiantamos

que ele resulta na necessidade de *umentar* os desvios padrões dos parâmetros estimados.

Já o caso (ii) obriga-nos a buscar outra forma para  $h(\mathbf{x}, \mathbf{p})$ . De qualquer maneira, este teste **apenas** não permite distinguir se o que está errado é a função ou o conjunto das variâncias.

## 5.9 Teste para $\chi^2$ baixo

Testar a hipótese (5.25) quando pretendemos avaliar se houve superestimação das variâncias corresponde a escolher como região crítica o intervalo

$$[0, \chi_{critico}^2].$$

O nível de significância do teste é

$$\alpha = P(\chi^2 \leq \chi_{critico}^2 \text{ com } \ell \text{ g.l.}) = \int_0^{\chi_{critico}^2} F_{\ell}(\chi^2) d\chi^2. \quad (5.28)$$

A tabela 5.4 apresenta alguns valores de  $\chi_{critico}^2$  para o nível de significância de

$\ell$	$\chi_{critico}^2$			$\ell$	$\chi_{critico}^2$		
	0,1%	1%	5%		0,1%	1%	5%
1	$1,5710^{-6}$	0,00016	0,0039	8	0,856	1,65	2,74
2	0,002	0,0201	0,104	9	1,152	2,09	3,32
3	0,024	0,115	0,351	10	1,48	2,56	3,94
4	0,920	0,297	0,712	15	3,48	5,23	7,26
5	0,210	0,554	1,145	20	5,92	8,26	10,9
6	0,384	0,872	1,63	25	8,65	11,5	14,6
7	0,595	1,24	2,17	30	11,6	15,0	18,5

Tabela 5.4: Valores críticos de  $\chi^2$  para o nível de significância de 0,1%, 1% e 5%, num teste baseado na cauda esquerda da f.d.p. de qui-quadrado, para alguns valores do número de graus de liberdade,  $\ell$

0,1%, 1% e 5%. Caso chegue-se à conclusão que a hipótese deva ser rejeitada, excluindo-se a possibilidade de um erro tipo I, devemos buscar a origem da superestimação das variâncias. Por um lado pode-se efetuar os testes descritos na seção precedente. Por outro, deve-se considerar a possibilidade dos dados serem *covariantes*. Pode-se mostrar que a correlação entre dados experimentais

tendem, frequentemente, a ser positiva e veremos no capítulo 9 que  $Q_{\text{minimo}}$  aumenta com o aumento da correlação (equação 9.58).

Um exemplo de tratamento de dados que conduz a uma superestimação das variâncias é o da medida de intensidade corrigida por uma calibração de eficiência que é função da variável independente. É um engano comum propagar a incerteza de calibração, que afeta todo o conjunto dos  $y_i$ , nas variâncias  $\sigma_i^2$ . Covariâncias em geral têm o mesmo efeito sobre o valor de  $Q_{\text{minimo}}$ . Se esses erros de calibração afetam todos os pontos de maneira regular, são erros sistemáticos e não devemos aumentar as variâncias individuais para levá-los em conta. Já se afetam diferentemente os vários pontos, é preciso calcular essas covariâncias e incluí-las no ajuste, o que aprenderemos a fazer no capítulo 9.

## 5.10 Teste de $\chi^2$ : nível de significância

Quando temos apenas um ajuste a considerar, é habitual considerar-se adequado se

$$0,05 \leq P(\chi_\ell^2 \geq Q_{\text{minimo}}) \leq 0,95$$

Entretanto, ao efetuarmos centenas de ajustes, mesmo escolhendo um nível relativamente pequeno como 0,01, provavelmente haverá um, ou alguns, ajustes com  $\chi^2$  dentro da região crítica sem que a hipótese H seja falsa, ou seja, sem que haja algo errado. Simplesmente, o número médio de erros tipo I cresce com o número de testes efetuados.

Ao escolher o nível de significância do teste de  $\chi^2$ , é preciso levar em conta o número de ajustes. Em particular, se temos muitos ajustes, normalmente precisamos reduzir esse nível. Vamos discutir este assunto através de um exemplo.

