Notas de apoio da disciplina de Probabilidades e Estatística

(Licenciaturas e Mestrados Integrados em Engenharia)

Manuel Cabral Morais

Lisboa, 13 de Fevereiro de 2012

Índice

Capítulo 1

Nota introdutória

1.1 Enquadramento da disciplina de Probabilidades e Estatística nas licenciaturas e mestrados integrados em Engenharia

A importância da disciplina de *Probabilidades e Estatística* na formação de alunas/os de Engenharia do IST é inegável. Basta pensar que, no seu plano curricular, elas/es estudarão fenómenos de natureza aleatória e terão que avaliar o desempenho de sistemas regidos por leis não determinísticas. Não surpreende pois que se trate de uma disciplina de nível básico e obrigatória nas licenciaturas em Engenharia do IST há várias décadas e que ela tenha *sobrevivido* ao *Processo de Bolonha*, constando da lista de disciplinas estruturantes de Matemática de todas as licenciaturas e mestrados integrados do IST.

Ao presumir que os programas das disciplinas de *Matemática* do Ensino Secundário são cumpridos, a disciplina de *Probabilidades e Estatística* proporciona aquele que é segundo contacto com esta área.

A disciplina de *Probabilidades e Estatística* tem carácter semestral e é introduzida no plano curricular no primeiro ou no segundo semestre do segundo ano lectivo. Esta introdução aparentemente tardia da disciplina no plano curricular das licenciaturas e mestrados integrados em Engenharia do IST encontra justificação no facto de o conteúdo programático da disciplina exigir que as/os alunas/os que a frequentam possuam alguma formação em Cálculo Diferencial e Integral, em particular, que estejam familiarizadas/os com:

- sucessões, funções reais de variável real, diferenciabilidade, primitivação, cálculo integral em R, fórmulas de integração por partes e por substituição, funções transcendentes elementares, séries numéricas;

- diferenciabilidade, derivadas parciais, estudo de extremos, integrais duplos.

Com efeito estabelece-se como desejável que as/os alunas/os tenham obtido aprovação às disciplinas de *Cálculo Diferencial e Integral I* e *II* de cujos programas se plasmaram os dois blocos de tópicos listados acima.[1]

Posto isto, o facto de a disciplina poder ser introduzida no primeiro semestre do segundo ano de algumas licenciaturas e de as/os alunas/os poderem ainda não ter obtido aprovação à disciplina de *Cálculo Diferencial e Integral II* requer alguns cuidados especiais na leccionação de alguns tópicos, nomeadamente pares aleatórios contínuos.

Importa referir que a disciplina de *Probabilidades e Estatística* é a primeira e última disciplina da área leccionada pela Secção de Probabilidades e Estatística em licenciaturas e mestrados integrados em Engenharia do IST, salvo raríssimas excepções.

Realce-se que a disciplina de *Probabilidades e Estatística* e os conceitos nela apreendidos abrem, no entanto, as portas a outras disciplinas que surgem posteriormente no plano curricular das licenciaturas e mestrados integrados do IST e que podem ter carácter complementar na área de Probabilidades e Estatística ou estarem directamente ligadas a aplicações específicas em Engenharia. A título meramente exemplificativo ocorre nomear

- uma disciplina cujo programa assenta em área de *Análise Multivariada*,

- outras duas mais especializadas e de cujos programas constam *cadeias de Markov* e *simulação estocástica e de Monte Carlo*, num dos casos, *processos estocásticos*, *probabilidades de erro* e *canais gaussianos*, noutro caso.

São elas as disciplinas de:

- *Análise de Dados e Avaliação* da área de especialização em *Transportes, Sistemas e Infra-Estruturas* do Mestrado Integrado em Engenharia Civil (5o. ano, 1o. semestre);

- *Modelação e Simulação* e *Fundamentos de Telecomunicações*, ambas do Mestrado Integrado em Eng. Electrotécnica e de Computadores (3o. ano, 2o. semestre).

[1]Estes tópicos são aqui mencionados pela ordem em surgem naqueles programas e não pela ordem em que são necessários na disciplina de *Probabilidades e Estatística*.

1.2 Objectivos operacionais

A disciplina de *Probabilidades e Estatística* tem por objectivo a *iniciação ao estudo da teoria das probabilidades e inferência estatística, tendo em vista a compreensão e aplicação dos seus principais conceitos e métodos.*[2]

Após a aprovação à disciplina as/os alunas/os devem ser capazes de:

- identificar eventos e calcular as respectivas probabilidades por recurso a resultados como as leis da probabilidade composta e da probabilidade total e o teorema de Bayes; averiguar a (in)dependência de eventos;

- destrinçar as variáveis aleatórias discretas das contínuas; identificar as diversas distribuições discretas e contínuas e as circunstâncias em que devem ser usadas; calcular probabilidades de eventos e momentos que lhes digam respeito; averiguar a (in)dependência de variáveis aleatórias e avaliar a associação entre elas;

- identificar as distribuições exactas ou aproximadas de combinações lineares de variáveis aleatórias, tirando partido das propriedades de fecho de algumas famílias de distribuições e do teorema do limite central;

- obter estimadores de parâmetros desconhecidos pelo método da máxima verosimilhança e avaliar as suas propriedades; obter uma variável fulcral para um parâmetro desconhecido, como o valor esperado, e a partir dela uma estatística de teste sobre esse mesmo parâmetro nos mais variados contextos distribucionais;

- construir um intervalo de confiança e efectuar testes de hipóteses sobre um parâmetro desconhecido em diversas situações distribuicionais; averiguar a adequação de uma distribuição ou de uma família de distribuições a um conjunto de dados; efectuar testes de hipóteses, recorrendo ao procedimento geral ou, em alternativa, por recurso ao *p-value*;

- estimar os diversos parâmetros desconhecidos do modelo de regressão linear simples; obter intervalos de confiança e efectuar testes de hipóteses sobre tais parâmetros; avaliar a qualidade do ajustamento da recta de regressão ao conjunto de dados.

De modo a atingir plenamente estes objectivos operacionais parece-nos essencial que a estrutura de apresentação dos capítulos destas notas de apoio respeite a filosofia *aprender*

[2] Ver, por exemplo, o *link* da disciplina de *Probabilidades e Estatística* no plano curricular do Mestrado integrado em Eng. Electrotécnica e de Computadores.

por exemplos e fazendo. Assim, a matéria é motivada, os resultados são enunciados, raramente demonstrados mas sempre ilustrados com exemplos ou exercícios trabalhados em conjunto com as/os alunas/os, como os que se seguem.

Exemplo 1.1 — Eventos e probabilidades

Um sistema de comunicação binária transmite "zeros" e "uns" com probabilidade 0.5 em qualquer dos casos. Devido ao ruído existente no canal de comunicação há erros na recepção: transmitido um "um" ele pode ser recebido como um "zero" com probabilidade 0.1, ao passo que um "zero" pode ser recebido como um "um" com probabilidade 0.05.

Determine a probabilidade de se receber um "zero".

- **Quadro de eventos e probabilidades**

Evento	Probabilidade
TZ = transmitir um "zero"	$P(TZ) = 0.5$
$\overline{TZ}$ = transmitir um "um"	$P(\overline{TZ}) = 0.5$
RZ = receber um "zero"	$P(RZ) =?$
$RZ\vert\overline{TZ}$ = receber um "zero" dado que foi transmitido um "um"	$P(RZ\vert\overline{TZ}) = 0.1$
$\overline{RZ}\vert TZ$ = receber um "um" dado que foi transmitido um "zero"	$P(\overline{RZ}\vert TZ) = 0.05$

- **Prob. pedida**

  Aplicando o teorema da probabilidade total, tem-se

$$\begin{aligned} P(RZ) &= P(RZ|TZ) \times P(TZ) + P(RZ|\overline{TZ}) \times P(\overline{TZ}) \\ &= [1 - P(\overline{RZ}|TZ)] \times P(TZ) + P(RZ|\overline{TZ})P(\overline{TZ}) \\ &= (1 - 0.05) \times 0.5 + 0.1 \times 0.5 \\ &= 0.525. \end{aligned}$$

•

Exemplo 1.2 — Duas variáveis aleatórias discretas

Um computador possui um número elevado de componentes de um mesmo tipo que falham de modo independente. O número de componentes desse tipo que falham por mês é uma variável aleatória com distribuição de Poisson com variância igual a um. Admita que o computador só falha se pelo menos doze dessas componentes falharem.

Calcule a probabilidade de o computador não ter falhado ao fim de um ano.

- **Variável aleatória**

  $X_1 =$ número de componentes que falham em um mês

- **Distribuição de X_1**

  $X_1 \sim \text{Poisson}(\lambda)$

- **Parâmetro**

$$\begin{aligned} \lambda \quad : \quad & V(X) = 1 \\ & \lambda = 1 \end{aligned}$$

- **Nova variável aleatória**

  $X_{12} =$ número de componentes que falham num ano (12 meses)

- **Distribuição de X_{12}**

  Tirando partido do facto de as componentes falharem de modo independente e recorrendo à propriedade reprodutiva da distribuição de Poisson, pode concluir-se que:

  $X_{12} \sim \text{Poisson}(\lambda_{12} = 12 \times \lambda = 12)$

- **Função de probabilidade de X_{12}**

$$P(X_{12} = x) = \frac{e^{-12} 12^x}{x!}, \; x = 0, 1, 2, \ldots$$

- **Probabilidade pedida**

$$\begin{aligned} P(\text{comp. não falhar num ano}) &= P(X_{12} \leq 11) \\ &= F_{Poisson(12)}(11) \\ &\stackrel{tabela}{\simeq} 0.4616. \end{aligned}$$

Exemplo 1.3 — Duas variáveis aleatórias contínuas
A resistência eléctrica[3] (X) de um objecto e a sua condutância eléctrica[4] (Y) estão relacionadas do seguinte modo: $Y = X^{-1}$.

Assuma que

$$P(X \leq x) = \begin{cases} 0, & x < 900 \\ \frac{x-900}{200}, & 900 \leq x \leq 1100 \\ 1, & x > 1100 \end{cases}$$

e determine a probabilidade de a condutância eléctrica exceder $10^{-3} mho$.

- **Variável aleatória**

  $X =$ resistência eléctrica

- **Nova variável aleatória**

  $Y = X^{-1} =$ condutância eléctrica

- **Probabilidade pedida**

$$\begin{aligned} P(Y > 10^{-3}\, mho) &= P\left(\frac{1}{X} > 10^{-3}\right) \\ &= P\left(X < 10^3\right) \\ &= \frac{1000 - 900}{200} \\ &= \frac{1}{2}. \end{aligned}$$

•

Importa fazer um reparo sobre a resolução dos exemplos/exercícios das notas de apoio. Ela é apresentada em pequenas secções com cabeçalho logo tem um carácter aparentemente repetitivo que se tem revelado, por sinal, útil para que as/os alunas/os aprendam a estruturar devidamente a resolução de qualquer exercício da disciplina de *Probabilidades e Estatística*.

[3] A resistência eléctrica é a capacidade de um corpo qualquer se opor à passagem de corrente elétrica pelo mesmo; de acordo com o Sistema Internacional de Unidades (SI), a resistência eléctrica é medida em *ohm* (http://pt.wikipedia.org/wiki/Resistência_elétrica).

[4] A condutância eléctrica mede a facilidade com que a corrente eléctrica flui através de uma componente eléctrica, logo trata-se do recíproco da resistência eléctrica; de acordo com o SI, a condutância eléctrica é medida em *siemens* ou *mho* (http://pt.wikipedia.org/wiki/Condutância_elétrica).

Estas notas de apoio constituem também um manual para a/o aluna/o desejosa/o de um rápido progresso na aprendizagem de *Probabilidades e Estatística* e disposta/o a estudar sozinha/o e capaz de combinar o material que aqui encontra com outro proveniente de fontes tão importantes como o livro de texto recomendado para a disciplina.

Expresso desde já os meus agradecimentos às/aos minhas/meus várias/os colegas da Secção de Probabilidades e Estatística do Departamento de Matemática do Instituto Superior Técnico que leccionaram a disciplina de *Probabilidades e Estatística* (das licenciaturas e mestrados integrados em Engenharia) e da então disciplina de *Estatística* (da licenciatura em *Matemática Aplicada e Computação*), por terem directa ou indirectamente contribuído para estas notas de apoio.

Os erros e imprecisões eventualmente existentes nestas notas são, naturalmente, aleatórios — muitas das vezes fruto de operações de *copy/paste* — e da inteira responsabilidade do autor, que muito agradece que eles lhe sejam comunicados por e-mail para *maj@math.ist.utl.pt*.

Boa leitura e bom trabalho.

Manuel Cabral Morais
Lisboa, 8 de Setembro de 2011

Capítulo 2

Noções básicas de probabilidade

Palavras como

- provável (provavelmente)
- probabilidade
- acaso
- sorte

pertencem ao vocabulário corrente e são utilizadas com extrema frequência por todos, em parte por termos a convicção de que a natureza é mutável e incerta, de que o futuro encerra em si inúmeras possibilidades e de que o acaso governa o mundo.

Na formalização matemática actual, a probabilidade é um termo medindo o grau de possibilidade ou de credibilidade de ocorrência de um acontecimento.

2.1 Experiências aleatórias. Espaço de resultados. Acontecimentos.

A formalização moderna de Probabilidade assenta nas nocões de

- experiência aleatória e seus possíveis resultados e de
- acontecimento.

Definição 2.1 — Experiência aleatória (E.A.)
Experiência cujo resultado exacto não pode ser predito antes da realização da mesma devido à intervenção do acaso. •

Nota 2.2 — Experiência aleatória
No caso de a experiência aleatória poder ser repetida um grande número de vezes, em condições mais ou menos semelhantes, os resultados globais apresentam certa "regularidade estatística"... •

Exemplo 2.3 — Experiências aleatórias

Designação	Experiência aleatória
E_1	Registo do número de viaturas que atingem os $100Km/h$ em menos de 6 segundos, em 7 viaturas testadas
E_2	Contagem do número anual de acidentes de automóvel na A1
E_3	Medição da resistência de uma mola da suspensão de uma viatura

•

Definição 2.4 — Espaço de resultados
Conjunto de todos os resultados possíveis de uma E.A. É conhecido antes de a E.A. se realizar e é usualmente representado pela letra grega Ω. •

Nota 2.5 — Espaço de resultados
Ω diz-se:

- discreto — caso $\#\Omega$ seja finito ou infinito numerável;
- contínuo — se $\#\Omega$ for infinito não numerável. •

Exemplo 2.6 — Espaços de resultados
Na tabela seguinte figuram os espaços de resultados das três e.a. apresentadas no Exemplo 2.3:

E.A.	**Espaço de resultados** (Ω)	**Classificação de** Ω
E_1	$\{0,1,2,3,4,5,6,7\}$	Discreto (finito)
E_2	$\{0,1,2,\ldots\}$	Discreto (infinito numerável)
E_3	$I\!R^+$	Contínuo (infinito não numerável)

•

Definição 2.7 — Evento (acontecimento)
Designação dada a qualquer subconjunto do espaço de resultados. •

Nota 2.8 — Evento
Em relação a uma dada E.A. diz-se que o evento A ocorreu sse o resultado da E.A. pertencer a A. •

Exemplo 2.9 — Eventos
De seguida apresentam-se alguns eventos associados às três e.a. descritas no Exemplo 2.3:

E.A.	**Evento**
E_1	$A =$ nenhuma das 7 viaturas testadas atingiu os $100Km/h$ em menos de 6 segundos $= \{0\}$ $B =$ pelo menos 4 das 7 viaturas testadas atingiu os $100Km/h$ em menos de 6 segundos $= \{4,5,6,7\}$
E_2	$C =$ registo de mais de 5 acidentes anuais na A1 $= \{6,7,\ldots\}$
E_3	$D =$ resistência superior a 8 unidades $= (8,+\infty)$

•

Nota 2.10 — Classificação de eventos
O evento A diz-se:

- elementar — quando constituído por um único elemento de Ω, i.e., $\#A = 1$;
- certo — se $A = \Omega$;
- impossível — caso $A = \emptyset$. •

Definição 2.11 — Eventos disjuntos
Os eventos A e B dizem-se disjuntos (ou mutuamente exclusivos, ou incompatíveis) sse

$$A \cap B = \emptyset, \tag{2.1}$$

i.e., se a realização simultânea de A e B for impossível. •

Definição 2.12 — Inclusão de eventos
Quando o evento A está contido (incluso) em B — $A \subset B$ — verifica-se:

$$\text{Realização de } A \;\Rightarrow\; \text{Realização de } B \tag{2.2}$$

$$\text{Realização de } A \;\not\Leftarrow\; \text{Realização de } B, \tag{2.3}$$

i.e., a realização de A implica a de B mas a implicação no sentido contrário não é necessariamente verdadeira. •

Uma vez que os eventos não passam de (sub)conjuntos é possível efectuar **operações sobre eventos** já nossas conhecidas como são o caso da intersecção, da reunião, etc. Descreveremos o seu significado em termos de realizações de eventos quer verbalmente, quer à custa de um diagrama de Venn.

Sejam

- Ω o espaço de resultados de uma E.A. e
- A e B dois eventos.

Então podemos efectuar as seguintes operações sobre A e B:

Operação	Notação	Descrição verbal	Diagrama de Venn
Intersecção	$A \cap B$	Realização simultânea de A e de B	
Reunião	$A \cup B$	Realização de A ou de B, i.e., de pelo menos um dos dois eventos	
Diferença	$B \backslash A$	Realização de B sem que se realize A (B excepto A)	
	$A \backslash B$	Realização de A sem que se realize B (A excepto B)	
Complementar	$\overline{A}$	Não realização de A	

As operações sobre eventos gozam de **propriedades** bem conhecidas como a associatividade, comutatividade, etc., que convém recordar:

Propriedade	Descrição matemática
Associatividade	$(A \cap B) \cap C = A \cap (B \cap C)$
	$(A \cup B) \cup C = A \cup (B \cup C)$
Comutatividade	$A \cap B = B \cap A$
	$A \cup B = B \cup A$
Distributividade	$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
Idempotência	$A \cap A = A$
	$A \cup A = A$
Absorção	$A \subset B \Rightarrow A \cap B = A$
	$A \subset B \Rightarrow A \cup B = B$
Modulares	$A \cap \Omega = A$
	$A \cup \Omega = \Omega$
	$A \cap \emptyset = \emptyset$
	$A \cup \emptyset = A$
Leis de De Morgan	$\overline{A \cap B} = \overline{A} \cup \overline{B}$
	$\overline{A \cup B} = \overline{A} \cap \overline{B}$
Dupla negação	$\overline{\overline{A}} = A$

2.2 Noção de probabilidade. Interpretações de Laplace, frequencista e subjectivista. Axiomas e teoremas decorrentes.

A probabilidade é um conceito extraordinariamente complexo e, como teremos ocasião de ver daqui a pouco, somos capazes de adiantar algumas noções de probabilidade que se revelarão insatisfatórias devido a limitações a elas subjacentes.

Definição 2.13 — Probabilidade clássica de Laplace
Considere-se uma E.A. com espaço de resultados Ω com as seguintes particularidades: Ω é constituído por

- n eventos elementares ($\#\Omega = n$)
- distintos
- igualmente prováveis[1] e em
- número finito.

Considere-se ainda que a realização do evento A passa pela ocorrência de m dos n eventos elementares, i.e., $\#A = m$. Então a probabilidade de realização de A é dada por:

$$\begin{aligned} P(A) &= \frac{\text{número de casos favoráveis à ocorrência de A}}{\text{número de casos possíveis}} \\ &= \frac{\#A}{\#\Omega} \\ &= \frac{m}{n}. \end{aligned} \tag{2.4}$$

•

Nota 2.14 — Limitações da probabilidade clássica de Laplace
Esta definição só é válida quando

- $\#\Omega < +\infty$ (ou seja, o número de eventos elementares é finito) e
- Ω é constituído por eventos elementares igualmente prováveis,

pressupostos estes frequentemente violados na prática. •

[1] Nada leva a crer que a ocorrência de algum dos eventos é privilegiada em relação à dos restantes.

Exemplo 2.15 — Probabilidade clássica de Laplace
Admita que num stand se encontram 353 viaturas de somente duas marcas (A e B). Destas:

- 201 são da marca A;
- 57 possuem direcção assistida;
- 37 são da marca A e possuem direcção assistida.

Calcule a probabilidade de uma viatura seleccionada ao acaso ser da marca A.

- **Evento**

  $A =$ viatura seleccionada ao acaso ser da marca A

- **No. casos favoráveis**

  $m = 201$

- **No. casos possíveis**

  $n = 353$

- **Probabilidade pedida**

$$P(A) = \frac{m}{n} = \frac{201}{353}.$$

•

Antes de passarmos a uma outra noção de probabilidade é conveniente adiantarmos a definição de frequência relativa de um evento bem como as propriedades algébricas dessa mesma frequência.