Imagine uma situação onde realizamos  $M=150$  ajustes de funções. Em cada ajuste calculamos  $\chi_{\text{minimo}}^2$  e aceitamos ou rejeitamos a hipótese (5.25) – portanto o ajuste – comparando este valor obtido com o valor crítico,  $\chi_{\text{critico}}^2$ . Escolhendo esse valor crítico de maneira que a probabilidade de obter-se um  $\chi^2$  acima desse valor corresponda ao nível de probabilidade  $p = 5\%$ ,

$$P(\chi_\ell^2 \geq \chi_{\text{critico}}^2) = \int_{\chi_{\text{critico}}^2}^{\infty} f_\ell(\chi^2) d\chi^2 = p = 0,05 \quad ,$$

a função de probabilidade do número  $m$  de casos que cairão na faixa crítica de  $\chi^2$ ,  $[\chi_{critico}^2, \infty[$ , é dada pela binomial

$$P_{M,p}(m) = \binom{M}{m} 0,05^m 0,95^{M-m} \text{ com } M=150,$$

que nos diz que teremos algo em torno de 7 a 8 ajustes nessa faixa sem que nem  $\sigma_i^2$  nem a função sejam inadequadas. Esse ajustes seriam, portanto, rejeitados equivocadamente (erro tipo I). Neste caso, o erro tipo II corresponde à aceitação de um ajuste inadequado mas que passou pelo teste porque a flutuação estatística deixou os pontos experimentais  $y_i$  de maneira a assemelharem-se à função  $h(\mathbf{x}, \mathbf{p})$ .

É claro que um aumento do valor crítico de  $\chi^2$  tem dois efeitos : reduzir a probabilidade de erro tipo I e aumentar a probabilidade de erro tipo II<sup>5</sup>. Não podemos, então, aumentar impunemente o valor crítico de  $\chi^2$ . Uma possibilidade consiste em escolher um valor crítico tal que a probabilidade *global* de ocorrer um certo número de erros tipo I seja prefixada. Um critério já utilizado corresponde em fixar em 50% a probabilidade de não haver nenhum erro tipo I - o que equivale em fixar também em 50% a probabilidade de haver ao menos um erro tipo I. Neste exemplo, efetuando os cálculos, queremos

$$P(n = 0) = 0,5 \Leftrightarrow$$

$$\binom{M}{0} (1-p)^M = 0,5 \Leftrightarrow$$

$$M \ln(1-p) = \ln(0,5) \Leftrightarrow (\text{supondo } p \ll 1)$$

$$p \cong 0,7/M \cong 0,5\%, \text{ substituindo } M=150.$$

Portanto, se fixarmos em 0,5% a probabilidade de erro tipo I em um único ajuste, teremos fixado em 50% a probabilidade de não cometermos nenhum erro tipo I na avaliação do conjunto dos 150 ajustes. Em 150 ajustes, então, esperamos ou não obter *nenhum*  $Q_{minimo}$  na região crítica ou obter *um* ou *dois* ajustes resultando em  $Q_{minimo}$  acima de  $\chi_{critico}$ .

---

<sup>5</sup>A probabilidade de cometer um erro tipo II depende da existência de uma hipótese alternativa. Uma vez conhecida esta hipótese, pode-se calcular a probabilidade associada.

## 5.11 Utilização do qui-quadrado reduzido.

Sabemos que  $\langle \chi^2 \rangle = \ell$  (fórmula 2.44), onde  $\ell$  é o número de graus de liberdade, o que origina a regra habitual segundo a qual o *ajuste é bom quando o qui-quadrado reduzido,  $\chi_{red}^2$*

$$\chi_{red}^2 = \frac{\chi^2}{\ell} \quad (5.29)$$

é próximo de 1, uma vez que este é seu valor médio. Em muitos casos, entretanto, obter o valor de qui-quadrado reduzido igual a 1, ou melhor, numa faixa estreita em torno de 1, digamos no intervalo  $[0,9; 1,1]$  é pouco provável. É necessário avaliar o quanto o valor de qui-quadrado reduzido pode diferir de 1 por flutuação estatística.