Definição 2.16 — Frequência relativa
Sejam:

- N o número de realizações (nas mesmas condições) de certa E.A.;
- $n_N(A)$ o número de vezes que o evento A ocorreu nas N realizações da E.A. (i.e., representa a frequência absoluta do evento A).

Então a frequência relativa do evento A é dada por

$$f_N(A) = \frac{n_N(A)}{N}. \tag{2.5}$$

•

Nota 2.17 — Propriedades algébricas
A frequência relativa satisfaz as seguintes propriedades:

- $0 \leq f_N(A) \leq 1$;
- $f_N(\Omega) = 1$;
- $f_N(A \cup B) = f_N(A) + f_N(B)$, se $A \cap B = \emptyset$;
- $f_N(A)$ estabiliza à medida que N aumenta. •

Não surpreende pois a seguinte noção de probabilidade.

Definição 2.18 — Probabilidade frequencista
A probabilidade do evento A é igual ao limite da frequência relativa da ocorrência do evento A:

$$P(A) = \lim_{N \to +\infty} \frac{n_N(A)}{N} = \lim_{N \to +\infty} f_N(A). \tag{2.6}$$

•

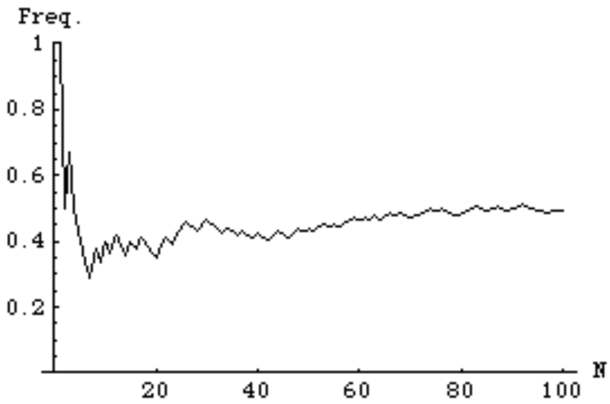
Exemplo 2.19 — Probabilidade frequencista
Foram registados os resultados respeitantes a um total de 100 lançamentos de uma moeda equilibrada. Assim, nas colunas da tabela abaixo podem encontrar-se

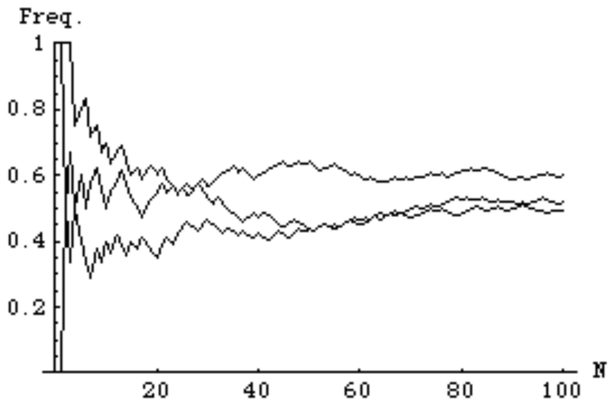
- o número do lançamento (N),
- o resultado do $N-$ésimo lançamento ($0 = \text{coroa}, 1 = \text{cara}$) e
- a frequência relativa do evento $A =$ sair cara até ao $N-$ésimo lançamento ($f_N(A)$),

respectivamente.

N	(0=coroa, 1=cara)	$f_N(A)$	$\cdots$	N	(0=coroa, 1=cara)	$f_N(A)$
1	1	$\frac{1}{1}$	$\cdots$	91	1	$\frac{46}{91}$
2	0	$\frac{1}{2}$	$\cdots$	92	1	$\frac{47}{92}$
3	1	$\frac{2}{3}$	$\cdots$	93	0	$\frac{47}{93}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
10	1	$\frac{2}{5}$	$\cdots$	100	0	$\frac{49}{100}$

A esta tabela segue-se o correspondente gráfico da frequência relativa $f_N(A)$ (à esquerda) e o deste e de outros dois conjuntos de 100 lançamentos (à direita). Em ambos é evidente a estabilização da frequência relativa em torno de 0.5 à medida que o número total de lançamentos (N) aumenta.





•

Nota 2.20 — Limitações da probabilidade frequencista
Esta noção de probabilidade não é razoável caso a E.A. não possa ser realizada mais do que uma vez ou quando ela é hipotética (por exemplo, uma ida a Marte). •

Definição 2.21 — Probabilidade subjectiva
Uma pessoa pode atribuir a um evento um número real no intervalo $[0, 1]$ a que se dará o nome de *probabilidade do evento* e que expressa um grau de credibilidade pessoal na ocorrência do evento. •

Exemplo 2.22 — Probabilidade subjectiva
Ao perguntar-se qual a probabilidade de visitar-se o planeta Marte antes do ano 2030 obteve-se as seguintes respostas de duas pessoas:

- funcionário da NASA $\rightarrow$ 0.5;
- taxista $\rightarrow$ 0. •

Este exemplo leva a considerar uma outra noção de probabilidade que deverá ser precedida pela apresentação da noção de $\sigma-$álgebra de eventos.

Definição 2.23 — $\sigma-$**álgebra de eventos**
Trata-se de uma colecção não vazia de eventos (probabilizáveis), $\mathcal{A}$, que verifica as seguintes propriedades:

1. $\Omega \in \mathcal{A}$;
2. $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$;
3. $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$ para qualquer colecção contável de eventos de $\mathcal{A}$, seja ela $\{A_1, A_2, \ldots\}$. •

Exemplo 2.24 — $\sigma-$**álgebra de eventos**

- $\mathcal{A}_1 = \{\emptyset, \Omega\}$;
- $\mathcal{A}_2 = \mathbb{P}(\Omega)$ que representa as "partes de Ω", i.e., a colecção de todos os subconjuntos de Ω. •

Definição 2.25 — **Função de probabilidade** (no sentido de Kolmogorov)
Função que possui por domínio a $\sigma-$álgebra de eventos e por contradomínio o intervalo $[0, 1]$ — i.e.,

$$P : \mathcal{A} \to [0, 1] \tag{2.7}$$

— e com a particularidade de respeitar os seguintes...

Axiomas

1. $P(\Omega) = 1$.
2. $0 \leq P(A) \leq 1, \forall A \in \mathcal{A}$.
3. Seja $\{A_1, A_2, \ldots\}$ uma colecção contável de eventos mutuamente exclusivos de $\mathcal{A}$ (i.e. $A_i \cap A_j = \emptyset, i \neq j$). Então
$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i). \tag{2.8}$$

•

Proposição 2.26 — **Consequências elementares dos axiomas**
Os axiomas não nos ensinam a calcular probabilidades mas estabelecem regras para o seu cálculo — muito em particular algumas das suas seguintes consequências elementares:

1. $P(\emptyset) = 0$;
2. $P(\overline{A}) = 1 - P(A)$;
3. $A \subset B \Rightarrow P(A) \leq P(B)$;
4. $P(B \backslash A) = P(B) - P(A \cap B)$. •

Exercício 2.27 — Demonstre as consequências elementares dos axiomas. •

Exemplo 2.28 — **Consequências elementares dos axiomas**
Uma companhia de telecomunicações classifica as chamadas como sendo do tipo:

- V, caso haja transmissão de voz;
- D, caso haja transmissão de dados (por modem ou fax).

Com base em informação da companhia pode adiantar-se que:

Evento	**Probabilidade**
$V =$ transmissão de voz	$P(V) = 0.7$
$D =$ transmissão de dados	$P(D) = 0.5$
$V \cap D =$ transmissão simultânea de voz e dados	$P(V \cap D) = 0.2$

(a) Calcule a probabilidade de a transmissão não ser de voz.

- **Evento**
  $\overline{V} =$ transmissão não ser de voz
- **Probabilidade pedida**

$$\begin{aligned} P(\overline{V}) &= 1 - P(V) \\ &= 1 - 0.7 \\ &= 0.3. \end{aligned}$$

(b) Obtenha a probabilidade de haver exclusivamente transmissão de voz.

- **Evento**

  $V \backslash D$ = transmissão exclusiva de voz

- **Probabilidade pedida**

$$
\begin{aligned}
P(V \backslash D) &= P(V) - P(V \cap D) \\
&= 0.7 - 0.2 \\
&= 0.5.
\end{aligned}
$$

•

Nota 2.29 — Um evento pode ainda ser classificado de:

- quase-certo — se $P(A) = 1$ no entanto $A \neq \Omega$;
- quase-impossível — caso $P(A) = 0$ mas $A \neq \emptyset$. •

Exercício 2.30 — Dê exemplos de eventos quase-certos e quase-impossíveis. •

De seguida são enunciados dois resultados que permitirão o cálculo da probabilidade da reunião de dois e de três eventos.

Proposição 2.31 — **Reunião de dois eventos**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{2.9}$$

Demonstração:

$$
\begin{aligned}
P(A \cup B) &= P[(A \backslash B) \cup (A \cap B) \cup (B \backslash A)] \\
&= P(A \backslash B) + P(A \cap B) + P(B \backslash A) \\
&= [P(A) - P(A \cap B)] + P(A \cap B) + [P(B) - P(A \cap B)] \\
&= P(A) + P(B) - P(A \cap B). \qquad (2.10)
\end{aligned}
$$

•

Proposição 2.32 — **Reunião de três eventos**

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
&\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&\quad +P(A \cap B \cap C). \qquad (2.11)
\end{aligned}
$$

•

Exercício 2.33 — Demonstre a regra de adição (2.11). •

Exemplo 2.34 — Regras de adição

Retome o Exemplo 2.15 respeitante ao stand com 353 viaturas.

(a) Organize um quadro com os eventos já descritos e com as respectivas probabilidades.

- **Quadro de eventos e probabilidades**

Evento	Casos favor.	Casos poss.	Probabilidade
$A =$ viat. marca A	201	353	$P(A) = \frac{201}{353}$
$D =$ viat. com dir. assist.	57	353	$P(D) = \frac{57}{353}$
$A \cap D =$ viat. marca A com dir. assist.	37	353	$P(A \cap D) = \frac{37}{353}$

(b) Obtenha a probabilidade de uma viatura seleccionada ao acaso ser da marca A ou possuir direcção assistida.

- **Evento**

  $A \cup D$

- **Probabilidade pedida**

$$
\begin{aligned}
P(A \cup D) &= P(A) + P(D) - P(A \cap D) \\
&= \frac{201}{353} + \frac{57}{353} - \frac{37}{353} \\
&= \frac{221}{353}.
\end{aligned}
$$

•

2.3 Probabilidade condicionada.

Motivação 2.35 — **Probabilidade condicionada**
Suponha que dispomos de um baralho de 52 cartas (13 de cada naipe) do qual extraímos uma carta ao acaso.

(a) Qual a probabilidade de ter saído o rei de copas? **1/52**

(b) Qual a probabilidade de ter saído o rei de copas sabendo à partida que a carta extraída é uma carta de paus? **0**

(c) E se soubéssemos de antemão que a carta é de copas? **1/13** •

Nota 2.36 — **Probabilidade condicionada**
Como pudemos ver, a probabilidade do evento *sair o rei de copas* foi sucessivamente avaliada à medida que nos foi adiantada informação. •

A questão que se coloca naturalmente é: de que forma a obtenção de informação adicional (correspondente à ocorrência de eventos) pode vir a influenciar a cálculo de probabilidades?

Definição 2.37 — **Probabilidade condicionada**
Sejam:

- Ω o espaço de resultados;
- P a função de probabilidade.

Então a probabilidade do evento A condicionada pela ocorrência do evento B é dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{2.12}$$

desde que $P(B) > 0$. •

Nota 2.38 — **Probabilidade condicionada**
Esta probabilidade também pode ler-se do seguinte modo *probabilidade de A dado B* ou *probabilidade de A sabendo que B ocorreu* e representa uma reavaliação da probabilidade de A face ao facto de B ter ocorrido. •

Exemplo 2.39 — Probabilidade condicionada (cont.)
Qual a probabilidade de a carta seleccionada ao acaso ser o rei de copas sabendo à partida que se trata de uma carta de copas?

- **Quadro de eventos e probabilidades**

Evento	**Probabilidade**
$A =$ Sair o rei de copas	$P(A) = \frac{1}{52}$
$B =$ Sair uma carta de copas	$P(B) = \frac{13}{52}$

- **Evento**

  $A|B$

- **Prob. pedida**

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)}{P(B)} \\ &= \frac{1/52}{13/52} \\ &= \frac{1}{13}. \end{aligned} \tag{2.13}$$

•

Nota 2.40 — Probabilidade condicionada
$P(\ldots|B)$ é uma função de probabilidade (no sentido de Kolmogorov), como tal respeita os três axiomas seguintes:

1. $P(\Omega|B) = 1$.
2. $0 \leq P(A|B) \leq 1$, $\forall A \in \mathcal{A}$.
3. Seja $\{A_1, A_2, \ldots\}$ uma colecção contável de eventos mutuamente exclusivos de $\mathcal{A}$ (i.e. $A_i \cap A_j = \emptyset, i \neq j$). Então

$$P\left[\left(\bigcup_{i=1}^{+\infty} A_i\right)\middle| B\right] = \sum_{i=1}^{+\infty} P(A_i|B). \tag{2.14}$$

Para além disso, verifica as consequências elementares destes mesmos axiomas enunciadas na Proposição 2.26. •

Exemplo 2.41 — Um grupo de alunos do 1o. ano de Mecânica elaborou 100 programas. Constatou-se que

- 20% possuem erros de Sintaxe (S),
- 30% possuem erros de Acesso à Memória (AM) e
- 10% possuem erros de Sintaxe e de Acesso à Memória.

Admitamos que foi escolhido ao acaso um programa e que este possuía erros de Sintaxe. Neste caso qual a probabilidade do programa possuir (também) erros de Acesso à Memória?

- **Quadro de eventos e probabilidades**

Evento	**Probabilidade**
$S =$ programa com erros de Sintaxe	$P(S) = 0.2$
$AM =$ programa com erros de Acesso à Memória	$P(AM) = 0.3$
$S \cap AM =$ programa com erros de Sintaxe e de Acesso à Memória	$P(S \cap AM) = 0.1$

- **Evento**

  $AM|S$

- **Prob. pedida**

$$\begin{aligned} P(AM|S) &= \frac{P(S \cap AM)}{P(S)} \\ &= \frac{0.1}{0.2} \\ &= \frac{1}{2} \end{aligned} \tag{2.15}$$

•

2.4 Leis das probabilidades compostas e da probabilidade total. Teorema de Bayes.

Motivação 2.42 — **Lei das probabilidades compostas**
Suponha que se conhece $P(A|B)$ e $P(B)$. Será que podemos calcular $P(A \cap B)$? A resposta é afirmativa:

$$P(A \cap B) = P(B)P(A|B). \tag{2.16}$$

A generalização deste resultado para a intersecção de n eventos constitui a lei das probabilidades compostas (uma de duas regras da multiplicação). •

Proposição 2.43 — **Lei das probabilidades compostas**
Considere-se uma colecção de n eventos $\{A_i\}_{i=1,\ldots,n}$ tal que $P(A_i) > 0$ e $P(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$. Então

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) &= P(A_1) \times P(A_2|A_1) \times P[A_3|(A_1 \cap A_2)] \\ &\quad \ldots \times P[A_n|(A_1 \cap A_2 \cap \ldots \cap A_{n-1})]. \end{aligned} \tag{2.17}$$

•

Exercício 2.44 — Demonstre a lei das prob. compostas enunciada na Prop. 2.43. •

Com o exemplo que se segue, veremos que a lei das probabilidades compostas é de extrema utilidade especialmente quando pretendemos calcular a probabilidade de sequências de eventos em experiências aleatórias.

Exemplo 2.45 — **Lei das probabilidades compostas**
Considere-se um lote de 100 molas do sistema de suspensão de automóvel. Destas, 20 são consideradas defeituosas (D) por violarem a lei de Hooke quando se aplica uma força superior a $35 \times 10^4 N$.
Responda às questões seguintes sabendo que foram recolhidas 3 molas ao acaso e sem reposição deste lote.

(a) Qual a probabilidade das 3 molas não serem defeituosas?

- **Evento**
  $\overline{D}_1 \cap \overline{D}_2 \cap \overline{D}_3$ = 1a., 2a. e 3a. molas não defeituosas

- **Prob. pedida**

$$\begin{aligned}P(\overline{D}_1 \cap \overline{D}_2 \cap \overline{D}_3) &\stackrel{\textit{lei prob. comp.}}{=} P(\overline{D}_1) \times P(\overline{D}_2|\overline{D}_1) \times P[\overline{D}_3|(\overline{D}_1 \cap \overline{D}_2)] \\ &= \frac{80}{100} \times \frac{80-1}{100-1} \times \frac{80-1-1}{100-1-1} \\ &= \frac{80}{100} \times \frac{79}{99} \times \frac{78}{98}.\end{aligned}$$

(b) Qual a probabilidade de, nessa mesma recolha, obtermos uma mola defeituosa somente à 3a. extracção?

- **Evento**

  $\overline{D}_1 \cap \overline{D}_2 \cap D_3 =$ 1a. e 2a. molas não defeituosas e 3a. mola defeituosa

- **Prob. pedida**

$$\begin{aligned}P(\overline{D}_1 \cap \overline{D}_2 \cap D_3) &\stackrel{\textit{lei prob. comp.}}{=} P(\overline{D}_1) \times P(\overline{D}_2|\overline{D}_1) \times P[D_3|(\overline{D}_1 \cap \overline{D}_2)] \\ &= \frac{80}{100} \times \frac{80-1}{100-1} \times \frac{20}{100-1-1} \\ &= \frac{80}{100} \times \frac{79}{99} \times \frac{20}{98}.\end{aligned}$$

•

A noção de partição do espaço de resultados enunciada já a seguir é necessária para enunciar não só a lei da probabilidade total nesta secção como o Teorema de Bayes enunciado ainda neste capítulo.

Definição 2.46 — Partição de Ω
A colecção de n eventos $\mathcal{P}_\Omega = \{A_i\}_{i=1,\ldots,n}$ diz-se uma partição de Ω sse:

- $A_i \cap A_j = \emptyset, i \neq j$ (eventos disjuntos dois a dois);
- $\bigcup_{i=1}^n A_i = \Omega$;
- $P(A_i) > 0, i = 1, \ldots, n.$[2] •

[2] As partições com que lidaremos são de um modo geral constituídas por um número finito de eventos. Estes também podem ser em número infinito numerável. No âmbito desta disciplina não se considerarão partições com um número infinito não numerável de eventos, daí a notação.

Exemplo 2.47 — Partição de Ω

- E.A. — Registo do número de acidentes na A1 durante um ano
- $\Omega = \{0, 1, 2, \ldots\}$
- $\mathcal{P}_\Omega = \{\{i\}\}_{i=0,1,2,\ldots}$, partição constituída por todos os eventos elementares de Ω
- $\mathcal{P}'_\Omega = \{\{0, 2, 4, 6, \ldots\}, \{1, 3, 5, \ldots\}\}$, partição constituída pelos eventos "registo de número par" e "registo de número ímpar". •

Proposição 2.48 — Lei da probabilidade total
Sejam:

- B um evento;
- $\mathcal{P}_\Omega = \{A_i\}_{i=1,\ldots,n}$ uma partição de Ω.

Então

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \tag{2.18}$$

•

Nota 2.49 — Lei da probabilidade total
Este resultado reveste-se de grande importância por permitir calcular a probabilidade de um evento B quando se dispõe das probabilidades do evento B condicionado a eventos A_i (que constituem uma partição de Ω) e das probabilidades destes eventos que condicionam B. •

Exercício 2.50 — Lei da probabilidade total
Prove a lei da probabilidade total enunciada anteriormente na Proposição 2.48. •

Exemplo 2.51 — Lei da probabilidade total [3]
Com o objectivo de aumentar a segurança de crianças em automóveis, estão a ser testados dois dispositivos de retenção A e B. As simulações mostraram que, em caso de acidente grave, o dispositivo A (resp. B) é eficaz em 95% (resp. 96%) dos casos. Admita que no mercado só passarão a existir estes dois dispositivos, instalados em automóveis exactamente na mesma proporção.
Qual a probabilidade do dispositivo de retenção instalado num automóvel seleccionado ao acaso vir a ser eficaz em caso de acidente grave?

[3] Adaptado do Exame de 1a. Época, 24 de Junho de 2006.

- **Quadro de eventos e probabilidades**

Evento	Probabilidade
A = dispositivo do tipo A	$P(A) = 0.5$
B = dispositivo do tipo B	$P(B) = 0.5$
E = dispositivo eficaz em caso de acidente grave	$P(E) = ?$
$E\|A$ = dispositivo eficaz... dado que é do tipo A	$P(E\|A) = 0.95$
$E\|B$ = dispositivo eficaz... dado que é do tipo B	$P(E\|B) = 0.96$

- **Evento**

  E = dispositivo eficaz em caso de acidente grave

- **Prob. pedida**

$$\begin{aligned} P(E) &\overset{\textit{lei prob. total}}{=} P(E|A) \times P(A) + P(E|B) \times P(B) \\ &= 0.95 \times 0.5 + 0.96 \times 0.5 \\ &= 0.955. \end{aligned}$$

  (Por que razão $P(E)$ é igual à média aritmética de $P(E|A)$ e $P(E|B)$?) •

Motivação 2.52 — **Teorema de Bayes**
Uma vez conhecida a probabilidade $P(B|A)$ poderá avaliar-se $P(A|B)$? A resposta a esta questão é afirmativa —

$$P(A|B) = P(B|A) \times \frac{P(A)}{P(B)} \tag{2.19}$$

— e é enunciada no teorema que se segue. •

Teorema 2.53 — **Teorema de Bayes**
Sejam:

- B um evento tal que $P(B) > 0$;

- $\mathcal{P}_\Omega = \{A_1, \ldots, A_n\}$ uma partição de Ω.

Então

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}. \tag{2.20}$$

Recorrendo à lei da probabilidade total pode adiantar-se que

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}. \tag{2.21}$$

•

Nota 2.54 — Este resultado permite que a reavaliação das probabilidades se faça tanto num sentido como noutro:

- A_i sabendo B e
- B sabendo A_i. •

Exemplo 2.55 — Teorema de Bayes [4]
Retome o Exemplo 2.51 e calcule agora a probabilidade de o dispositivo ser do tipo A sabendo que foi eficaz em caso de acidente grave.

- **Evento**

  $A|E =$ dispositivo do tipo A dado que foi eficaz em caso de acidente grave

- **Prob. pedida**

$$\begin{aligned} P(A|E) &\overset{teorema\ Bayes}{=} \frac{P(E|A) \times P(A)}{P(E)} \\ &= \frac{0.95 \times 0.5}{0.955} \\ &= 0.4974. \end{aligned}$$

•

Exemplo 2.56 — Lei da probabilidade total e teorema de Bayes [5]
Quatro instrumentos de corte, um de cada uma das marcas M_1, M_2, M_3 e M_4, funcionam satisfatoriamente com probabilidade 0.9, 0.8, 0.6, 0.4, respectivamente para cada marca.

(a) Determine a probabilidade de um instrumento, seleccionado ao acaso desses quatro, funcionar satisfatoriamente.

[4] Adaptado do Exame de 1a. Época, 24 de Junho de 2006.
[5] Exame de Época Especial, 8 de Setembro de 2004.

- **Quadro de eventos e probabilidades**

Evento	**Probabilidade**
$M_i =$ instrum. ser da marca i	$P(M_i) = 0.25,\ i = 1,2,3,4$
$S =$ instrum. func. satisf.	$P(S) = ?$
$S\|M_i =$ instrum. func. satisf. dado que é da marca i	$P(S\|M_i) = \begin{cases} 0.9,\ i=1 \\ 0.8,\ i=2 \\ 0.6,\ i=3 \\ 0.4,\ i=4 \end{cases}$

- **Evento**

  $S =$ instrum. seleccionado funcionar satisfatoriamente

- **Prob. pedida**

$$
\begin{aligned}
P(S) &\overset{lei\ prob.\ total}{=} \sum_{i=1}^{4} P(S|M_i) \times P(M_i) \\
&= (0.9 + 0.8 + 0.6 + 0.4) \times 0.25 \\
&= 0.675.
\end{aligned}
$$

(b) Sabendo que o instrumento seleccionado ao acaso não funciona satisfatoriamente, qual a probabilidade de se ter selecionado um dos dois instrumentos menos fiáveis?

- **Evento**

  $(M_3 \cup M_4)|\overline{S} =$ instrum. seleccionado ser das duas marcas menos fiáveis dado que não funcionou satisfatoriamente

- **Prob. pedida**

$$
\begin{aligned}
P[(M_3 \cup M_4)|\overline{S}] &= P(M_3|\overline{S}) + P(M_4|\overline{S}) \\
&\overset{teorema\ Bayes}{=} \frac{P(\overline{S}|M_3) \times P(M_3)}{P(\overline{S})} + \frac{P(\overline{S}|M_4) \times P(M_4)}{P(\overline{S})} \\
&= \frac{[1 - P(S|M_3)] \times P(M_3)}{1 - P(S)} \\
&\quad + \frac{[1 - P(S|M_4)] \times P(M_4)}{1 - P(S)} \\
&= \frac{(1 - 0.6) \times 0.25}{1 - 0.675} + \frac{(1 - 0.4) \times 0.25}{1 - 0.675} \\
&= 0.769.
\end{aligned}
$$

- **Importante**

  De realçar que:

$$P[(M_3 \cup M_4)|\overline{S}] \neq 1 - P[(M_3 \cup M_4)|S].$$

  Com efeito, aplicando mais uma vez o teorema de Bayes, tem-se :

$$\begin{aligned}
1 - P[(M_3 \cup M_4)|S] &= 1 - [P(M_3|S) + P(M_4|S)] \\
&= 1 - \left[\frac{P(S|M_3) \times P(M_3)}{P(S)} + \frac{P(S|M_4) \times P(M_4)}{P(S)}\right] \\
&= 1 - \left(\frac{0.6 \times 0.25}{0.675} + \frac{0.4 \times 0.25}{0.675}\right) \\
&\simeq 0.630.
\end{aligned}$$

•

2.5 Acontecimentos independentes.

Definição 2.57 — Eventos independentes
Os eventos A e B dizem-se independentes sse

$$P(A \cap B) = P(A) \times P(B), \tag{2.22}$$

Neste caso é usual escrever-se $A \perp\!\!\!\perp B$. •

Exemplo 2.58 — Lei da probabilidade total; eventos independentes[6]
75% da população de Nicosia (Chipre) é grega e 25% turca. Apurou-se também que 20% dos gregos e 10% dos turcos falam inglês.

(a) Qual a percentagem da população de Nicosia que fala inglês?

- **Quadro de eventos e probabilidades**

Evento	Probabilidade
G = habitante grego	$P(G) = 0.75$
T = habitante turco	$P(T) = 0.25$
I = habitante falar inglês	$P(I) = ?$
$I\|G$ = habitante falar inglês dado que é grego	$P(I\|G) = 0.20$
$I\|T$ = habitante falar inglês dado que é turco	$P(I\|T) = 0.10$

- **Evento**

  I = habitante falar inglês

- **Prob. pedida**

$$\begin{aligned} P(I) &\overset{lei\ prob.\ total}{=} P(I|G) \times P(G) + P(I|T) \times P(T) \\ &= 0.20 \times 0.75 + 0.10 \times 0.25 \\ &= 0.175. \end{aligned}$$

[6] Adaptado do Teste A, 22 de Abril de 2006.

(b) Serão os eventos “ser grego” e “falar inglês” eventos independentes?

- **Averiguação de independência**

$$\begin{array}{rcl} P(G \cap I) & = & P(I|G) \times P(G) \\ & = & 0.20 \times 0.75 \\ & = & 0.15 \\ & \neq & \\ P(G) \times P(I) & = & 0.75 \times 0.175 \\ & = & 0.13125. \end{array}$$

Já que $P(G \cap I) \neq P(G) \times P(I)$ pode afirmar-se que G e I não são eventos independentes como, aliás, seria de esperar. •

Proposição 2.59 — **Consequências**

1. Sejam A e B eventos independentes tais que $P(A) > 0$ e $P(B) > 0$. Então:

   - $P(A|B) = P(A)$;
   - $P(B|A) = P(B)$.

   Isto é, o conhecimento de B não afecta a reavaliação da probabilidade de A, e vice-versa.

2. Sejam A e B dois eventos tais que:

   - $A \cap B = \emptyset$ (eventos disjuntos);
   - $P(A) > 0$ e $P(B) > 0$.

   Então $A \not\perp\!\!\!\perp B$, i.e., A e B não são independentes.

3. Tem-se, para qualquer evento A:

   - $A \perp\!\!\!\perp \emptyset$;
   - $A \perp\!\!\!\perp \Omega$.

4. Sejam A e B dois eventos independentes. Então:

   - $\overline{A} \perp\!\!\!\perp B$;
   - $A \perp\!\!\!\perp \overline{B}$;
   - $\overline{A} \perp\!\!\!\perp \overline{B}$. •

Exercício 2.60 — Eventos independentes
Demonstre a Proposição 2.59. •

Nota 2.61 — Independência completa (três eventos)
Os eventos A, B e C dizem-se completamente independentes sse

- $P(A \cap B) = P(A) \times P(B)$;
- $P(A \cap C) = P(A) \times P(C)$;
- $P(B \cap C) = P(B) \times P(C)$;
- $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$. •

Nota 2.62 — Independência completa (n eventos)
Os eventos $A_1, \ldots, A_n$ dizem-se completamente independentes sse o forem dois a dois, três a três, …, n a n. •

Capítulo 3

Variáveis aleatórias e distribuições discretas

Em algumas situações os resultados das e.a. são numéricos, como é o caso de medições, contagens, etc. Noutras os resultados possíveis constituem um espaço não numérico.[1] Basta pensar na classificação de 2 artigos, seleccionados da produção de uma fábrica, quanto a serem defeituosos ou não defeituosos.

Ao realizar esta e outras e.a. é frequente não estarmos interessados nos resultados detalhados da mesma[2] mas somente numa quantidade numérica determinada pelo resultado da e.a., por exemplo, o número de artigos defeituosos. Neste caso atribui-se um número a cada resultado da e.a., sendo que tal atribuição pode ser puramente convencional...

Passemos à definição (informal) de variável aleatória: a de função — com características especiais — que transforma eventos em números ou, mais genericamente, em vectores.

Definição 3.1 — **Variável aleatória**

A função X diz-se uma variável aleatória (v.a.) se possuir

- domínio Ω,
- contradomínio $I\!R_X$ contido em $I\!R^n$

e tiver a particularidade de verificar uma condição de mensurabilidade.[3] Assim,

[1] Leia-se: Ω não é um subconjunto de $I\!R^n$.

[2] Neste caso particular os resultados são pares ordenados: (D_1, D_2), $(\overline{D}_1, D_2)$, etc., onde D_i ($\overline{D}_i$) representa a classificação do artigo i de defeituoso (não defeituoso, respectivamente).

[3] Esta condição será explicitada dentro em breve.

$$X : \Omega \to I\!R_X \subset I\!R^n, \tag{3.1}$$

i.e., a imagem de um evento $\omega \in \Omega$ é um vector $X(\omega) \in I\!R_X \subset I\!R^n$. •

Nota 3.2 — Tipos de variáveis aleatórias
Dependendo do valor da constante n estaremos a lidar com v.a.s com carácteres distintos:

- se $n = 1$, a v.a. diz-se unidimensional;
- se $n = 2$, a v.a. diz-se bidimensional;
- se $n > 2$, a v.a. diz-se multidimensional.

No âmbito desta disciplina debruçar-nos-emos em particular nos casos unidimensional e bidimensional.

Dependendo do cardinal do conjunto de valores possíveis da v.a. estaremos a lidar grosso modo com os seguintes tipos de v.a.:

- se $\#I\!R_X = \#I\!N$ (ou seja, $I\!R_X$ é finito ou infinito numerável), a v.a. diz-se discreta;
- se $\#I\!R_X = \#I\!R$ (i.e., infinito não numerável) e..., a v.a. diz-se contínua. •

Nota 3.3 — Condição de mensurabilidade
Considere-se que uma e.a. possui espaço de resultados Ω associado à $\sigma-$álgebra $\mathcal{A}$. E seja X uma v.a. unidimensional.

A condição de mensurabilidade prende-se neste caso com a existência de imagem inversa (segundo X) de qualquer intervalo do tipo $(-\infty, x]$ na $\sigma-$álgebra $\mathcal{A}$:

$$\forall x \in I\!R, X^{-1}((-\infty, x]) \in \mathcal{A},$$

onde $X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\}$. •

3.1 Variáveis aleatórias discretas.

Definição 3.4 — **V.a. discreta**
A v.a. X diz-se discreta, caso tome valores em $I\!R_X = \{x_1, x_2, \ldots\}$, em número finito ou infinito numerável, e

$$P(X \in I\!R_X) = P(X \in \{x_1, x_2, \ldots\}) = 1. \tag{3.2}$$

•

Exemplo/Exercício 3.5 — **Variável aleatória discreta**
Considere-se um teste americano com 3 questões, cujas respostas são dadas de forma independente. A resposta a cada questão pode estar correcta (C) com probabilidade $P(C) = 0.5$, ou incorrecta ($\overline{C}$) com probabilidade $P(\overline{C}) = 0.5$.

(a) Identifique a e.a. e o respectivo espaço de resultados.

- **E.a.**

  Classificação de 3 questões em teste americano

- **Eventos-chave**

  C =questão classificada como correcta

  $\overline{C}$ =questão classificada como incorrecta

- **Espaço de resultados**

  $\Omega = \{\overline{C}\overline{C}\overline{C}, C\overline{C}\overline{C}, \overline{C}C\overline{C}, \overline{C}\overline{C}C, CC\overline{C}, C\overline{C}C, \overline{C}CC, CCC\}$.

  Este espaço de resultados é constituído por 8 eventos elementares.

  Por exemplo, $\overline{C}\overline{C}\overline{C}$ representa de forma abreviada a classificação das 3 questões como incorrectas e $\overline{C}C\overline{C}$ a classificação da segunda questão como correcta e as restantes como incorrectas.

(b) Defina uma v.a. que considere pertinente e determine o conjunto dos seus valores possíveis (i.e., o seu contradomínio). Represente a v.a. esquematicamente enquanto função que transforma eventos em números reais.

- **V.a.**

  X = número de respostas correctas no teste americano

- **Contradomínio e representação esquemática de X**

  $I\!R_X = \{0, 1, 2, 3\}$

$$X(\overline{C}\,\overline{C}\,\overline{C}) = 0$$

$$X(C\overline{C}\,\overline{C}) = X(\overline{C}C\overline{C}) = X(\overline{C}\,\overline{C}C) = 1, \text{ etc.}$$

3.2 Função de probabilidade.

A definição de uma v.a. discreta só fica completa a partir do momento em que se define a probabilidade de ocorrência de cada um dos elementos de $I\!R_X$, i.e., a partir do momento em que se define a função de probabilidade (ou função massa de probabilidade) da v.a. discreta X.

Definição 3.6 — **Função de probabilidade**
Seja X uma v.a. discreta com contradomínio $I\!R_X = \{x_1, x_2, \ldots\}$ (necessariamente finito ou infinito numerável). Então a função de probabilidade (f.p.) de X é dada por

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}) = \begin{cases} P(X = x_i), \text{ se } x = x_i \in I\!R_X \\ 0, \text{ c.c.} \end{cases} \tag{3.3}$$

e satisfaz:

- $P(X = x) > 0, \ \forall x \in I\!R_X$;
- $\sum_{x \in I\!R} P(X = x) = \sum_{x \in I\!R_X} P(X = x) = \sum_i P(X = x_i) = 1.$ •

Exemplo/Exercício 3.7 — **Função de probabilidade**
Retome o Exemplo 3.5 e defina a f.p. da v.a. X e desenhe o respectivo gráfico.

- **V.a.**

  $X =$ número de respostas correctas no teste americano

- **F.p. de X**

$$\begin{aligned}
P(X = 0) &= P(\overline{C}\overline{C}\overline{C}) \\
&= P(\overline{C}_1\overline{C}_2\overline{C}_3) \\
&\overset{ev.\ indep}{=} P(\overline{C}_1) \times P(\overline{C}_2) \times P(\overline{C}_3) \\
&\overset{ev.\ equiprov.}{=} P(\overline{C}) \times P(\overline{C}) \times P(\overline{C}) \\
&= (1/2)^3 \\
P(X = 1) &= P(C\overline{C}\overline{C}) + P(\overline{C}C\overline{C}) + P(\overline{C}\overline{C}C) \\
&= \ldots \\
&= 3 \times (1/2)^3
\end{aligned}$$

$$\begin{aligned}
P(X=2) &= P(CC\overline{C}) + P(C\overline{C}C) + P(\overline{C}CC) \\
&= \ldots \\
&= 3 \times (1/2)^3 \\
P(X=3) &= P(CCC) \\
&= \ldots \\
&= (1/2)^3.
\end{aligned}$$

Ou de forma resumida

$$P(X=x) = \begin{cases} 1/8, & x = 0,3 \\ 3/8, & x = 1,2 \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Gráfico da f.p. de** X

De notar que esta função possui 4 pontos de descontinuidade, tantos quantos o número de valores possíveis da v.a. X. •

3.3 Função de distribuição.

Motivação 3.8 — É usual falar-se na probabilidade da v.a. X tomar valores não superiores a um número real arbitrário x. É com o objectivo de obter tal probabilidade que definiremos a função de distribuição da v.a. discreta X. •

Definição 3.9 — **Função de distribuição**
A função de distribuição (f.d.) de uma v.a. X é dada por

$$F_X(x) = P(X \leq x),\ x \in I\!R, \tag{3.4}$$

independentemente do seu carácter ser discreto ou contínuo. •

Nota 3.10 — **Função de distribuição**
Relembre-se que, por tratar-se da probabilidade da v.a. X atribuir valores menores ou iguais ao real arbitrário x, tem-se

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}),\ x \in I\!R, \tag{3.5}$$

ou seja, $F_X(x)$ é igual à probabilidade da colecção de todos os eventos (da $\sigma-$álgebra de Ω) aos quais a v.a. X atribui valores no intervalo $(-\infty, x]$. •

Definição 3.11 — **Função de distribuição de v.a. discreta**
A f.d. da v.a. discreta X (com contradomínio $I\!R_X = \{x_1, x_2, \ldots\}$) pode escrever-se à custa da f.p. de X:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i),\ x \in I\!R. \tag{3.6}$$

•

Nota 3.12 — **Função de distribuição**
Retenha-se também que a f.d. possui domínio real e toma valores no intervalo $[0,1]$ já que se trata de uma probabilidade, ou seja:

$$F_X : I\!R \rightarrow [0,1], \tag{3.7}$$

quer X seja uma v.a. discreta ou contínua. •

Esta é uma de diversas propriedades satisfeitas pela f.d. de qualquer v.a. discreta, também patente na resolução do exercício que se segue.

Exemplo/Exercício 3.13 — Função de distribuição de v.a. discreta
Determine a f.d. da v.a. X definida no Exemplo 3.5 e desenhe o gráfico de $F_X(x)$ e nomeie algumas características desta função.

- **V.a.**

  X = número de respostas correctas no teste americano

- **F.p. de** X

$$P(X = x) \quad = \quad \begin{cases} 1/8, & x = 0 \\ 3/8, & x = 1 \\ 3/8, & x = 2 \\ 1/8, & x = 3 \\ 0, & \text{outros valores de } x. \end{cases}$$

- **F.d. de** X

  Comece-se por preencher a tabela abaixo — de forma justificada — com alguns valores da f.d. de X.

x	$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$	Obs.
-0.5	$F_X(-0.5) = P(X \leq -0.5) = 0$	
0	$F_X(0) = P(X \leq 0) = P(X = 0) = 1/8$	
0.3	$F_X(0.3) = P(X \leq 0.3) = P(X = 0) = 1/8$	
1	$F_X(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 1/2$	
1.4	$F_X(1.4) = P(X \leq 1.4) = \ldots$	
2	$F_X(2) = \ldots$	
2.8	$F_X(2.8) = \ldots$	
3	$F_X(3) = \ldots$	
10.5	$F_X(10.5) = \ldots$	

  Pode então concluir-se que:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{x_i \leq x} P(X = x_i) \end{aligned}$$

$$= \begin{cases} 0, & x < 0 \\ 1/8, & 0 \leq x < 1 \\ 1/8 + 3/8 = 1/2, & 1 \leq x < 2 \\ 1/8 + 3/8 + 3/8 = 7/8, & 2 \leq x < 3 \\ 1/8 + 3/8 + 3/8 + 1/8 = 1, & x \geq 3. \end{cases}$$

- **Gráfico da f.d. de** X

  (Procure identificar graficamente os diversos valores da f.p. de X.)