Podemos utilizar o desvio padrão como medida da flutuação estatística. Na seção 2.9, fórmula (2.45), vemos que o desvio padrão da f.d.p. de  $\chi^2$  é  $\sqrt{2\ell}$ . Usando como orientação genérica que os valores mais prováveis são aqueles numa faixa em torno do valor médio de mais, ou menos, dois desvios padrões, teríamos como valores mais prováveis de qui-quadrado

$$[\ell - 2\sqrt{2\ell}; \ell + 2\sqrt{2\ell}] \quad (5.30)$$

o que corresponde à faixa de qui-quadrado reduzido

$$\left[ 1 - 2\sqrt{\frac{2}{\ell}}; 1 + 2\sqrt{\frac{2}{\ell}} \right] \quad (5.31)$$

Assim, para um número pequeno de graus de liberdade, a faixa de qui-quadrado reduzido provável de obter-se é larga e, para um número grande, é estreita. Por exemplo, para 10 graus de liberdade, esta faixa é  $[0,1; 1,9]$  e para 200 graus de liberdade é  $[0,8; 1,2]$ .

Esta regra deve ser utilizada com muita cautela para um número pequeno de graus de liberdade (menor que 30) devido a assimetria da f.d.p. de  $\chi^2$ , mas ela é preferível ao hábito de utilizar-se apenas o número 1 como guia.

Embora aproximada, a regra (5.31) acima fornece intervalos razoavelmente adequados. O que não se pode fazer é tentar estabelecer um intervalo de qui-quadrado reduzido que valha para **qualquer** número de graus de liberdade. O que muitos experimentadores utilizam é que, em suas medidas, o número de graus de liberdade é mais ou menos constante e, como o intervalo de qui-quadrado reduzido varia pouco com o número de graus de liberdade **para**

**valores próximos e razoavelmente grandes**, é possível estabelecer uma faixa de qui-quadrado reduzido aceitável para **suas condições experimentais particulares**.

## EXERCÍCIOS

1. Em um laboratório didático 6 grupos de estudantes mediram a aceleração da gravidade local obtendo os resultados 8,5; 8,0; 9,2; 8,7; 8,8; 10,3; (em  $m/s^2$ ). Supondo que esses resultados obedeam a uma mesma f.d.p. gaussiana, use o teste  $t$  para verificar se esses dados são compatíveis com o valor  $g = 9,807m/s^2$  dentro de níveis de significância de 5%, 1% e 0,1%.
2. Outros 6 grupos de estudantes obtiveram, para a aceleração da gravidade no mesmo local (e também em  $m/s^2$ ) os valores 9,2; 9,9; 9,9; 9,5; 10,5; 10,0.
  - (a) Use o teste F para verificar se este conjunto de dados pode ter o mesmo desvio padrão que o conjunto de dados do exercício anterior (com nível de significância de 5%).
  - (b) Supondo que o desvio padrão destes dados e o do exercício anterior sejam iguais, use o teste  $t$  para verificar se ambas as medidas podem corresponder a uma mesma grandeza física.
3. Uma das características de um nuclídeo instável é o espectro de radiação gama que emite. Normalmente, uma espécie radioativa particular emite fótons com algumas energias características, sendo as energias e também as intensidades de emissão características particulares da sua espécie, servindo de identificação. Além disso, cada espécie radioativa apresenta um tempo característico de decaimento, normalmente descrito pela *meia vida*.

Numa medida de espectroscopia gama em que diversos nuclídeos estão presentes, as espécies podem ser identificadas, então, pelas energias das radiações emitidas, sendo que as intensidades relativas dessas radiações não dependem do tempo. Isto permite que um conjunto de medidas consecutivas, separadas por intervalos de tempo comparáveis à meia-vida, defina se um certo conjunto de radiações provém do mesmo nuclídeo – na hipótese delas provirem de uma mesma espécie radioativa: as intensidades relativas dessas radiações devem permanecer constantes ao longo da sequência de medidas.

Uma fonte radioativa, contendo uma ou talvez duas espécies radioativas, foi observada emitindo radiações das seguintes energias: 127, 196, 238, 307 e 507 keV. Mediu-se a intensidade relativa à transição de 307 keV, que sabe-se ser característica do nuclídeo  $^{164m}\text{Ho}$ , em cinco espectros. A tabela abaixo resume os resultados obtidos em medidas consecutivas, identificadas pelas letras A a E.