- **Algumas propriedades da f.d. de** X

  Função em escada;

  Função monótona não decrescente (ou crescente em sentido lato); etc. •

Proposição 3.14 — Propriedades da f.d. de v.a. discreta
A f.d. da v.a. discreta X, $F_X(x)$, é uma:

1. Função em escada que possui tantos pontos de descontinuidade quantos os valores distintos de $I\!R_X$;

2. Função contínua à direita;[4]

3. Função monótona não decrescente de x.[5]

Entre outras propriedades da f.d. de uma v.a discreta X destaque-se também:

4. $0 \leq F_X(x) \leq 1$;

[4]Ou seja, $F_X(x) = F_X(x^+)$, $\forall x \in I\!R$, onde $F_X(x^+)$ representa o limite à direita da f.d., num ponto real x arbitrário. Recorde-se que este limite é definido por $F_X(x^+) = \lim_{h \to 0} F_X(x+h), h > 0$. Note-se também que este limite nem sempre é igual ao limite à esquerda no ponto x, $F_X(x^-) = \lim_{h \to 0} F_X(x - h), h > 0$, caso se esteja a lidar com uma v.a. discreta.

[5]I.e., $F_X(x) \leq F_X(x+h)$, $\forall h > 0, x \in I\!R$.

5. $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$;

6. $F_X(+\infty) = \lim_{x\to+\infty} F_X(x) = 1$;

7. $P(X < x) = F_X(x^-) = \lim_{h\to 0} F_X(x-h), h > 0$;

8. $P(X > x) = 1 - F_X(x)$;

9. $P(X \geq x) = 1 - F_X(x^-)$;

e ainda, para $a < b$,

10. $P(a < X \leq b) = F_X(b) - F_X(a)$,

11. $P(a < X < b) = F_X(b^-) - F_X(a)$,

12. $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$,

13. $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$.

Refira-se por fim que, a f.p. da v.a. discreta X, $P(X = x)$ pode escrever-se à custa da f.d. de X:

$$P(X = x) = F_X(x) - F_X(x^-), \tag{3.8}$$

que corresponde ao salto da f.d. de X no ponto x. •

Importa notar que todas as propriedades aqui referidas — à excepção de 1. — são partilhadas pela f.d. de qualquer v.a. contínua. Refira-se também que algumas destas propriedades serão rescritas de modo a reflectir o carácter contínuo da v.a.

Exemplo 3.15 — Relação entre a função de probabilidade e a função de distribuição
Retome o Exemplo 3.5.

(a) Obtenha a f.p. da v.a. X que representa o número de questões correctas no teste americano à custa da f.d. desta v.a.

- **F.p. de** X

  $P(X=0) = F_X(0) - F_X(0^-) = 1/8 - 0 = 1/8$
  $P(X=1) = F_X(1) - F_X(1^-) = F_X(1) - F_X(0) = 1/2 - 1/8 = 3/8$
  $P(X=2) = F_X(2) - F_X(2^-) = F_X(2) - F_X(1) = 7/8 - 1/2 = 3/8$
  $P(X=3) = F_X(3) - F_X(3^-) = F_X(3) - F_X(2) = 1 - 7/8 = 1/8.$

(b) Determine $P(0 < X < 3)$ recorrendo quer à f.p. de X, quer à f.d. de X.

- **Probabilidade pedida**

  $P(0 < X < 3) = P(X=1) + P(X=2) = 3/8 + 3/8 = 3/4$
  $P(0 < X < 3) = F_X(3^-) - F_X(0) = F_X(2) - F_X(0) = 7/8 - 1/8 = 3/4.$

  Como se pode constatar os resultados são iguais. •

3.4 Valor esperado, variância e algumas das suas propriedades. Moda e quantis.

Para descrever completamente o comportamento probabilístico de uma v.a. X é necessário recorrer à função de distribuição. Caso X seja uma v.a. discreta, pode recorrer-se em alternativa à função de probabilidade.

Pode, no entanto, optar-se por uma caracterização parcial da v.a. X.

Definição 3.16 — Parâmetro
Um parâmetro não passa de um indicador ou medida sumária capaz de caracterizar — embora parcialmente — algum aspecto de uma v.a. X. Entre eles destacaremos os:

- Parâmetros de localização central
  - valor esperado
  - moda
  - mediana;
- Parâmetros de localização não central
  - quantil de probabilidade (ou quantil de ordem) p;
- Parâmetros de dispersão
  - variância
  - desvio-padrão
  - coeficiente de variação. •

Motivação 3.17 — Valor esperado de v.a. discreta
O conceito de valor esperado teve a sua origem nos jogos de acaso e ao que parece a sua introdução deve-se a Christiaan Huygens (1629–95).

O valor esperado de X corresponde ao seu centro de gravidade. Esta analogia física tem a sua razão de ser: para tal basta pensar num sistema de corpos com

- massas m_i
- posições x_i,

cujo centro de massa é dado por $\frac{\sum_i x_i\, m_i}{\sum_i m_i}$. Assim, o valor esperado de X mais não é que o centro de "massa de probabilidade" desta v.a. •

Definição 3.18 — Valor esperado de v.a. discreta
O valor esperado da v.a. discreta X é dado por

$$E(X) = \sum_x x\, P(X = x). \tag{3.9}$$

•

Nota 3.19 — Valor esperado de v.a. discreta

1. $E(X) = \sum_{x \in I\!R} x\, P(X = x) = \sum_{x \in I\!R_X} x\, P(X = x)$.

2. $E(X)$ existe sse $\sum_{x \in I\!R_X} |x|\, P(X = x)$, i.e., sse a série for absolutamente convergente.

3. De um modo geral $E(X) \not\in I\!R_X$, i.e., o valor esperado de X não pertence ao conjunto de valores possíveis de X, caso X seja uma v.a. discreta.

4. É usual escolher $E(X)$ como o valor representativo da v.a. X, em particular se X possuir distribuição com tendência central uma vez que neste caso é em torno desse ponto que mais se concentram as probabilidades.

5. O valor esperado de X é também representado/designado por μ, μ_X, média, valor médio, esperança matemática, etc. •

Proposição 3.20 — Propriedades do valor esperado
O valor esperado da v.a. X satisfaz as seguintes propriedades.

1. $E(b) = b$, $\forall b \in I\!R$.

2. $E(aX + b) = aE(X) + b$, $\forall a, b \in I\!R$, ou seja, o valor esperado é um operador linear.

3. Seja $Y = \psi(X)$ uma v.a. função (mensurável) de X. Então

$$E(Y) = E[\psi(X)] = \sum_x \psi(x)\, P(X = x), \tag{3.10}$$

   caso X seja uma v.a. discreta.

4. De um modo geral

$$E[\psi(X)] \neq \psi[E(X)]. \tag{3.11}$$

5. Seja X uma v.a. inteira não negativa, i.e., $\mathbb{R}_X = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Então

$$E(X) = \sum_{x=0}^{+\infty} P(X > x) = \sum_{x=0}^{+\infty} [1 - F_X(x)]. \tag{3.12}$$

•

Exercício 3.21 — Demonstre a Proposição 3.20. •

Exemplo 3.22 — **Valor esperado de v.a. discreta**

(a) Calcule o valor esperado do número de respostas correctas no teste americano descrito no Exemplo 3.5.

- **V.a.**

  $X =$ número de respostas correctas no teste americano

- **F.p. de X**

$$P(X = x) = \begin{cases} 1/8, & x = 0, 3 \\ 3/8, & x = 1, 2 \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Valor esperado de X**

$$\begin{aligned} E(X) &= \sum_{x=0}^{3} x \times P(X = x) \\ &= 0 \times 1/8 + 1 \times 3/8 + 2 \times 3/8 + 3 \times 1/8 \\ &= 1.5. \end{aligned}$$

  Importa notar que $E(X)$ coincide, neste caso, com o ponto de simetria da f.p. de X.

(b) O número de neutrinos registados em intervalos de 12 segundos, aquando da primeira observação da supernova S1987a por astrónomos, é bem modelado pela v.a. X com f.p.

$$P(X = x) = \begin{cases} \frac{e^{-0.8} \times 0.8^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{outros valores de } x. \end{cases}$$

Obtenha o valor esperado de X.[6]

[6]Adaptado do Teste A, 22 de Abril de 2006.

- **V.a.**

  $X =$ número de neutrinos registados num intervalo de 12 segundos

- **F.p. de X**

$$P(X = x) = \begin{cases} \frac{e^{-0.8} \times 0.8^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Valor esperado de X**

$$\begin{aligned} E(X) &= \sum_{x=0}^{+\infty} x \times P(X = x) \\ &= \sum_{x=0}^{+\infty} x \times \frac{e^{-0.8} \times 0.8^x}{x!} \\ &= \sum_{x=1}^{+\infty} \frac{e^{-0.8} \times 0.8^x}{(x-1)!} \\ &= 0.8 \times e^{-0.8} \sum_{x=0}^{+\infty} \frac{0.8^x}{x!} \\ &= 0.8. \end{aligned}$$

  Aqui temos outro exemplo de um valor esperado que não pertence ao conjunto de valores possíveis da v.a. X. •

Definição 3.23 — Moda de v.a. discreta
É importante saber qual o valor da v.a. X que ocorre com mais frequência. Este valor será designado por $mo = mo(X)$ e corresponde ao ponto de máximo da f.p. de X, i.e.,

$$mo: \; P(X = mo) = \max_x P(X = x), \tag{3.13}$$

ou, equivalentemente, $mo = \arg\max_x P(X = x)$. •

Nota 3.24 — Moda de v.a. discreta
A moda de uma v.a. nem sempre é única como ilustra o exemplo seguinte e já agora refira-se que

- X se diz bimodal se possuir duas modas. •

Exemplo 3.25 — **Moda de v.a. discreta**

Obtenha a moda do:

(a) Número de respostas correctas no teste americano (Exemplo 3.5).

- **V.a.**

  $X =$ número de respostas correctas no teste americano

- **F.p. de** X

$$P(X = x) \quad = \quad \begin{cases} 1/8, & x = 0, 3 \\ 3/8, & x = 1, 2 \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Moda de** X

  $mo = mo(X) = 1$ e 2 já que $P(X = 1) = P(X = 2) = \max_x P(X = x)$.

  Logo a v.a. X é bimodal.

(b) Número de neutrinos registado num intervalo de 12 segundos (Exemplo 3.22).

- **V.a.**

  $X =$ número de neutrinos registados num intervalo de 12 segundos

- **F.p. de** X

$$P(X = x) \quad = \quad \begin{cases} \frac{e^{-0.8} \times 0.8^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Moda de** X

  A obtenção do ponto de máximo de uma f.p. não passa pela determinação de um ponto de estacionaridade (de uma função contínua e diferenciável) mas sim pela determinação de

$$mo = mo(X) \in I\!R_X \quad : \quad \begin{cases} P(X = mo) \geq P(X = mo - 1) \\ P(X = mo) \geq P(X = mo + 1) \end{cases}$$
$$\begin{cases} \frac{P(X=mo)}{P(X=mo-1)} \geq 1 \\ \frac{P(X=mo+1)}{P(X=mo)} \leq 1. \end{cases} \tag{3.14}$$

  Ao substituir-se a f.p. de X nas duas desigualdades acima, conclui-se que

$$mo = mo(X) \in I\!R_X : \begin{cases} \frac{0.8}{mo} \geq 1 \\ \frac{0.8}{mo+1} \leq 1, \end{cases}$$

  i.e., $mo = mo(X) = 0$. •

Motivação 3.26 — Mediana de v.a. discreta

A mediana da v.a. X, $me = me(X)$, tem a particularidade de verificar

$$\begin{cases} P(X \leq me) \geq \frac{1}{2} \\ P(X \geq me) \geq \frac{1}{2}, \end{cases} \tag{3.15}$$

pelo que a probabilidade de registarmos valores da v.a. X não superiores (não inferiores) à mediana é de pelo menos 50%. •

Definição 3.27 — Mediana de v.a. discreta

A mediana da v.a. discreta X, $me = me(X)$,[7] verifica a dupla desigualdade

$$me : \frac{1}{2} \leq F_X(me) \leq \frac{1}{2} + P(X = me), \tag{3.16}$$

por sinal equivalente a (3.15) e já agora equivalente a $F_X(me^-) \leq \frac{1}{2} \leq F_X(me)$.[8] •

Nota 3.28 — Mediana de v.a. discreta

Ao lidarmos com v.a. discretas a mediana pode não ser única, passando nesse caso a falar-se em classe mediana.[9] •

Exemplo 3.29 — Mediana de v.a. discreta

Determine a mediana do número de respostas correctas no teste americano (Exemplo 3.5).

- **V.a.**

  X = número de respostas correctas no teste americano

- **F.d. de X**

  Esta função foi determinada previamente e é igual a:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \begin{cases} 0, & x < 0 \\ 1/8, & 0 \leq x < 1 \\ 1/8 + 3/8 = 1/2, & 1 \leq x < 2 \\ 1/8 + 3/8 + 3/8 = 7/8, & 2 \leq x < 3 \\ 1/8 + 3/8 + 3/8 + 1/8 = 1, & x \geq 3. \end{cases} \end{aligned}$$

[7] me é também representada por $F_X^{-1}(1/2)$.

[8] Esta talvez seja a forma mais expedita de identificar a(s) mediana(s) de uma v.a. discreta.

[9] Ou, de acordo com alguns autores, considera-se que a mediana corresponde ao menor dos valores da classe mediana.

- **Mediana de X**

  Tirando partido da expressão de $F_X(x)$ e da definição de mediana de uma v.a. discreta pode construir-se a seguinte tabela que servirá para identificar a(s) mediana(s) de X:

Candidato a me	$\frac{1}{2} \leq F_X(me) \leq \frac{1}{2} + P(X = me)$	Obs.
1	$\mathbf{\frac{1}{2}} \leq F_X(1) = \mathbf{\frac{1}{2}} \leq \frac{1}{2} + P(X = 1) = \frac{1}{2} + \frac{3}{8} = \mathbf{\frac{7}{8}}$	Prop. verd.
1.5	$\mathbf{\frac{1}{2}} \leq F_X(1.5) = \mathbf{\frac{1}{2}} \leq \frac{1}{2} + P(X = 1.5) = \frac{1}{2} + 0 = \mathbf{\frac{1}{2}}$	Prop. verd.
2	$\mathbf{\frac{1}{2}} \leq F_X(2) = \mathbf{\frac{7}{8}} \leq \frac{1}{2} + P(X = 2) = \frac{1}{2} + \frac{3}{8} = \mathbf{\frac{7}{8}}$	Prop. verd.
2.1	$\mathbf{\frac{1}{2}} \leq F_X(2.1) = \mathbf{\frac{7}{8}} \leq \frac{1}{2} + P(X = 2.1) = \frac{1}{2} + 0 = \mathbf{\frac{1}{2}}$	Prop. FALSA.

  Deste modo conclui-se que mediana da v.a. X não é única e que qualquer valor no intervalo $[1, 2]$ é mediana de X. Não surpreende neste caso que se designe o intervalo $[1, 2]$ de classe mediana.

  Mais, ao notar-se que $me : F_X(me^-) \leq 1/2 \leq F_X(me)$ evita-se o recurso à f.p. de X e identifica(m)-se a(s) mediana(s) desta v.a. discreta de uma forma expedita. (Averigue...) •

Motivação 3.30 — Quantil de probabilidade p de v.a. discreta
A mediana é um caso particular de um outro parâmetro de localização não central mais genérico, o quantil de probabilidade (ou quantil de ordem) p $(0 < p < 1)$, χ_p, que verifica

$$\begin{cases} P(X \leq \chi_p) \geq p \\ P(X \geq \chi_p) \geq 1 - p. \end{cases} \tag{3.17}$$

Assim sendo a probabilidade de registarmos valores da v.a. X não superiores (não inferiores, resp.) ao quantil de probabilidade p é de pelo menos $p \times 100\%$ ($(1-p) \times 100\%$, resp.). •

Definição 3.31 — Quantil de probabilidade p de v.a. discreta
O quantil de probabilidade p $(0 < p < 1)$ da v.a. X, χ_p,[10] satisfaz

$$\chi_p : p \leq F_X(\chi_p) \leq p + P(X = \chi_p), \tag{3.18}$$

obviamente equivalente a (3.17) e a $F_X(\chi_p^-) \leq p \leq F_X(\chi_p)$. •

[10] χ_p é também representado por $F_X^{-1}(p)$.

Nota 3.32 — Quantil de probabilidade p de v.a. discreta
A mediana da v.a. X corresponde a $\chi_{1/2} = F_X^{-1}(1/2)$. Outros quantis importantes:

- $\chi_{1/4} = F_X^{-1}(1/4) =$ 1o. quartil;
- $\chi_{3/4} = F_X^{-1}(3/4) =$ 3o. quartil;
- $\chi_{1/100} = F_X^{-1}(1/100) =$ 1o. percentil;
- $\chi_{n/100} = F_X^{-1}(n/100) = n-$ésimo percentil, $n = 1, 2, \ldots, 99$;
- $\chi_{0.1} = F_X^{-1}(1/10) =$ 1o. decil. •

Exemplo 3.33 — Quantil de ordem p de v.a. discreta
O número de navios que chegam diariamente a um porto (X) é uma v.a. com f.p.

$$P(X = x) = e^{-2} \times \frac{2^x}{x!}, x = 0, 1, 2, \ldots$$

As condições do porto impedem que atraquem mais de 3 navios por dia, sendo os restantes navios reencaminhados para outro porto.[11]

(a) Qual a probabilidade de serem reencaminhados um ou mais navios para o outro porto (num dia escolhido arbitrariamente)?

- **V.a.**
  $X =$ número de navios que chegam diariamente ao porto
- **F.p. de X**
  $P(X = x) = e^{-2} \times \frac{2^x}{x!}, x = 0, 1, 2, \ldots$
- **Probabilidade pedida**
  Seja R o evento que representa o reencaminhamento de um ou mais navios para o outro porto. Então

$$\begin{aligned} P(R) &= P(X > 3) \\ &= 1 - P(X \leq 3) \\ &= 1 - \sum_{x=0}^{3} P(X = x) \\ &= 1 - \sum_{x=0}^{3} e^{-2} \times \frac{2^x}{x!} \\ &= 1 - 0.8571 \\ &= 0.1429. \end{aligned}$$

[11] Adaptado de Exame de Época Especial, 13 de Setembro de 2002.

(b) Qual deve ser a capacidade mínima do porto de modo que o reencaminhamento de um ou mais navios ocorra no máximo em 5% dos dias?

- **Obtenção da capacidade mínima** (c')

  c' é o menor inteiro c que verifica $P(\text{reencaminhamento navios}) \leq 0.05$, i.e.,

$$\begin{aligned} c \ : \ & P(X > c) \leq 0.05 \\ & 1 - P(X \leq c) \leq 0.05 \\ & 1 - F_X(c) \leq 0.05 \\ & F_X(c) \geq 0.95 \\ & c \geq F_X^{-1}(0.95), \end{aligned}$$

  onde $F_X^{-1}(0.95)$ representa o quantil de ordem 0.95 da v.a. X e, como seria de esperar, a capacidade mínima c'.

  A obtenção de c' passa pela determinação de $F_X(c)$ para vários valores de c. Ora, tirando partido do resultado da alínea anterior ($F_X(3) = 0.8571$ e da monotonia não decrescente de qualquer f.d., bastará considerar valores de c maiores que 3, tal como ilustramos na tabela seguinte:[12]

c	$F_X(c) = \sum_{x=0}^{c} e^{-2} \times \frac{2^x}{x!}$	$F_X(c) \geq 0.95$?
4	$F_X(4) = F_X(3) + P(X=4) = 0.8571 + e^{-2} \times \frac{2^4}{4!} = 0.9473$	Não
5	$F_X(5) = F_X(4) + P(X=5) = 0.9473 + e^{-2} \times \frac{2^5}{5!} = 0.9834$	SIM

  Deste modo, conclui-se que $c' = F_X^{-1}(0.95) = 5$. •

[12] Veremos mais tarde que a obtenção de c' poderia fazer-se por recurso às tabelas da f.d. daquilo que chamaremos de v.a. de Poisson.

Motivação 3.34 — Variância
É necessário definir um parâmetro que dê ideia da dispersão dos valores da v.a. em torno do centro de gravidade da mesma.

Este parâmetro corresponde, fisicamente, ao momento de inércia de um sistema em relação a um eixo perpendicular ao eixo das abcissas e passando pelo centro de gravidade (Murteira (1990, p. 184)). •

Definição 3.35 — Variância
A variância da v.a. X é dada por

$$V(X) = E(X^2) - E^2(X), \tag{3.19}$$

independentemente do carácter da v.a. ser discreto ou contínuo. Contudo se X for uma v.a. discreta então $V(X)$ obtém-se recorrendo ao facto de neste caso $E(X^2) = \sum_x x^2 P(X = x)$ e $E^2(X) = [\sum_x xP(X = x)]^2$. •

Nota 3.36 — Variância

1. A variância da v.a. X é igual ao seguinte valor esperado $E\{[X - E(X)]^2\}$. No entanto, esta fórmula é de longe muito menos conveniente, pelo que faremos muito pouco uso da mesma.

2. $V(X)$ é por vezes representado por σ^2, σ^2_X, $Var(X)$, etc. •

Exercício 3.37 — Variância
Prove que:

$$\begin{aligned} V(X) &= E\left\{[X - E(X)]^2\right\} = E(X^2) - E^2(X); \\ E(X^2) &= V(X) + E^2(X). \end{aligned}$$ •

Proposição 3.38 — Propriedades da variância
A variância de uma v.a. quer discreta, quer contínua, goza das propriedades:

1. $V(b) = 0, \forall b \in I\!R$;

2. $V(X) \geq 0$, qualquer que seja a v.a. X;

3. $V(aX + b) = a^2 V(X), \forall a, b \in I\!R$. •

Exercício 3.39 — Propriedades da variância

Demonstre a Proposição 3.38. •

Exemplo 3.40 — Variância de uma v.a. discreta

Retome o Exemplo 3.33 e assuma que por cada navio que chega diariamente ao porto a administração do mesmo factura 1200 Euros. Determine o valor esperado e a variância da facturação diária.

- **V.a.**

  $X =$ número de navios que chegam diariamente ao porto

- **F.p. de** X

  $P(X = x) = e^{-2} \times \frac{2^x}{x!}, x = 0, 1, 2, \ldots$

- **Valor esperado de** X

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{+\infty} x \times P(X = x) \\
&= \sum_{x=0}^{+\infty} x \times \frac{e^{-2}\, 2^x}{x!} \\
&= \sum_{x=1}^{+\infty} e^{-2} \times \frac{2^x}{(x-1)!} \\
&= 2\, e^{-2} \times \sum_{x=0}^{+\infty} \frac{2^x}{x!} \\
&= 2
\end{aligned}
$$

- **Variância de** X

  Tendo em conta que

$$
V(X) = E(X^2) - E^2(X), \tag{3.20}
$$

  onde

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{+\infty} x^2 \times P(X = x) \\
&= \sum_{x=0}^{+\infty} x^2 \times \frac{e^{-2}\, 2^x}{x!} \\
&= \sum_{x=1}^{+\infty} x[(x-1)+1] \times \frac{e^{-2}\, 2^x}{x!}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{x=2}^{+\infty} \frac{e^{-2}\, 2^x}{(x-2)!} + \sum_{x=0}^{+\infty} x \times \frac{e^{-2}\, 2^x}{x!} \\
&= 2^2\, e^{-2} \times \sum_{x=0}^{+\infty} \frac{2^x}{x!} + E(X) \\
&= 2^2 + 2,
\end{aligned}
$$

conclui-se que

$$
\begin{aligned}
V(X) &= 2^2 + 2 - 2^2 \\
&= 2 \\
&= E(X).
\end{aligned}
$$

- **Nova v.a.**

  $Y = 1200\, X =$ facturação diária.

  Para a obtenção de $E(Y)$ e $V(Y)$ tiraremos partido do facto de Y ser uma função linear de X e de algumas propriedades do valor esperado e da variância.

- **Valor esperado de Y**

$$
\begin{aligned}
E(Y) &= E(1200\, X) \\
&= 1200 \times E(X) \\
&= 1200 \times 2 \\
&= 2400
\end{aligned}
$$

- **Variância de Y**

$$
\begin{aligned}
V(Y) &= V(1200\, X) \\
&= 1200^2 \times V(X) \\
&= 1200^2 \times 2.
\end{aligned}
$$

•

Motivação 3.41 — Desvio-padrão

A variância não é expressa nas mesmas unidades que a v.a. X, pelo que é costume recorrer a outra medida de dispersão absoluta. Trata-se do desvio-padrão definido a seguir. •

Definição 3.42 — Desvio-padrão

O desvio-padrão da v.a. X é a raiz quadrada positiva da variância de X:

$$DP(X) = +\sqrt{V(X)}, \qquad (3.21)$$

independentemente de X ser uma v.a. discreta ou contínua. •

Motivação 3.43 — Coeficiente de variação

Quer a variância, quer do desvio-padrão são medidas de dispersão absoluta que têm em conta a ordem de grandeza dos valores que X toma. E parece-nos óbvio que um desvio-padrão igual a 10 unidades tenha necessariamente significado distinto caso os valores de X sejam da ordem das dezenas ou da ordem das dezenas de milhar. Uma saída possível será optar pela medida de dispersão relativa que se define imediatamente a seguir. •

Definição 3.44 — Coeficiente de variação

O coeficiente de variação é igual a

$$CV(X) = \frac{DP(X)}{|E(X)|}, \tag{3.22}$$

quer X seja uma v.a. discreta, quer seja contínua, desde que $E(X) \neq 0$. •

Exemplo 3.45 — Coeficiente de variação de uma v.a. discreta

Obtenha o coeficiente de variação da facturação diária definida no Exemplo 3.40.

- **V.a.**

  $Y = 1200\,X =$ facturação diária.

- **Coeficiente de variação de Y**

$$\begin{aligned} CV(Y) &= \frac{DP(Y)}{|E(Y)|} \\ &= \frac{\sqrt{V(1200\,X)}}{|E(1200\,X)|} \\ &= \frac{\sqrt{1200^2 \times V(X)}}{1200 \times |E(X)|} \\ &= \frac{DP(X)}{|E(X)|} \\ &= CV(X) \\ &= \frac{\sqrt{2}}{2}. \end{aligned}$$

•

Este exemplo permite afirmar que o coeficiente de variação de um múltiplo de X é igual ao coeficiente de variação de X,

- $CV(aX) = CV(X)$, $a \neq 0$,

i.e., que o coeficiente de variação é invariante a dilatações e contracções de v.a.

Importa notar que esta propriedade não é satisfeita por qualquer das duas medidas de dispersão absoluta até agora consideradas, a variância e o desvio-padrão.

3.5 Distribuição uniforme discreta.

De salientar que a apresentação desta distribuição e das que se seguem respeitará um mesmo figurino: em primeiro lugar, alude-se à notação utilizada para representar a distribuição, de seguida, são referido(s), por esta ordem: o(s) parâmetro(s) da distribuição; o contradomínio da v.a. (i.e., os valores possíveis da v.a.); a função de probabilidade; o valor esperado; e a variância da v.a. Toda esta informação será condensada numa tabela.

Motivação 3.46 — **Distribuição uniforme discreta**
Esta distribuição é razoável quando a v.a. discreta toma n valores distintos, todos com a mesma probabilidade. Sem perda de generalidade considere-se que esta v.a. toma n valores distintos, $x_1, x_2, \ldots, x_n$, em que $x_1 < x_2 < \ldots < x_n$. •

Definição 3.47 — **Distribuição uniforme discreta**
A v.a. discreta X diz-se ter distribuição uniforme discreta no conjunto $\{x_1, x_2, \ldots, x_n\}$, caso a sua f.p. seja igual a

$$P(X = x) = \begin{cases} \frac{1}{n}, & x = x_1, x_2, \ldots, x_n \\ 0, & c.c. \end{cases} \tag{3.23}$$

Neste caso escreve-se

$$X \sim \text{uniforme discreta}(\{x_1, x_2, \ldots, x_n\}), \tag{3.24}$$

onde "$X \sim \ldots$" se lê, naturalmente, "X tem distribuição...".

Uniforme discreta	
Notação	$X \sim$ uniforme discreta$(\{x_1, x_2, \ldots, x_n\})$
Parâmetro	$\{x_1, x_2, \ldots, x_n\}$ $(x_i \in I\!R, i = 1, \ldots, n)$
Contradomínio	$\{x_1, x_2, \ldots, x_n\}$
F.p.	$P(X = x) = \begin{cases} \frac{1}{n}, & x = x_1, x_2, \ldots, x_n \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$
Variância	$V(X) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^2$

•

Exemplo 3.48 — Distribuição uniforme discreta

Um conjunto de n amostras de solo — das quais só uma está contaminada por uma perigosa substância química — chega a um laboratório. Admita ainda que a amostra de solo contaminado não foi etiquetada previamente. Considere agora a v.a. X que representa o número total de amostras inspeccionadas sem reposição até ser identificada a amostra de solo contaminado.

(a) Identifique a distribuição de X e desenhe o gráfico da f.p. desta v.a.

- **V.a.**

  X = número de amostras inspeccionadas sem reposição até à detecção da amostra contaminada

- **Contradomínio de** X

  $\mathbb{R}_X = \{1, 2, \ldots, n\}$

- **F.p. de** X

  $P(X = 1) = \frac{1}{n}$

  $P(X = 2) = \frac{n-1}{n} \times \frac{1}{n-1} = \frac{1}{n}$

  $P(X = 3) = \frac{n-1}{n} \times \frac{n-2}{n-1} \times \frac{1}{n-2} = \frac{1}{n}$

  $\vdots$

  $$P(X = x) = \begin{cases} \frac{1}{n}, & x = 1, 2, \ldots, n \\ 0, & c.c. \end{cases}$$

  De notar que se teve em consideração que a inspecção é feita sem reposição e que:

  - $X = 1$ se a 1a. amostra inspeccionada estiver contaminada;
  - $X = 2$ se a 1a. amostra inspeccionada não estiver contaminada mas a 2a. o estiver;

  e assim por diante.

- **Distribuição de** X

  $X \sim$ uniforme discreta$(\{1, 2, \ldots, n\})$

- **Gráfico da f.p. de** X

(b) Obtenha a f.d. de X e esboce o respectivo gráfico.

- **F.d. de X**

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= \begin{cases} 0, & x < 1 \\ \frac{1}{n}, & 1 \leq x < 2 \\ \frac{2}{n}, & 2 \leq x < 3 \\ \vdots & \\ \frac{n-1}{n}, & n-1 \leq x < n \\ 1, & x \geq n \end{cases} \\
&= \begin{cases} 0, & x < 1 \\ \frac{[x]}{n}, & 1 \leq x < n \\ 1, & x \geq n, \end{cases}
\end{aligned}
$$

onde $[x]$ representa a parte inteira do real x.

- **Gráfico da f.d. de X**

(c) Calcule o valor esperado e a variância desta v.a.

- **Nota**

Relembra-se para o efeito que:

$$
\begin{aligned}
\sum_{x=1}^{n} x &= \frac{n(n+1)}{2}; \\
\sum_{x=1}^{n} x^2 &= \frac{n(n+1)(2n+1)}{6}.
\end{aligned}
$$

- **Valor esperado de X**

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{n} x \times P(X = x) \\
&= \frac{\sum_{x=1}^{n} x}{n} \\
&= \frac{n(n+1)/2}{n} \\
&= \frac{n+1}{2}
\end{aligned}
$$

- **Variância de X**

$$
\begin{aligned}
V(X) &= E(X^2) - E^2(X) \\
&= \frac{\sum_{x=1}^{n} x^2}{n} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{n(n+1)(2n+1)/6}{n} - \frac{(n+1)^2}{4} \\
&= \frac{n+1}{2}\left(\frac{2n+1}{3} - \frac{n+1}{2}\right) \\
&= \frac{n+1}{2}\,\frac{4n+2-3n-3}{6} \\
&= \frac{n^2-1}{12}
\end{aligned}
$$

•

Dada a simplicidade da distribuição uniforme discreta a sua f.p., o seu valor esperado e a sua variância não se encontram no formulário disponível no final destas notas de apoio, ao contrário do que acontece com qualquer das distribuições apresentadas já de seguida.

3.6 Distribuição binomial.

Antes de passar à apresentação da distribuição binomial é necessário definir "prova de Bernoulli".[13]

Definição 3.49 — **Prova de Bernoulli**
Uma e.a. diz-se uma prova de Bernoulli se possuir apenas dois resultados possíveis:

- um sucesso, que ocorre com probabilidade p $(0 \leq p \leq 1)$;
- um insucesso, que ocorre com probabilidade $1-p$. •

Definição 3.50 — **Distribuição de Bernoulli**
A v.a.

- X = número de sucessos numa prova de Bernoulli

diz-se com distribuição de Bernoulli com parâmetro p e possui f.p. dada por

$$P(X=x)=\begin{cases} p^x\,(1-p)^{1-x}, & x=0,1 \\ 0, & c.c. \end{cases} \tag{3.25}$$

Bernoulli	
Notação	$X \sim \text{Bernoulli}(p)$
Parâmetro	$p = P(\text{sucesso})$ $(p \in [0,1])$
Contradomínio	$\{0,1\}$
F.p.	$P(X=x)=\begin{cases} p^x\,(1-p)^{1-x}, & x=0,1 \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X)=p$
Variância	$V(X)=p\,(1-p)$

•

Exercício 3.51 — Prove que $E(X) = p$ e $V(X) = p\,(1-p)$ quando $X \sim \text{Bernoulli}(p)$. •

[13] Jacques Bernoulli (1654–1705).

Motivação 3.52 — **Distribuição binomial**
A distribuição binomial é particularmente útil na caracterização probabilística do número de sucessos em n provas de Bernoulli realizadas de forma independente e com probabilidade de sucesso comum p. •

Definição 3.53 — **Distribuição binomial**
A v.a.

- X = número de sucessos num conjunto de n provas de Bernoulli independentes com probabilidade de sucesso comum e igual a p

diz-se com distribuição binomial de parâmetros (n, p) e possui f.p. igual a

$$P(X = x) = \begin{cases} \binom{n}{x}\, p^x\, (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & c.c., \end{cases} \tag{3.26}$$

onde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ representa as combinações de n elementos tomados x a x.

Binomial	
Notação	$X \sim \text{binomial}(n, p)$
Parâmetros	n =número de provas de Bernoulli ($n \in I\!\!N$)
	$p = P(\text{sucesso})$ ($p \in [0, 1]$)
Contradomínio	$\{0, 1, 2, \ldots, n\}$
F.p.	$P(X = x) = \begin{cases} \binom{n}{x}\, p^x\, (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = n\,p$
Variância	$V(X) = n\,p\,(1-p)$

•

A v.a. $X \sim \text{binomial}(n, p)$ está também associada à contagem do número de elementos com determinada característica (sucesso), num total de n elementos extraídos ao acaso e com reposição. Pode também ser entendida como a generalização natural da distribuição de Bernoulli.

Exercício 3.54 — Procure justificar a expressão da f.p. da v.a. com distribuição binomial(n, p) e demonstre que o valor esperado e a variância desta v.a. são iguais a $E(X) = n\,p$ e $V(X) = n\,p\,(1-p)$. •

Nota 3.55 — **F.d. da v.a. binomial**
A f.d. da v.a. $X \sim \text{binomial}(n,p)$ é dada por

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^{[x]} \binom{n}{i} p^i (1-p)^{n-i}, & 0 \leq x < n \\ 1, & x \geq n, \end{cases} \tag{3.27}$$

onde $[x]$ representa a parte inteira do real x.[14]

Esta função está tabelada para alguns pares de valores de n e p; refira-se que os valores de n e p não excedem (nas tabelas disponíveis para esta disciplina) 20 e 0.5, respectivamente. •

Exemplo 3.56 — **Utilização das tabelas da f.d. da v.a. binomial**

V.a.	x	Valor tabelado de $F_X(x) = P(X \leq x)$
$X \sim \text{binomial}(n = 5, p = 0.05)$	0	$F_X(0) = 0.7738$
$X \sim \text{binomial}(n = 9, p = 0.1)$	4	$F_X(4) = 0.9991$
$X \sim \text{binomial}(n = 10, p = 0.4)$	8	$F_X(8) = 0.9983$
$X \sim \text{binomial}(n = 20, p = 0.5)$	11	$F_X(11) = 0.7483$

•

Exemplo 3.57 — **Distribuição binomial**
A probabilidade de as leis de Kirchhoff virem a ser violadas, durante um teste laboratorial a que se submete certo tipo de indutor, é igual a 0.1.

Qual a probabilidade de esta lei vir a ser violada mais de 4 vezes em 9 destes testes laboratoriais?

- **V.a.**

  X = número de violações das leis de Kirchhoff em 9 testes laboratoriais

- **Distribuição de** X

  $X \sim \text{binomial}(n,p)$

- **Parâmetros**

  $n = 9$ testes

  $p = P(\text{violação leis Kirchhoff em teste laboratorial}) = 0.1$

[14]Relembre-se que a parte inteira do real x corresponde ao maior inteiro menor ou igual a x. Assim, $[0.3] = 0, [2.8] = 2, [-0.7] = -1$.

- **F.p. de X**

  $P(X = x) = \binom{9}{x}\, 0.1^x\, (1 - 0.1)^{9-x}$, $x = 0, 1, 2, \ldots, 9$

- **Probabilidade pedida**

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - F_X(4) \\ &= 1 - F_{binomial(9,0.1)}(4) \\ &\stackrel{tabela}{=} 1 - 0.9991 \\ &= 0.0009. \end{aligned}$$

•

Exercício 3.58 — Represente graficamente a f.d. da v.a. $X \sim \text{binomial}(4, 0.1)$, recorrendo às tabelas disponíveis, e obtenha a f.p. desta v.a. à custa dos valores obtidos para a f.d. •

Proposição 3.59 — **Distribuição binomial**
Seja X o número de sucessos em n provas de Bernoulli independentes com probabilidade de sucesso p, i.e., $X \sim \text{binomial}(n, p)$. Então o número de insucessos nessas mesmas n provas de Bernoulli, Y, verifica:

- $Y = n - X \sim \text{binomial}(n, 1 - p)$;
- $F_Y(y) = 1 - F_X(n - y - 1)$. •

Exercício 3.60 — Demonstre a segunda das propriedades da Proposição 3.59 e ilustre a sua utilização na obtenção de valores da f.d. de v.a. binomiais com probabilidade de sucesso superior a 0.5, fazendo uso das tabelas disponíveis.

- **V.a.**

  $X \sim \text{binomial}(n, p)$

- **Nova v.a.**

  $Y = n - X$

- **Demonstração da 2a. propriedade**

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(n - X \leq y) \\
&= P(X \geq n - y) \\
&= 1 - P(X \leq n - y - 1) \\
&= 1 - F_X(n - y - 1)
\end{aligned}
$$

- **Ilustração da 2a. propriedade**

V.a.	y	$F_Y(y) = P(X \leq n - y - 1)$
$Y \sim \text{bin}(5, 0.95)$	4	$F_Y(4) = 1 - F_{bin(5,1-0.95)}(5 - 4 - 1) = 1 - 0.7738 = 0.2262$
$Y \sim \text{bin}(10, 0.6)$	1	$F_Y(1) = 1 - F_{bin(10,1-0.6)}(10 - 1 - 1) = 1 - 0.9983 = 0.0017$

•

3.7 Distribuição geométrica.

Motivação 3.61 — Distribuição geométrica

A distribuição binomial(n, p) está associada à contagem do número de sucessos em n provas de Bernoulli independentes que possuem em qualquer dos casos probabilidade de sucesso igual a p. Caso estejamos interessados em contabilizar o

- o número total de provas de Bernoulli realizadas até ao registo do primeiro sucesso,

passamos a lidar com uma v.a. discreta com distribuição distinta da binomial. •

Definição 3.62 — Distribuição geométrica

Seja

- X = número de provas de Bernoulli (independentes com probabilidade de sucesso comum e igual a p) realizadas até à ocorrência do primeiro sucesso.

Então a v.a. X diz-se com distribuição geométrica com parâmetro p e a sua f.p. é dada por

$$P(X = x) = \begin{cases} (1-p)^{x-1}\, p, & x = 1, 2, 3, \ldots \\ 0, & c.c. \end{cases} \tag{3.28}$$

Geométrica	
Notação	$X \sim \text{geométrica}(p)$
Parâmetro	$p = P(\text{sucesso})$ $(p \in [0, 1])$
Contradomínio	$I\!N = \{1, 2, 3, \ldots\}$
F.p.	$P(X = x) = \begin{cases} (1-p)^{x-1}\, p, & x = 1, 2, 3, \ldots \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = \frac{1}{p}$
Variância	$V(X) = \frac{1-p}{p^2}$

•

A distribuição geométrica é por vezes designada por distribuição discreta do tempo de espera pelo primeiro sucesso.

Nota 3.63 — **F.d. da v.a. geométrica**
A f.d. da v.a. $X \sim$ geométrica(p) não está tabelada pois obtém-se sem grande dificuldade, por estar a lidar-se com uma série geométrica. Com efeito,

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \sum_{i=1}^{[x]} (1-p)^{i-1}\, p = 1 - (1-p)^{[x]}, & x \geq 1, \end{cases} \tag{3.29}$$

onde $[x]$ representa novamente a parte inteira do real x. •

Exercício 3.64 — Justifique as expressões da f.p. e da f.d. desta v.a. Obtenha também o seu valor esperado e a sua variância. •

Exemplo 3.65 — **Distribuição geométrica**
Estudos preliminares indicaram que a probabilidade de ser detectada a presença de alto teor de metais pesados numa amostra de solo proveniente de certo local é de 0.01.[15]

(a) Obtenha o valor esperado do número total de amostras seleccionadas ao acaso até que seja detectada a primeira amostra de solo com alto teor de metais pesados.