$E_\gamma$	A	B	C	D	E
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)

Tabela de intensidade no espectro (exercício 3)

Para cada energia, determine a intensidade média, seu desvio padrão e  $\chi^2$ . Tendo em vista o resultado do teste de  $\chi^2$ , a hipótese destas radiações originarem-se de um único nuclídeo é razoável? (Estes são dados reais, não são simulação)

4. As figuras I, II e III mostram os dados experimentais, com suas respectivas barras de incerteza iguais ao desvio padrão, obtidos pela observação da posição de um corpo em Movimento Retilíneo e Uniforme. A equação horária é

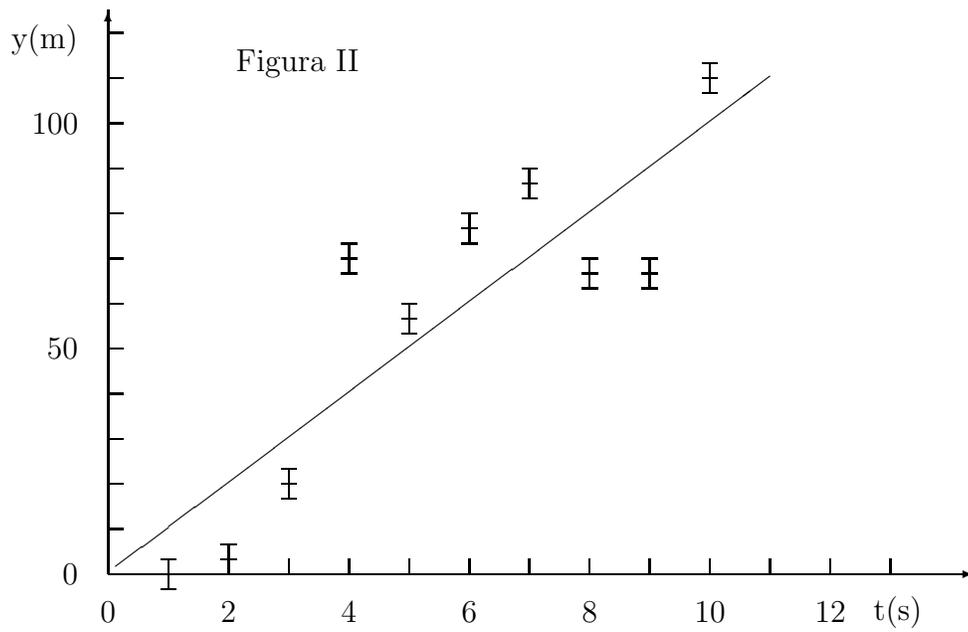
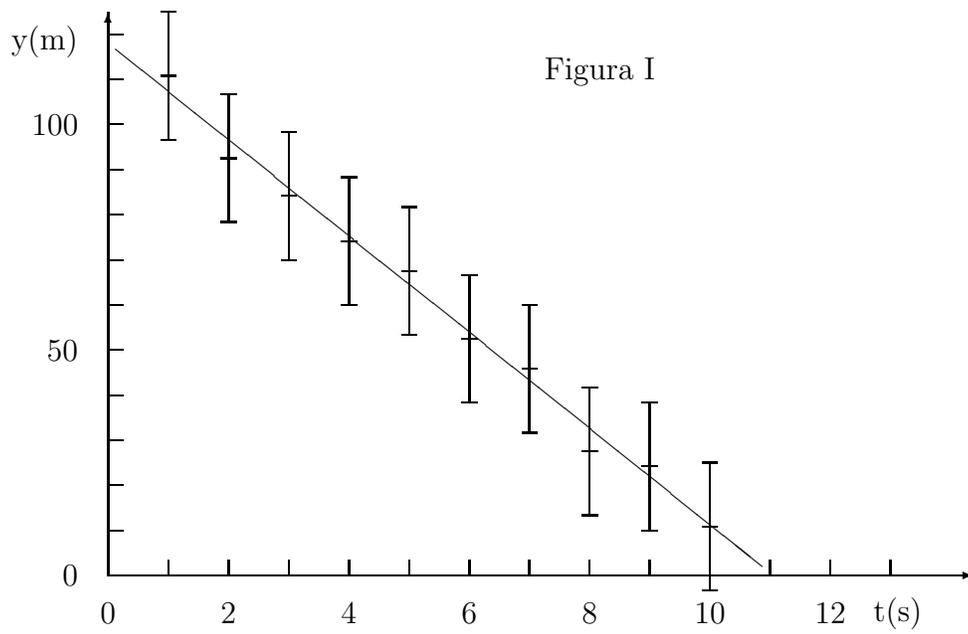
$$y = y_0 + v_0 t.$$