- **V.a.**
  $X =$ número total de amostras seleccionadas até que seja detectada a primeira amostra de solo com alto teor de metais pesados
- **Distribuição de** X
  $X \sim$ geométrica(p)
- **Parâmetro**
  $p = 0.01$
- **F.p. de** X
  $P(X = x) = (1 - 0.01)^{x-1} \times 0.01$, $x = 1, 2, 3, \ldots$
- **Valor esperado de** X

$$\begin{aligned} E(X) &\stackrel{form.}{=} \frac{1}{p} \\ &= \frac{1}{0.01} \\ &= 100 \text{ amostras.} \end{aligned}$$

[15] Adaptado do Exame de 2a. Época, 4 de Fevereiro de 2003.

(b) Determine a probabilidade de serem inspeccionadas mais de $100+50$ amostras até à detecção da primeira amostra de solo com alto teor de metais pesados, sabendo que já foram inspeccionadas mais de 100 amostras sem que semelhante detecção tivesse ocorrido.

- **Probabilidade pedida**

  Tirando partido da expressão geral da f.d. da v.a. geométrica tem-se sucessivamente:

$$
\begin{aligned}
P(X > 100+50 | X > 100) &= \frac{P(X > 100+50, X > 100)}{P(X > 100)} \\
&= \frac{P(X > 100+50)}{P(X > 100)} \\
&= \frac{1 - P(X \leq 100+50)}{1 - P(X \leq 100)} \\
&= \frac{1 - [1 - (1-0.01)^{100+50}]}{1 - [1 - (1-0.01)^{100}]} \\
&= \frac{(1-0.01)^{100+50}}{(1-0.01)^{100}} \\
&= (1-0.01)^{50} \\
&= P(X > 50).
\end{aligned}
$$

  Este resultado deve-se a uma propriedade desta distribuição que será enunciada de seguida. •

Proposição 3.66 — Falta de memória da distribuição geométrica
Seja $X \sim$ geométrica(p). Então

$$P(X > k + x | X > k) = P(X > x), \forall k, x \in I\!N, \tag{3.30}$$

Equivalentemente, a v.a. $X - k | X > k$, que representa o número de provas de Bernoulli adicionais sabendo que já foram efectuadas mais de k provas, também possui distribuição geométrica com parâmetro p:

$$X - k | X > k \sim \text{geométrica}(p), \forall k \in I\!N. \tag{3.31}$$

•

Esta propriedade é denominada de "falta de memória" uma vez que (3.31) sugere um recomeço probabilístico.

3.8 Distribuição hipergeométrica.

Motivação 3.67 — **Distribuição hipergeométrica**
A distribuição binomial(n,p) está associada à contagem do número de sucessos, em n extracções ao acaso com reposição. Ao considerar-se um

- processo de extracção casual sem reposição,

passamos a lidar com uma v.a. discreta com distribuição distinta da binomial. •

Definição 3.68 — **Distribuição hipergeométrica**
Considerem-se:

- N = número total de elementos de uma população (dimensão da pop.);
- M = número de elementos dessa população que possuem certa característica (sucesso);
- n = número de extracções sem reposição.

Então a v.a.

- X = número de elementos com certa característica (sucesso), em n extraídos ao acaso sem reposição da população acima

diz-se com distribuição hipergeométrica com parâmetros (N,M,n) e a sua f.p. pode encontrar-se na tabela abaixo a par de outras características desta distribuição.

Hipergeométrica	
Notação	$X \sim \text{hipergeométrica}(N,M,n)$
Parâmetros	N $(N \in I\!N)$ M $(M \in I\!N, M \leq N)$ n $(n \in I\!N, n \leq N)$
Contradomínio	$\{\max\{0, n-(N-M)\}, \ldots, \min\{n,M\}\}$
F.p.	$P(X=x) = \begin{cases} \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}, \\ \qquad x = \max\{0, n-(N-M)\}, \ldots, \min\{n,M\} \\ 0, \ c.c. \end{cases}$
Valor esperado	$E(X) = n\,\frac{M}{N}$
Variância	$V(X) = n\,\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$

•

Nota 3.69 — Distribuição hipergeométrica
X corresponde ao número de sucessos num conjunto de n provas de Bernoulli DEPENDENTES com probabilidade de sucesso comum e igual a $p = M/N$.

$E(X)$ e $V(X)$ fazem lembrar o valor esperado e a variância da distribuição binomial (n, p) com factores de correcção que se devem à não reposição das extracções. Com efeito, sob certas condições, a (f.p./f.d. da) v.a. hipergeométrica(N, M, n) pode ser aproximada pela (f.p./f.d. d)a v.a. binomial$(n, \frac{M}{N})$, como veremos mais tarde. •

Exemplo 3.70 — Distribuição hipergeométrica
Justifique a expressão geral da f.p., bem como o contradomínio da v.a. hipergeométrica.

Na fase de concepção de um sistema de controlo de qualidade do fabrico, foram escolhidos 100 cabos dos quais apenas dois apresentavam desvios superiores a 9.8 mícrons. Se desses 100 cabos forem seleccionados 10 ao acaso e sem reposição, qual é a probabilidade de mais do que um ter um desvio superior a 9.8 mícrons? Indique também o valor esperado do número de cabos, entre esses 10, com um desvio superior a 9.8 mícrons.[16]

- **V.a.**

  X = número de cabos com um desvio superior a 9.8 mícrons, em 10 cabos seleccionados sem reposição (de lote de 100 cabos dos quais apenas dois apresentam desvios superiores a 9.8 mícrons)

- **Distribuição de X**

  $X \sim$ hipergeométrica(N, M, n)

- **Parâmetros**

  $N = 100$

  $M = 2$

  $n = 10$

- **F.p. de X**

  $P(X = x) = \frac{\binom{2}{x}\binom{100-2}{10-x}}{\binom{100}{10}}, \ x = 0, 1, 2$

[16] Adaptado do Teste A, 11 de Novembro de 2006.

- **Probabilidade pedida**

$$\begin{aligned} P(X > 1) &= P(X = 2) \\ &= \frac{\binom{2}{2}\binom{100-2}{10-2}}{\binom{100}{10}} \\ &= \frac{1}{110} \end{aligned}$$

- **Valor esperado de X**

$$\begin{aligned} E(X) &\overset{form.}{=} n \times \frac{M}{N} \\ &= 10 \times \frac{2}{100} \\ &= 0.2 \end{aligned}$$

-

3.9 Distribuição de Poisson.

Motivação 3.71 — Distribuição de Poisson
A distribuição de Poisson é frequentemente usada na contagem de ocorrências de certo tipo de evento em períodos fixos de tempo,[17] eventos tais como: chegadas, partidas, acidentes, falhas de equipamento, testemunhos verdadeiros em tribunal, número de excedências de níveis elevados de pluviosidade/ondas/marés, número de colisões de detritos espaciais com diâmetro superior a $1cm$ num satélite numa região orbital abaixo dos $2.000Km$ de altitude, etc.

A distribuição de Poisson foi originalmente introduzida em 1837 por Siméon Dennis Poisson (1781–1840) como distribuição limite da distribuição binomial. Anos mais tarde von Bortkiewicz (1898) recorre à distribuição de Poisson para descrever o comportamento probabilístico do número de mortes por coices de cavalo no exército prussiano. •

Definição 3.72 — Distribuição de Poisson
A v.a. X com distribuição tem a particularidade de possuir o valor esperado e a variância iguais ao parâmetro que define a distribuição, λ, e f.p. na tabela abaixo

Poisson	
Notação	$X \sim \text{Poisson}(\lambda)$
Parâmetro	λ ($\lambda \in \mathbb{R}^+$)
Contradomínio	$\mathbb{N}_0$
F.p.	$P(X = x) = \begin{cases} e^{-\lambda}\,\frac{\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = \lambda$
Variância	$V(X) = \lambda$

•

Nota 3.73 — F.d. da v.a. de Poisson
A f.d. da v.a. $X \sim \text{Poisson}(p)$,

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^{[x]} e^{-\lambda}\,\frac{\lambda^i}{i!}, & x \geq 0 \end{cases} \tag{3.32}$$

(onde $[x]$ é a parte inteira do real x), está tabelada para alguns valores de λ. •

[17] Ou mesmo em áreas ou volumes.

Exemplo 3.74 — **Utilização das tabelas da f.d. da v.a. de Poisson**

V.a.	x	Valor tabelado de $F_X(x) = P(X \leq x)$
$X \sim \text{Poisson}(\lambda = 0.05)$	0	$F_X(0) = 0.9512$
$X \sim \text{Poisson}(\lambda = 3)$	1	$F_X(1) = 0.1991$
$X \sim \text{Poisson}(\lambda = 12)$	1	$F_X(1) = 0.0001$
$X \sim \text{Poisson}(\lambda = 20)$	14	$F_X(14) = 0.1049$

•

Exemplo 3.75 — **Distribuição de Poisson**

A procura semanal de uma luxuosa marca de automóvel segue uma lei de Poisson. Sabe-se ainda que a probabilidade de numa semana não existir procura é igual a e^{-3}.[18]

(a) Determine a probabilidade de a procura semanal exceder pelo menos 2 automóveis.

- **V.a.**

  $X =$ procura semanal de automóveis da referida marca

- **Distribuição de X**

  $X \sim \text{Poisson}(\lambda)$

- **F.p. de X**

  $P(X = x) = e^{-\lambda}\,\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots$

- **Parâmetro**

$$\begin{aligned}\lambda \quad : \quad & P(X = 0) = e^{-3} \\ & e^{-\lambda}\,\frac{\lambda^0}{0!} = e^{-3} \\ & \lambda = 3\end{aligned}$$

- **Probabilidade pedida**

$$\begin{aligned}P(X \geq 2) \quad & = \quad 1 - P(X < 2) \\ & = \quad 1 - P(X \leq 1) \\ & = \quad 1 - F_{Poisson(3)}(1) \\ & \stackrel{tabela}{=} \quad 1 - 0.1991 \\ & = \quad 0.8009.\end{aligned}$$

[18] Adaptado do Exame de 2a. Época, 21 de Julho de 2001.

(b) Qual a probabilidade de a procura em 4 semanas ser de pelo menos 2 automóveis?

- **V.a.**

  $Y =$ procura de automóveis da referida marca em 4 semanas

- **Distribuição de** Y

  $Y \sim \text{Poisson}(4 \times \lambda)$

- **F.p. de** Y

  $P(Y = y) = e^{-12}\,\frac{12^y}{y!},\ y = 0, 1, 2, \ldots$

- **Probabilidade pedida**

$$\begin{aligned} P(Y \geq 2) &= 1 - F_{Poisson(12)}(1) \\ &\overset{tabela}{=} 1 - 0.0001 \\ &= 0.9999. \end{aligned}$$

  Esta alínea ilustra a propriedade reprodutiva da v.a. de Poisson (ver Cap. 5).

-

3.10 Algumas notas sobre análise combinatória

- **Permutações de n elementos** $(n!)$

  Número de formas distintas de preenchimento de caixa com n compartimentos, dispondo de n elementos e não havendo a possibilidade de repetição no preenchimento.

  Sequências de n elementos distintos...

- **Arranjos com repetição de n elementos tomados x a x** (n^x)

  Número de formas distintas de preenchimento de caixa com x compartimentos, dispondo de n elementos e havendo a possibilidade de repetição no preenchimento.

  Sequências de x elementos...

- **Arranjos sem repetição de n elementos tomados x a x** $\left(\frac{n!}{(n-x)!}\right)$

  Número de formas distintas de preenchimento de caixa com x compartimentos, dispondo de n elementos e não havendo a possibilidade de repetição no preenchimento.

  Sequências de x elementos distintos...

- **Combinações de n elementos tomados x a x** $\left(\binom{n}{x} = \frac{n!}{x!(n-x)!}\right)$

  Número de conjuntos de cardinal x (logo com elementos distintos) que podem ser formados com n elementos.

  Conjuntos de cardinal x...

- **Permutações de n elementos de k tipos distintos** $\left(\frac{n!}{n_1!\, n_2! \ldots n_k!}\right)$

  Número de formas distintas de preenchimento de caixa com n compartimentos, dispondo de n elementos de k tipos distintos, onde n_i representa o número de elementos do tipo i, $i = 1, 2, \ldots, k$, e $\sum_{i=1}^{k} n_i = n$.

  Sequências de n elementos de k tipos...

- **Binómio de Newton**

  $(a+b)^n = \sum_{x=1}^{n} \binom{n}{x} a^x\, b^{n-x}$.

Capítulo 4

Variáveis aleatórias e distribuições contínuas

4.1 Variáveis aleatórias contínuas.

Motivação 4.1 — V.a. contínua
Muitas quantidades de interesse são expressas por valores numéricos. São disso exemplo

- a vibração produzida por um motor (em hertz por unidade de tempo), a deflexão de uma mola (em metros), o tempo de reparação de peça mecânica,
- a intensidade da corrente eléctrica em certa zona de um circuito (em amperes), a impedância de um sistema (em ohm),
- a temperatura, o volume da voz, a concentração de um poluente, etc.

Em qualquer dos casos é perfeitamente razoável assumir que

- o conjunto de valores possíveis é infinito não numerável,

por exemplo, um intervalo real $[a, b]$, ou $I\!R$, ou $I\!R^+$. •

O facto de o contradomínio da v.a. X ser infinito não numerável é manifestamente insuficiente para descrever rigorosamente uma v.a. contínua como veremos já de seguida.

Definição 4.2 — V.a. contínua
A v.a. X diz-se contínua, caso

- possua f.d. $F_X(x) = P(X \leq x)$ contínua em $I\!R$

e exista uma função real de variável real, $f_X(x)$, que verifique:

- $f_X(x) \geq 0, \forall x \in I\!R$
- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$. •

A função $f_X(x)$ é denominada de função densidade de probabilidade e goza de um conjunto de propriedades descritas na secção seguinte.

4.2 Função de densidade de probabilidade.

Motivação 4.3 — **Função de densidade de probabilidade**
Ao lidarmos com uma v.a. discreta podemos calcular a f.p. $P(X = x)$. No entanto, tal cálculo não faz qualquer sentido ao lidar-se com a v.a. contínua X. É razoável sim

- calcular a probabilidade de X pertencer a um intervalo

e definir o análogo contínua da f.p., a função de densidade de probabilidade (f.d.p.). •

Proposição 4.4 — **Função de densidade de probabilidade**
Tendo em conta as propriedades da f.d. e a relação entre esta função e a f.d.p., deduzem-se as seguintes propriedades para $f_X(x)$. Para além de verificar

- $f_X(x) \geq 0, \forall x \in \mathbb{R}$
- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$,

a f.d.p. satisfaz

- $\int_{-\infty}^{+\infty} f_X(x)dx = 1$
- $\int_a^b f_X(x)dx = P(a < X \leq b)$, $\forall a < b$. •

Nota 4.5 — **F.d.p.**
Para v.a. contínuas tem-se:

- $P(X = x) = 0$, $\forall x \in \mathbb{R}$, i.e., a probabilidade de a v.a. tomar o valor real x é nula (o evento é quase-impossível!)
- $P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b)$ $= \int_a^b f_X(x)dx = F_X(b) - F_X(a)$, $\forall a < b$,

  correspondendo à área sob o gráfico da f.d.p. entre a e b (gráfico!). •

Nota 4.6 — **Interpretação geométrica da f.d.p.**
De notar que

$$\begin{aligned} P\left(a - \frac{\epsilon}{2} < X \leq a + \frac{\epsilon}{2}\right) &= \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_X(x)\,dx \\ &\simeq \epsilon \times f_X(a). \qquad (4.1) \end{aligned}$$

$f_X(a)$ dá, portanto, uma "ideia" da "possibilidade" de ocorrência de valores próximos do ponto a.[1] •

[1]Que nunca deve confundir-se com a probabilidade de ocorrer $\{X = a\}$, probabilidade que se sabe ser nula para v.a. contínuas.

4.3 Função de distribuição.

Motivação 4.7 — F.d. de v.a. contínua
Tal como no caso discreto justifica-se o cálculo de probabilidades do tipo $P(X \leq x)$, para qualquer $x \in I\!R$, ou seja, é pertinente definir a f.d. da v.a. contínua X. •

Definição 4.8 — F.d. de v.a. contínua
Seja X uma v.a. contínua. Então a sua f.d. é dada por

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(t)dt, \; x \in I\!R \\ &= \text{Área sob o gráfico da f.d.p. de } X \text{ entre } -\infty \text{ e } x. \end{aligned} \tag{4.2}$$

•

Exemplo 4.9 — F.d. de v.a. contínua
Assuma que o tempo (em anos) entre duas colisões consecutivas de detritos espaciais com diâmetro maior que $1mm$ num satélite em MEO[2] é uma v.a. contínua com f.d.p.

$$f_X(x) = \begin{cases} 0, & x < 0 \\ 0.4\, e^{-0.4x}, & x \geq 0. \end{cases} \tag{4.3}$$

(a) Após ter feito o gráfico da f.d.p. de X, calcule a probabilidade do tempo entre duas colisões consecutivas exceder 1 ano e três meses.

- **V.a.**
  X = tempo entre duas colisões consecutivas... (em anos)
- **F.d.p. de** X

$$f_X(x) \quad = \quad \begin{cases} 0, & x < 0 \\ 0.4\, e^{-0.4x}, & x \geq 0. \end{cases}$$

- **Gráfico da f.d.p. de** X

[2] *Medium Earth Orbit*, associada a altitudes entre $2.000Km$ e $34.786Km$.

- **Probabilidade pedida**

$$\begin{aligned} P(X > 1.25) &= \int_{1.25}^{+\infty} f_X(x)\,dx \\ &= \int_{1.25}^{+\infty} 0.4\,e^{-0.4x}\,dx \\ &= -e^{-0.4x}\Big|_{1.25}^{+\infty} \\ &= -0 + e^{-0.4\times 1.25} \\ &= e^{-0.5}. \end{aligned}$$

(b) Determine a f.d. de X e esboce o respectivo gráfico.

- **F.d. de** X

Tal como no caso discreto talvez não seja má ideia começar-se por preencher a tabela abaixo com alguns valores da f.d. de X para depois determinar-se a expressão geral da f.d. de X.

x	$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$	Esquema
-1.5	$F_X(-1.5) = \int_{-\infty}^{-1.5} f_X(t)dt$ $= \int_{-\infty}^{-1.5} 0\,dt$ $= 0$	
1.25	$F_X(1.25) = \int_{-\infty}^{1.25} f_X(t)dt$ $= \int_{-\infty}^{0} 0\,dt + \int_{0}^{1.25} 0.4\,e^{-0.4t}dt$ $= -e^{-0.4t}\big\|_{0}^{1.25}$ $= 1 - e^{-0.5}$	
$x > 0$	$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ $= \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 0.4\,e^{-0.4t}dt$ $= -e^{-0.4t}\big\|_{0}^{x}$ $= 1 - e^{-0.4\,x}$	

Deste modo conclui-se que:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-0.4x}, & x \geq 0. \end{cases}$$

- **Gráfico da f.d. de** X

•

Proposição 4.10 — Propriedades da f.d. de v.a. contínua
A f.d. da v.a. contínua X, $F_X(x)$, é uma:

1. Função contínua, logo contínua quer à direita,[3] quer à esquerda, ou seja, $F_X(x) = F_X(x^+) = F_X(x^-)$, $\forall x \in \mathbb{R}$;

2. Função monótona não decrescente de x.

A f.d. de uma v.a contínua X verifica também:

3. $0 \leq F_X(x) \leq 1$;

4. $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$;

5. $F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$;

6. $P(X \leq x) = P(X < x) = F_X(x)$;

7. $P(X > x) = 1 - F_X(x)$;

e ainda, para $a < b$,

8. $P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$, $\forall a < b$,

como, aliás, já se tinha referido. •

Nota 4.11 — Relação entre a f.d.p. e a f.d.
É possível obter a f.d.p. derivando a f.d.:

$$f_X(x) = \frac{d\,F_X(x)}{dx}. \tag{4.4}$$

•

[3]Tal como acontece com a f.d. de qualquer v.a. discreta.

Exemplo 4.12 — **F.d. de v.a. contínua** *(e não só)*

O tempo que uma viatura de uma luxuosa marca leva a atingir 100 Km/h (em segundos) é uma variável aleatória X com f.d.p.[4]

$$f_X(x) = \begin{cases} \frac{2\times 4.5^2}{x^3}, & x \geq 4.5 \\ 0, & \text{c.c.} \end{cases}$$

(a) Determine a probabilidade de uma viatura seleccionada ao acaso atingir 100 Km/h em mais de 7 segundos e como tal necessitar de uma afinação.