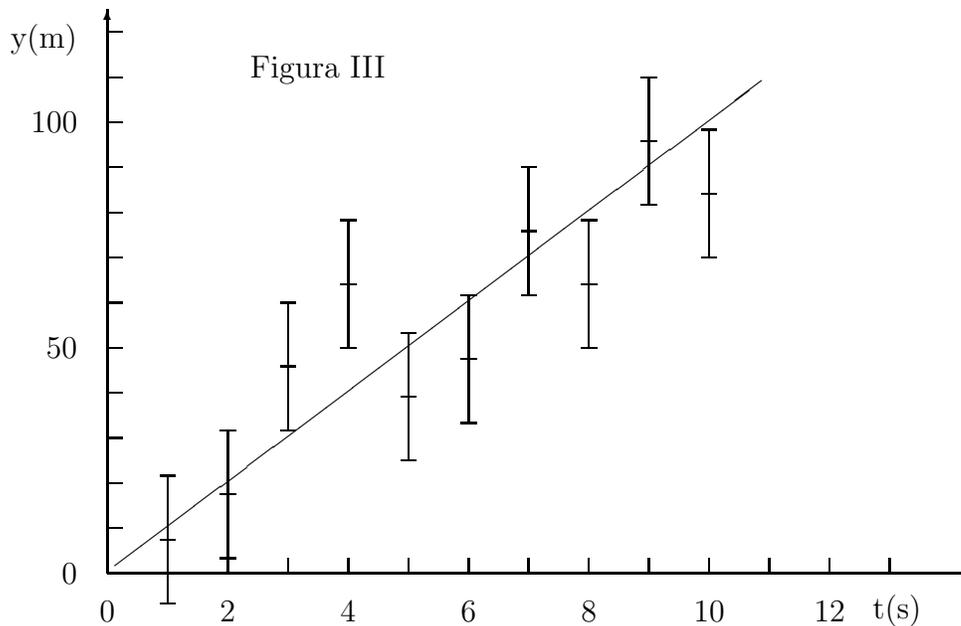
Em cada gráfico, está representada a reta

$$y = \hat{y} + \hat{v}t,$$

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

5. A relação entre  $y$  e  $x$  é da forma  $y = ax$ . Para estimar o valor de  $a$ , foram medidos valores de  $y$  para  $x = 1$ ,  $x = 2$  e  $x = 3$ , obtendo-se 2,3; 2,5 e 4,5, respectivamente, todos com mesmo desvio padrão 1,0 e covariâncias nulas.





- (a) Determine o valor de  $a$  pelo método dos mínimos quadrados. Não deixe de determinar o desvio padrão de  $a$ , também.
- (b) Realize um teste de qui-quadrado, utilizando o resultado obtido no ajuste efetuado no item acima.
- (c) Supondo que a relação entre  $y$  e  $x$  é  $y = 3x - 3$ , realize um novo teste de qui-quadrado. Compare as duas funções do ponto de vista de qualidade do ajuste.
6. Um experimentador ajustou uma reta  $y = a + b \cdot x$  aos pontos experimentais  $(x_i; y_i; \sigma_i)$  :

$$\{(1; 4, 12; 0, 15), (2; 5, 32; 0, 15), (3; 6, 31; 0, 15)\}.$$

- (a) Calcule as estimativas de  $a$  e de  $b$ , bem como seus respectivos desvios padrões.
- (b) Calcule  $\chi^2$  e a probabilidade desse  $\chi^2$  ser excedido. Este teste suporta a hipótese da dependência linear entre  $y$  e  $x$ ?