- **V.a.**

  $X =$ tempo (em segundos) que uma viatura leva a atingir $100Km/h$

- **F.d.p. de** X

$$f_X(x) = \begin{cases} \frac{2\times 4.5^2}{x^3}, & x \geq 4.5 \\ 0, & \text{c.c.} \end{cases}$$

- **Probabilidade pedida**

$$\begin{aligned} P(X > 7) &= \int_7^{+\infty} f_X(x)\,dx \\ &= \int_7^{+\infty} \frac{2\times 4.5^2}{x^3}\,dx \\ &= -\left.\frac{4.5^2}{x^2}\right|_7^{+\infty} \\ &= \frac{4.5^2}{49} \\ &\simeq 0.4133. \end{aligned}$$

(b) Qual a probabilidade de pelo menos duas viaturas necessitarem de afinação, de entre um conjunto de 10 seleccionadas ao acaso da produção diária?

- **Nova v.a.**

  $Y =$ número de viaturas que necessitam de afinação, em 10 seleccionadas ao acaso[5]

- **Distribuição de** Y

  $Y \sim \text{binomial}(n, p)$

[4]Adaptado do Teste A, 22 de Maio de 2006.

[5]Admitindo independência entre os tempos a as diferentes viaturas atingem os $100Km/h$.

- **Parâmetros**

  n =10 viaturas

  $p = P(\text{viatura necessitar de afinação}) = P(X > 7) \simeq 0.4133$

- **F.p. de** Y

  $P(Y = y) = \binom{10}{y}\, 0.4133^y\, (1 - 0.4133)^{10-y}$, $y = 0, 1, 2, \ldots, 10$

- **Probabilidade pedida**

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y \leq 1) \\ &= 1 - \sum_{y=0}^{1} \binom{10}{y} 0.4133^y\, (1 - 0.4133)^{10-y} \\ &= 1 - \left[(1 - 0.4133)^{10} + 10 \times 0.4133 \times (1 - 0.4133)^9\right]) \\ &= 0.9612. \end{aligned}$$

  Uma vez que a probabilidade obtida é elevadíssima, alerta-se para a necessidade de uma melhoria do processo de produção, de modo a diminuir o número esperado de viaturas que necessitam de afinação $E(Y) = n\,p = 10 \times 0.4133 = 4.133$ viaturas. •

4.4 Valor esperado, variância e algumas das suas propriedades. Moda e quantis.

Motivação 4.13 — Parâmetros
Tal como no caso das v.a. discretas é importante calcular medidas sumárias de

- Localização central
- Localização não central
- Dispersão

capazes de caracterizar — embora parcialmente — algum aspecto de uma v.a. X contínua. A definição de qualquer destes parâmetros é análoga ao caso discreto: onde tínhamos

- $\sum_{x \in I\!R}$ ou $P(X = x)$

passamos a ter

- $\int_{-\infty}^{+\infty} \; dx$ ou $f_X(x)$, respectivamente. •

Ao invocar esta analogia escusar-nos-emos a fazer qualquer tipo de motivação adicional a estes parâmetros ou repetir comentários já efectuados sobre estes parâmetros no capítulo anterior.

Definição 4.14 — Valor esperado de v.a. contínua
O valor esperado da v.a. contínua X é igual a

$$E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx. \tag{4.5}$$

•

Nota 4.15 — Valor esperado de v.a. contínua

1. $E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_{x \in I\!R_X} x\, f_X(x)\, dx$.
2. $E(X)$ existe sse $\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx$, i.e., sse o integral for absolutamente convergente.
3. Ao lidarmos com v.a. contínuas $E(X) \in I\!R_X$, i.e., o valor esperado de X pertence, de um modo geral, ao conjunto de valores possíveis de X.[6] •

[6] Há eventuais excepções: basta pensar numa v.a. contínua que tome valores nos intervalos disjuntos $[0, 10]$ e $[20, 30]$ e possua f.d.p. constante e igual a $\frac{1}{20}$ em qualquer dos dois intervalos.

Proposição 4.16 — Propriedades do valor esperado de v.a. contínua
Tal como no caso discreto, o valor esperado da v.a. contínua X goza das duas seguintes propriedades:

1. $E(b) = b, \forall b \in \mathbb{R}$;

2. $E(aX + b) = aE(X) + b, \forall a, b \in \mathbb{R}$ (operador linear).

E para além disso satisfaz as propriedades:

3. Seja $Y = \psi(X)$ uma v.a. função (mensurável) de X. Então
$$E(Y) = E[\psi(X)] = \int_{-\infty}^{+\infty} \psi(x)\, f_X(x)\, dx. \tag{4.6}$$

4. Seja X uma v.a. real não negativa (resp. positiva), i.e., $\mathbb{R}_X = \mathbb{R}_0^+ = [0, +\infty)$ (resp. $\mathbb{R}_X = \mathbb{R}^+ = (0, +\infty)$). Então
$$E(X) = \int_0^{+\infty} P(X > x)\, dx = \int_0^{+\infty} [1 - F_X(x)]\, dx. \tag{4.7}$$

•

Exemplo 4.17 — Valor esperado de v.a. contínua
Calcule o valor esperado do tempo entre colisões consecutivas, definido no Exemplo 4.9.

- **V.a.**

  X =tempo (em anos) entre duas colisões consecutivas...

- **F.d.p. de X**

$$f_X(x) = \begin{cases} 0, & x < 0 \\ 0.4\, e^{-0.4x}, & x \geq 0. \end{cases}$$

- **Valor esperado de X**

  Integrando por partes[7] obtém-se:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \times f_X(x)\, dx \\ &= \int_{-\infty}^{0} 0\, dx + \int_0^{+\infty} x \times 0.4\, e^{-0.4x}\, dx \end{aligned}$$

---

[7] $\begin{cases} u = x \\ v' = 0.4\, e^{-0.4x} \end{cases} \quad \begin{cases} u' = 1 \\ v = -e^{-0.4x} \end{cases}$

$$
\begin{aligned}
&= -x \times e^{-0.4x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-0.4x}\,dx \\
&= 0 - \left.\frac{e^{-0.4x}}{0.4}\right|_0^{+\infty} \\
&= \frac{1}{0.4} \\
&= 2.5 \text{ anos.}
\end{aligned}
$$

•

Definição 4.18 — Moda de v.a. contínua

A moda da v.a. contínua X, mo, está associada ao ponto de máximo da f.d.p. de X, ou seja,

$$
mo: \; f_X(mo) = \max_x f_X(x). \tag{4.8}
$$

•

Nota 4.19 — Moda de v.a. contínua

De referir que:

- tal como no caso das v.a. discretas, a moda de uma v.a. contínua pode não ser única como é o caso das v.a. definidas em intervalos finitos e com f.d.p. constante nesses mesmos intervalos;

- a moda de uma v.a. contínua obtém-se recorrendo de um modo geral às técnicas de maximização de funções:

$$
mo: \begin{cases} \left.\frac{df_X(x)}{dx}\right|_{x=mo} = 0 \text{ (ponto de estacionaridade)} \\ \left.\frac{d^2 f_X(x)}{dx^2}\right|_{x=mo} < 0 \text{ (ponto de máximo).} \end{cases} \tag{4.9}
$$

•

Definição 4.20 — Mediana de v.a. contínua

Tal como acontece no caso discreto, a mediana da v.a. contínua X, me, verifica a dupla desigualdade $\frac{1}{2} \leq F_X(me) \leq \frac{1}{2} + P(X = me)$. Mas como $P(X = me) = 0$ no caso contínuo a mediana é definida por

$$
me: \; F_X(me) = \frac{1}{2}. \tag{4.10}
$$

•

Definição 4.21 — Quantil de probabilidade p de v.a. contínua
Analogamente, o quantil de probabilidade (ou quantil de ordem) p $(0 < p < 1)$ da v.a. contínua X, χ_p, define-se à custa da equação

$$\chi_p : F_X(\chi_p) = p. \tag{4.11}$$

•

Exemplo 4.22 — Moda, mediana e terceiro quartil de v.a. contínua
Retome o Exemplo 4.9 e obtenha a moda, a mediana e o terceiro quartil do tempo entre colisões consecutivas de detritos espaciais num satélite em MEO.

- **V.a.**

  X = tempo entre duas colisões consecutivas... (em anos)

- **F.d.p. de X**

$$f_X(x) \;=\; \begin{cases} 0, & x < 0 \\ 0.4\, e^{-0.4x}, & x \geq 0. \end{cases}$$

- **Moda de X**

  $mo = mo(X) = 0$ já que $f_X(x)$ é uma função decrescente em $\mathbb{R}_0^+$.

- **Mediana de X**

  Tirando partido do facto da f.d. de X ser igual a

$$F_X(x) \;=\; \begin{cases} 0, & x < 0 \\ 1 - e^{-0.4x}, & x \geq 0, \end{cases}$$

  conclui-se que a mediana da v.a. X é dada por

$$\begin{aligned} me: \quad F_X(me) &= 1/2 \\ 1 - e^{-0.4\,me} &= 1/2 \\ me &= -\tfrac{1}{0.4}\ln(1 - 1/2) \\ &\simeq 1.73 \text{ anos} \end{aligned}$$

  e pode afirmar-se que se calcula que metade dos tempos entre colisões consecutivas não excederá 1.73 anos.

- **Terceiro quartil de X**

  Designe-se este quartil por $F_X^{-1}(3/4)$. Então

  $$\begin{array}{rrcl} F_X^{-1}(3/4): & F_X[F_X^{-1}(3/4)] & = & 3/4 \\ & F_X^{-1}(3/4) & = & -\frac{1}{0.4}\ln(1-3/4) \\ & & \simeq & 3.47 \text{ anos} \end{array}$$

  e note-se que $P[X \leq F_X^{-1}(3/4)] = \mathbf{3/4}$ e $P[X \geq F_X^{-1}(3/4)] = \mathbf{1/4}$. •

Definição 4.23 — **Variância de v.a. contínua**
A variância da v.a. contínua X é igual a

$$V(X) = E(X^2) - E^2(X), \tag{4.12}$$

onde $E(X^2) = \int_{-\infty}^{+\infty} x^2\, f_X(x)\, dx$ e $E^2(X) = [\int_{-\infty}^{+\infty} x\, f_X(x)\, dx]^2$. •

Nota 4.24 — **Propriedades da variância**
Relembre-se que a variância de uma v.a. quer discreta, quer contínua, goza das propriedades:

1. $V(b) = 0, \forall b \in \mathbb{R}$
2. $V(X) \geq 0$, qualquer que seja a v.a. X
3. $V(aX+b) = a^2 V(X), \forall a, b \in \mathbb{R}$. •

Definição 4.25 — **Desvio-padrão de v.a. contínua**
Escusado será dizer que o desvio-padrão da v.a. contínua X é a raiz quadrada positiva da variância de X:

$$DP(X) = +\sqrt{V(X)}. \tag{4.13}$$

•

Definição 4.26 — **Coeficiente de variação**
Relembre-se também que o coeficiente de dispersão é igual a

$$CV(X) = \frac{DP(X)}{|E(X)|}, \tag{4.14}$$

caso X seja uma v.a. contínua, desde que $E(X) \neq 0$. •

Exercício 4.27 — Retome o Exercício 4.9 e prove que a variância, o desvio-padrão e o coeficiente de variação do tempo entre colisões consecutivas são iguais a $1/0.4^2$, $1/0.4$ e 1 respectivamente. •

4.5 Distribuição uniforme contínua.

Motivação 4.28 — Distribuição uniforme contínua
Esta distribuição é o análogo contínuo da distribuição uniforme discreta, tal como sugere o seu nome. Não surpreende pois que se trate de distribuição adequada a descrever o comportamento probabilístico de v.a. cujos valores possíveis se crê terem todos o mesmo "peso", para além de constituirem um conjunto conhecido e limitado (quer à esquerda, quer à direita). •

Definição 4.29 — Distribuição uniforme contínua
A v.a. contínua X possui distribuição uniforme contínua no intervalo $[a, b]$ (onde $a < b$), caso a sua f.d.p. seja dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & c.c. \end{cases} \tag{4.15}$$

Uniforme	
Notação	$X \sim \text{uniforme}(a, b)$
Parâmetros	a (extremo inferior do intervalo; $a \in I\!R$) b (extremo superior do intervalo; $b \in I\!R$)
Contradomínio	$[a, b]$
F.d.p.	$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = \frac{a+b}{2}$
Variância	$V(X) = \frac{(b-a)^2}{12}$

•

Exercício 4.30 — Distribuição uniforme contínua
Esboce o gráfico da f.d.p. da v.a. $X \sim \text{uniforme}(a, b)$ e deduza o seu valor esperado e a sua variância.

•

Nota 4.31 — **F.d. da v.a. uniforme contínua**
A f.d. da v.a. $X \sim \text{uniforme}(a,b)$ possui três troços e é igual a

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(t)dt \\ &= \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b. \end{cases} \end{aligned} \tag{4.16}$$

Graças à simplicidade desta função, ela não se encontra tabelada, à semelhança do que acontece com a da distribuição uniforme discreta. •

Exemplo 4.32 — **Distribuição uniforme contínua**
O diâmetro mínimo e máximo de pistões fornecidos a certo fabricante automóvel é de $18mm$ e $22mm$, respectivamente.
Um dispositivo de controlo, que mede o diâmetro de um pistão, alerta o operador assim que se registe um diâmetro superior a $21.5mm$.[8]

(a) Qual a probabilidade de o operador ser alertado aquando de uma medição, assumindo que o diâmetro dos pistões se distribui uniformemente?

- **V.a.**

  $X =$ diâmetro do pistão (em mm)

- **Distribuição de** X

  $X \sim \text{uniforme}(a,b)$

- **Parâmetros**

  $a =$ diâmetro mínimo $= 18mm$

  $b =$ diâmetro máximo $= 22mm$

- **F.d.p. de** X

$$f_X(x) = \begin{cases} \frac{1}{22-18} = \frac{1}{4}, & 18 \leq x \leq 22 \\ 0, & c.c. \end{cases}$$

[8] Adaptado do Teste A, 11 de Maio de 2002.

- **Probabilidade pedida**

  Seja A o evento que representa a emissão de alerta aquando de uma medição. Então

$$\begin{aligned}
P(A) &= P(X > 21.5) \\
&= \int_{21.5}^{+\infty} f_X(x)dx \\
&= \int_{21.5}^{22} \frac{1}{4}dx + \int_{22}^{+\infty} 0\, dx \\
&= \left.\frac{x}{4}\right|_{21.5}^{22} \\
&= \frac{1}{8}.
\end{aligned}$$

(b) Calcule a probabilidade de o primeiro alerta não ser emitido por nenhuma das 15 primeiras medições de diâmetros.

- **Nova v.a.**

  $Y =$número de medições até à emissão de um alerta

- **Distribuição de Y**

  $Y \sim$ geométrica(p)

- **Parâmetro**

  $p = P(A) = 1/8$

- **F.p. de Y**

  $P(Y = y) = (1-p)^{y-1} \times p = (1-1/8)^{y-1} \times 1/8,\ y = 1, 2, \ldots$

- **Probabilidade pedida**

$$\begin{aligned}
P(Y > 15) &= 1 - P(Y \leq 15) \\
&= 1 - \sum_{y=1}^{15} (1-p)^{y-1} \times p \\
&= 1 - p \times \frac{1-(1-p)^{15}}{1-(1-p)} \\
&= (1-p)^{15} \\
&\simeq 0.135.
\end{aligned}$$

•

Proposição 4.33 — Distribuição uniforme contínua

Considere-se que $X \sim \text{uniforme}(a,b)$. Então os intervalos com a mesma amplitude — desde que contidos em $[a,b]$ — são equiprováveis. Assim,

$$P(c < X \leq c+\Delta) = P(d < X \leq d+\Delta) = \frac{\Delta}{b-a}, \tag{4.17}$$

para $c, c+\Delta, d, d+\Delta \in [a,b]$. •

4.6 Distribuição normal.

Motivação 4.34 — Distribuição normal
Esta distribuição surge associada à modelação de observações relativas a medições de temperaturas, de velocidades, de erros, de volumes de ruído, etc. Surge também como distribuição aproximada, nomeadamente de algumas distribuições discretas[9] ou ainda de médias aritméticas ou somas de v.a.[10] O recurso à distribuição normal nestes casos prende-se com alguma evidência empírica e matemática, respectivamente. Importa referir que o uso desta distribuição deve ser evitado ou deve fazer-se com muita cautela quando a v.a. tem carácter (acentuadamente) assimétrico pois como teremos oportunidade de ver já de seguida esta v.a. tem f.d.p. simétrica.

A utilização da distribuição normal no domínio da Inferência Estatística[11] reveste-se de extrema importância e deve-se a uma série de propriedades matemáticas interessantes desta distribuição. •

Nota 4.35 — Distribuição normal (nota histórica)
A distribuição normal foi introduzida pelo matemático francês Abraham de Moivre (1667–1754) em 1733 e foi utilizada para aproximar probabilidades de eventos respeitantes a v.a. binomiais.[12] E é curioso notar que o trabalho de de Moivre esteve "perdido" por algum tempo, tendo Karl Friedrich Gauss (1777–1855) derivado a distribuição normal de forma independente cerca de 100 anos mais tarde, e que o nome alternativo mais usual para esta distribuição é *distribuição gaussiana*. •

Definição 4.36 — Distribuição normal
A v.a. contínua X diz-se com distribuição normal de parâmetros μ e σ^2 se a sua f.d.p. for igual a

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; -\infty < x < +\infty. \tag{4.18}$$

[9] Como poderemos ver no próximo capítulo.

[10] Graças ao Teorema do Limite Central, como teremos ocasião de descrever de forma mais detalhada no próximo capítulo.

[11] Assunto de que falaremos a partir do Capítulo 6.

[12] Este resultado foi posteriormente estendido por Pierre Simon Laplace (1749–1827) a outras v.a., no século XIX.

Normal	
Notação	$X \sim \text{normal}(\mu, \sigma^2)$
Parâmetros	μ ($\mu \in \mathbb{R}$) σ^2 ($\sigma^2 \in \mathbb{R}^+$)
Contradomínio	$\mathbb{R}$
F.d.p.	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < +\infty$
Valor esperado	$E(X) = \mu$
Variância	$V(X) = \sigma^2$

•

Nota 4.37 — **F.d.p. da v.a. normal**
A f.d.p. da distribuição normal é simétrica em torno de μ e tem a forma de sino.[13] É devido a estes dois factos que a mediana é igual à moda e ao valor esperado, i.e.,

- $me = mo = E(X) = \mu$.

Para além disso, o valor de σ^2 determina o achatamento da f.d.p. desta v.a. ou não se tratasse σ^2 de uma medida de dispersão (absoluta):

- σ^2 muito pequeno, f.d.p. muito alongada;
- σ^2 muito grande, f.d.p. muito achatada. •

Exercício 4.38 — Esboce e compare os gráficos das f.d.p. das v.a.:

- $X \sim \text{normal}(5, 0.5^2)$;
- $X \sim \text{normal}(10, 0.5^2)$;
- $X \sim \text{normal}(10, 3^2)$. •

Nota 4.39 — **F.d. da v.a. normal**
A f.d. da v.a. $X \sim \text{normal}(\mu, \sigma^2)$ é dada por

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \ -\infty < x < +\infty. \tag{4.19}$$

[13] Já agora adiante-se que a f.d.p. da v.a. normal(μ, σ^2) possui dois pontos de inflexão em $\mu \pm \sigma$.

No entanto, $F_X(x)$ não possui expressão fechada, pelo que só pode ser obtida numericamente ou por consulta de tabelas existentes. A consulta destas tabelas encontra justificação nas duas proposições seguintes. •

Proposição 4.40 — **Normal padrão** (ou normal reduzida, ou normal *standard*)
Seja $X \sim \text{normal}(\mu, \sigma^2)$. Então a v.a. $Z = \frac{X-E(X)}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}$ diz-se com distribuição *normal padrão*. Aliás,

$$X \sim \text{normal}(\mu, \sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \sim \text{normal}(0,1). \tag{4.20}$$

Para além disso, Z possui f.d. dada por

$$F_Z(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(z). \tag{4.21}$$

•

A função $\Phi(z)$ está por sinal tabelada para um grande número de valores de z como se ilustra no exemplo seguinte.

Exemplo 4.41 — **Consulta das tabelas da distribuição normal padrão**

- $\Phi(2.53) =$
- $\Phi(0) =$
- $\Phi(1.0) =$. •

Proposição 4.42 — **F.d. da v.a. normal**
A f.d. da v.a. $Z = \frac{X-\mu}{\sigma} \sim \text{normal}(0,1)$ encontra-se tabelada e, como vimos, é usualmente representada por Φ. É à custa desta f.d. e invocando a Proposição 4.40 que se obtém a f.d. da v.a. $X \sim \text{normal}(\mu, \sigma^2)$, para quaisquer valores admissíveis de μ e de σ^2. Com efeito:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned} \tag{4.22}$$

•

Nota 4.43 — Consulta das tabelas da distribuição normal (cont.)
Importa notar que não constam valores de $\Phi(-z)$ para $z > 0$ nas tabelas disponibilizadas para esta disciplina. No entanto, graças à simetria da f.d.p. da normal padrão em torno da origem pode concluir-se que

$$\Phi(-z) = 1 - \Phi(z), -\infty < z < +\infty. \tag{4.23}$$

Por exemplo:

- $\Phi(-2.53) = 1 - \Phi(2.53) =$.

Convém notar também que das tabelas disponibilizadas só constam valores de Φ para $z = 0.00 - 4.09\,(0.01)$. Contudo, ao tirar-se partido do facto de a f.d. de qualquer v.a. ser monótona não decrescente, conclui-se, por exemplo, que

- $\Phi(5.15) > \Phi(4.09) = 0.999978 \simeq 1.0000$

ou que

- $\Phi(-5.15) < \Phi(-4.09) = 1 - \Phi(4.09) = 1 - 0.999978 \simeq 0.0000.$ •

Exemplo 4.44 — Distribuição normal
Uma componente mecânica é constituída por uma biela e uma manivela. Esta componente possui massa total com distribuição normal com valor esperado $1.8Kg$ e variância $0.0049Kg^2$.

Calcule a probabilidade desta componente não respeitar as seguintes especificações: $1.8 \pm 0.075Kg$.

- **V.a.**

  X =massa total da componente

- **Distribuição de** X

  $X \sim \text{normal}(\mu, \sigma^2)$

- **Parâmetros**

  $\mu = E(X) = 1.8Kg$

  $\sigma^2 = V(X) = 0.0049Kg^2$

- **Probabilidade pedida**

  Sejam D o evento que representa respeitar-se as especificações e $Z = \frac{X-E(X)}{\sqrt{V(X)}}$ a v.a. normal padrão (i.e., $Z \sim \text{normal}(0,1)$). Assim, a probabilidade pedida é igual a

$$\begin{aligned}
P(\overline{D}) &= 1 - P(D) \\
&= 1 - P(1.8 - 0.075 \leq X \leq 1.8 + 0.075) \\
&= 1 - P\left[\frac{(1.8-0.075)-E(X)}{\sqrt{V(X)}} \leq \frac{X-E(X)}{\sqrt{V(X)}} \leq \frac{(1.8+0.075)-E(X)}{\sqrt{V(X)}}\right] \\
&= 1 - P\left[\frac{(1.8-0.075)-1.8}{\sqrt{0.0049}} \leq Z \leq \frac{(1.8+0.075)-1.8}{\sqrt{0.0049}}\right] \\
&= 1 - P\left(-\frac{0.075}{0.07} \leq Z \leq \frac{0.075}{0.07}\right) \\
&\simeq 1 - P(-1.07 \leq Z \leq 1.07) \\
&= 1 - [\Phi(1.07) - \Phi(-1.07)] \\
&= 1 - \{\Phi(1.07) - [1 - \Phi(1.07)]\} \\
&= 2 \times [1 - \Phi(1.07)] \\
&\stackrel{tabela}{=} 2 \times (1 - 0.8577) \\
&= 0.2846.
\end{aligned}$$

  Refira-se que a probabilidade de a componente mecânica não respeitar as especificações é elevada, pelo que é necessário melhor o processo de produção, nomeadamente, diminuir a variância da massa desta componente. •

Nota 4.45 — **Mais curiosidades acerca da distribuição normal**
Adiante-se que

Intervalo	Probabilidade
$\mu \pm \sigma$	$P(\mu - \sigma < X \leq \mu + \sigma) = \Phi(1) - \Phi(-1) = 0.8413 - (1 - 0.8413) = 0.6826$
$\mu \pm 2\sigma$	$P(\mu - 2\sigma < X \leq \mu + 2\sigma) = \ldots = 0.9544$
$\mu \pm 3\sigma$	$P(\mu - 3\sigma < X \leq \mu + 3\sigma) = \ldots = 0.9973$
$\mu \pm 1.9600\sigma$	$P(\mu - 1.9600\sigma < X \leq \mu + 1.9600\sigma) = \ldots = 0.9500$
$\mu \pm 2.5758\sigma$	$P(\mu - 2.5758\sigma < X \leq \mu + 2.5758\sigma) = \ldots = 0.9900$

independentemente dos valores do valor esperado (μ) e do desvio–padrão (σ) da distribuição normal.

De realçar que, embora o contradomínio da distribuição normal seja $(-\infty, +\infty)$, ao considerar-se valores positivos para μ suficientemente grandes quando comparados com o valor de σ (e.g. $\mu/\sigma >> 3$) a probabilidade de registar-se valores negativos é irrisória. Por exemplo, caso $X \sim \text{normal}(40, 10^2)$, tem-se

$$\begin{aligned} P(X < 0) &= \Phi[(0-40)/10] \\ &= \Phi(-4.00) \\ &= 1 - \Phi(4.00) \\ &\simeq 0.000000. \end{aligned} \tag{4.24}$$

Justifica-se assim a utilização da distribuição normal na modelação de quantidades numéricas intrinsecamente positivas como é o caso de tempos, volumes de ruídos, alturas, pesos, etc., caso estas v.a. tenham, bem entendido, carácter simétrico e valor esperado muito maior que o desvio-padrão. •

A distribuição normal satisfaz uma série de outras propriedades, mas, para já, enunciemos a seguinte, da qual a Proposição 4.40 é um caso particular.

Proposição 4.46 — **Fecho da distribuição normal para funções lineares**
Seja $X \sim \text{normal}(\mu, \sigma^2)$. Então qualquer sua função linear $Y = aX + b$ verifica:

$$Y = aX + b \sim \text{normal}(a\mu + b, a^2\sigma^2), \tag{4.25}$$

onde a e b são reais arbitrários. •

Veremos mais tarde que a soma de v.a. normais independentes ainda é uma v.a. normal...

Exercício 4.47 — **Fecho da distribuição normal para funções lineares**
Prove a Proposição 4.46. •

É frequente pretender-se resposta a questões como esta:

> *Ao considerar-se que X representa o peso de um indivíduo, quantos Kg se calcula que metade da população não excederá?*

A resposta para esta questão é claramente a mediana, $me = F_X^{-1}(0.5)$, já que $P(X \leq me) = 0.5$. Aliás, a questão poderia ser reformulada e passar-se a solicitar um quantil de probabilidade p.

Nota 4.48 — Quantis da distribuição normal padrão
Foi também disponibilizada uma tabela de quantis da distribuição normal padrão, i.e., respeitantes à v.a. $Z \sim \text{normal}(0, 1)$. A consulta da mesma faz-se sem grande dificuldade. Senão vejamos alguns exemplos de quantis de probabilidade p, com $0.500 \leq p \leq 1.000$, da distribuição normal padrão:

- $F_Z^{-1}(0.975) = \Phi^{-1}(0.975) = 1.9600$
- $\Phi^{-1}(0.974) = 1.9431$.

Mas atente-se que os quantis de probabilidade $p < 0.500$ são negativos.[14] Por exemplo,

- $\Phi^{-1}(0.023) = -\Phi^{-1}(1 - 0.023) = -1.9954$. •

Exemplo 4.49 — Quantis da distribuição normal
As especificações sobre o diâmetro de um certo tipo de cabo admitem um desvio máximo, face ao valor de referência, de ± 10 mícrons. Contudo, no seu fabrico apenas se consegue garantir que esse desvio tem distribuição normal de valor esperado igual a zero mícrons.

Que valor deve ter a variância para se poder garantir que em 95% dos cabos produzidos os desvios estão entre ± 9.8 mícrons?[15]

- **V.a.**

  X = desvio (em mícrons) do diâmetro de um certo tipo de cabo face ao valor de referência

- **Distribuição de** X

  $X \sim \text{normal}(\mu, \sigma^2)$

- **Parâmetros**

  $\mu = E(X) = 0$

  $\sigma^2 = V(X) = ?$

[14] Basta pensar que a f.d.p. da distribuição normal padrão é simétrica em relação à origem e como tal $P(Z \leq 0) = 0.5$.
[15] Adaptado do Teste A de 11 de Novembro de 2006.

- **Obtenção da variância**

  Ao considerar-se a v.a. $Z \sim \text{normal}(0,1)$, tem-se

$$
\begin{aligned}
\sigma^2 \quad : \quad & P(-9.8 \leq X \leq 9.8) = 0.95 \\
& P\left[\frac{-9.8 - E(X)}{\sqrt{V(X)}} \leq \frac{X - E(X)}{\sqrt{V(X)}} \leq \frac{9.8 - E(X)}{\sqrt{V(X)}}\right] = 0.95 \\
& P\left(\frac{-9.8 - 0}{\sigma} \leq Z \leq \frac{9.8 - 0}{\sigma}\right) = 0.95 \\
& \Phi\left(\frac{9.8}{\sigma}\right) - \Phi\left(-\frac{9.8}{\sigma}\right) = 0.95 \\
& 2 \times \Phi\left(\frac{9.8}{\sigma}\right) - 1 = 0.95 \\
& \frac{9.8}{\sigma} = \Phi^{-1}(0.975) \\
& \sigma \stackrel{tabela}{=} \frac{9.8}{1.96} \\
& \sigma^2 = 25.
\end{aligned}
$$

  Ilustra-se assim a utilidade dos quantis na obtenção de um parâmetro da distribuição normal. •

Exercício 4.50 — Quantis da distribuição normal padrão
Prove que o quantil de probabilidade p de uma v.a. $X \sim \text{normal}(\mu, \sigma^2)$ se escreve do seguinte modo à custa do quantil de probabilidade p da normal padrão, $\Phi^{-1}(p)$: $F^{-1}_{normal(\mu,\sigma^2)}(p) = \mu + \sigma \times \Phi^{-1}(p)$. •

4.7 Distribuição exponencial.

Motivação 4.51 — Distribuição exponencial
Trata-se certamente da distribuição contínua mais utilizada na caracterização da duração de equipamento, por exemplo, electrónico, naquilo que usualmente se designa de testes de vida. Esta distribuição surge também na prática no contexto da modelação dos tempos entre ocorrências consecutivas de eventos do mesmo tipo, e.g. chegadas de clientes a um sistema, falhas mecânicas, colisões, etc.

Constate-se que a popularidade da distribuição exponencial prende-se essencialmente com a simplicidade do modelo (e das inferências sobre este modelo), algumas das suas propriedades e alguma evidência empírica. •

Definição 4.52 — Distribuição exponencial
Lidaremos com uma v.a. X com distribuição exponencial de parâmetro λ, caso

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & c.c. \end{cases} \tag{4.26}$$

Exponencial	
Notação	$X \sim \text{exponencial}(\lambda)$
Parâmetro	λ ($\lambda \in I\!R^+$)
Contradomínio	$I\!R_0^+ = [0, +\infty)$
F.d.p.	$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & c.c. \end{cases}$
Valor esperado	$E(X) = \frac{1}{\lambda}$
Variância	$V(X) = \frac{1}{\lambda^2}$

•

Exercício 4.53 — F.d. da v.a. exponencial
Esboce o gráfico da f.d.p. da v.a. $X \sim \text{exponencial}(\lambda)$ e verifique que a f.d. de X é dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0, \end{cases} \tag{4.27}$$

que por sinal também não está tabelada devido à simplicidade da sua expressão geral. •

Exemplo 4.54 — **Distribuição exponencial**

Os *turbofan jet engines* começaram a ser usados há mais de 20 anos como meio de propulsão de aeronaves comerciais: constituem o que se considera uma forma económica e segura de transportar carga e passageiros.

O tempo entre falhas consecutivas (em horas) por parte de certo tipo destes motores possui distribuição exponencial de valor esperado igual a 500 horas.

(a) Obtenha a probabilidade do tempo entre falhas consecutivas exceder $E(X) = 500$ horas. Comente.

- **V.a.**

  $X =$ tempo entre falhas consecutivas (em horas)

- **Distribuição de** X

  $X \sim \text{exponencial}(\lambda)$

- **Parâmetro**

$$\begin{aligned} \lambda : E(X) &= 500 \\ \tfrac{1}{\lambda} &= 500 \\ \lambda &= 0.002 \end{aligned}$$

- **F.d.p. de** X

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda\, e^{-\lambda x}, & x \geq 0. \end{cases}$$

- **F.d. de** X

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

- **Probabilidade pedida**

$$\begin{aligned} P[X > E(X) = 500] &= 1 - F_X[E(X)] \\ &= 1 - F_X\left(1/\lambda\right) \\ &= 1 - \left(1 - e^{-\lambda \frac{1}{\lambda}}\right) \\ &= e^{-1} \end{aligned}$$

  é constante para qualquer valor (positivo, é claro) do parâmetro λ. Mais, $F^{-1}_{Exp(\lambda)}(1 - e^{-1}) = E(X) = 1/\lambda$.

(b) Sabendo que decorreram mais de 800 horas desde o registo da última falha, calcule a probabilidade de virmos a aguardar adicionalmente mais de 500 horas até que ocorra a próxima falha. Compare o valor desta probabilidade com o obtido em (a).

- **Probabilidade pedida**

$$
\begin{aligned}
P(X > 800 + 500|X > 800) &= \frac{P(X > 800 + 500, X > 800)}{P(X > 800)} \\
&= \frac{P(X > 800 + 500)}{P(X > 800)} \\
&= \frac{1 - F_X(800 + 500)}{1 - F_X(800)} \\
&= \frac{1 - \left(1 - e^{-\lambda \times (800+500)}\right)}{1 - (1 - e^{-\lambda \times 800})} \\
&= e^{-\lambda \times 500} \\
&= e^{-1} \\
&= P(X > 500).
\end{aligned}
$$

O valor desta probabilidade condicionada coincide com a probabilidade calculada na alínea anterior, ilustrando assim uma propriedade que se enuncia já de seguida. •

Proposição 4.55 — **Falta de memória da distribuição exponencial**
Considere-se que $X \sim \text{exponencial}(\lambda)$. Então

$$P(X > t + x|X > t) = P(X > x), \forall t, x \in \mathbb{R}_0^+. \tag{4.28}$$

Equivalentemente,

$$(X - t|X > t) \sim \text{exponencial}(\lambda), \forall t \in \mathbb{R}_0^+. \tag{4.29}$$

Assim sendo, ao considerar-se que X representa a vida de um vida de um item e que X tem distribuição exponencial(λ), a vida residual no instante t, $(X - t|X > t)$, possuirá exactamente a mesma distribuição e o mesmo parâmetro. Recorde-se que esta propriedade é denominada de "falta de memória". •

Nota 4.56 — **Falta de memória da distribuição exponencial**
A distribuição exponencial é a única distribuição contínua que satisfaz a propriedade de falta de memória, como aliás acontece com a distribuição geométrica entre as distribuições discretas. Por este facto a distribuição exponencial pode ser considerada o análogo contínuo da distribuição geométrica.

Importa referir que falta de memória da distribuição exponencial torna-a inadequada para modelar tempos de vida de equipamento sujeito a envelhecimento ou desgaste. •

Definição 4.57 — Processo de Poisson
Seja

- N_t o número de ocorrências de um evento no intervalo $(0, t]$,

onde $t > 0$.[16] Então a colecção de v.a. $\{N_t,\, t > 0\}$ diz-se um processo de Poisson de taxa λ caso possua as seguintes propriedades:

- o número de ocorrências de eventos em intervalos disjuntos são v.a. independentes;[17]
- os números de ocorrências de eventos em intervalos com a mesma amplitude são v.a. com a mesma distribuição;[18]
- $N_t \sim \text{Poisson}(\lambda \times t)$, $t > 0$. •

Proposição 4.58 — Relação entre as distribuições exponencial e de Poisson
Seja

- N_t o número de ocorrências de um evento no intervalo $(0, t]$, $t > 0$.
- X_i o tempo entre as ocorrências consecutivas $(i - 1)$ e i do evento, $i \in I\!N$.

Caso a colecção de v.a. $\{N_t,\, t > 0\}$ seja um processo de Poisson de taxa λ, é sabido que $N_t \sim \text{Poisson}(\lambda \times t)$ e pode concluir-se que

$$X_i \overset{i.i.d.}{\sim} \text{exponencial}(\lambda),\ i \in I\!N. \tag{4.30}$$

•

Exercício 4.59 — Prove que a v.a. $X \sim \text{exponencial}(\lambda)$ possui outra propriedade curiosa: tem coeficiente de variação unitário para qualquer valor de λ. •

Exemplo 4.60 — Relação entre as distribuições exponencial e de Poisson
Admita que as chegadas de mensagens electrónicas a um computador são regidas por um processo de Poisson caracterizado por intervalos de tempo entre chegadas consecutivas com distribuição exponencial com valor esperado igual a 2 minutos.

Determine a probabilidade de chegarem mais de 10 mensagens em meia-hora.

[16] Considera-se, naturalmente, que $N_0 = 0$.

[17] Neste caso é habitual dizer-se que o processo possui incrementos independentes, i.e., N_{t_1}, $N_{t_2} - N_{t_1}$, $N_{t_3} - N_{t_2}$, …, $N_{t_n} - N_{t_{n-1}}$ são v.a. independentes, ao considerar-se $0 < t_1 < \ldots < t_n$ e $n \geq 2$.

[18] Nesta situação é costume afirmar-se que o processo possui incrementos estacionários, ou seja, $N(t_2 + s) - N(t_1 + s) \sim N(t_2) - N(t_1)$, para $s > 0$ e $0 \leq t_1 < t_2$.

- **V.a.**

  $X =$ tempo (em minutos) entre chegadas consecutivas de mensagens electrónicas

- **Distribuição de X**

  $X \sim$ exponencial(λ)

- **Parâmetro**

$$\begin{array}{rcl} \lambda : E(X) & = & 2 \\ \frac{1}{\lambda} & = & 2 \\ \lambda & = & 0.5 \text{ (mensagens/minuto)} \end{array}$$

- **Nova v.a.**

  $N_t = N_{30} =$ número de mensagens chegadas em 30 minutos

- **Distribuição de N_{30}**

  Uma vez que se admite que as chegadas de mensagens electrónicas são regidas por um processo de Poisson, pode invocar-se a Proposição 4.58 e afirmar que

$$N_{30} \sim \text{Poisson}(\lambda \times t = 0.5 \times 30 = 15).$$

- **F.p. de N_{30}**

  $P(N_{30} = y) = e^{-15}\,\frac{15^y}{y!}$, $y = 0, 1, 2, \ldots$

- **Probabilidade pedida**

$$\begin{array}{rcl} P(N_{30} > 10) & = & 1 - P(N_{30} \leq 10) \\ & = & 1 - F_{Poisson(15)}(10) \\ & \stackrel{tabela}{=} & 1 - 0.1185 \\ & = & 0.8815. \end{array}$$

•

Exemplo 4.61 — Relação entre as distribuições exponencial e de Poisson (bis) Durante a hora de ponta dos dias úteis, o número de autocarros que passam por hora em certo local tem distribuição de Poisson.[19]

(a) Calcule o valor esperado dessa distribuição, sabendo que a probabilidade de não passar nenhum autocarro nesse local durante 20 minutos é igual a e^{-15}.

- **V.a.**

  N_1 = número de autocarros que passam em uma hora

  $N_{1/3}$ = número de autocarros que passam em 20 min. (i.e., em 1/3 de hora)

- **Distribuições de N_1 e $N_{1/3}$**

  $N_1 \sim \text{Poisson}(\lambda)$

  $N_{1/3} \sim \text{Poisson}(\lambda/3)$

- **F.p. de N_1 e $N_{1/3}$**

  $P(N_1 = y) \stackrel{form}{=} e^{-\lambda}\,\frac{\lambda^y}{y!},\ y = 0, 1, 2, \ldots$

  $P(N_{1/3} = y) = e^{-\lambda/3}\,\frac{(\lambda/3)^y}{y!},\ y = 0, 1, 2, \ldots$

- **Parâmetros**

$$\begin{aligned} \lambda : P(N_{1/3} = 0) &= e^{-15} \\ e^{-\lambda/3}\,\frac{(\lambda/3)^0}{0!} &= e^{-15} \\ \lambda/3 &= 15 \\ \lambda &= 45 \end{aligned}$$

- **Valor esperado pedido**

  $E(N_1) \stackrel{form}{=} \lambda = 45$

(b) Qual é a probabilidade do intervalo de tempo entre a passagem de dois autocarros por esse local exceder 5 minutos?

- **Nova v.a.**

  $N_{1/12}$ = número de autocarros que passam em 5 min. (i.e., em 1/12 de hora)

- **Distribuição de $N_{1/12}$**

  $N_{1/12} \sim \text{Poisson}(\lambda/12 = 3.75)$

- **F.p. de $N_{1/12}$**

  $P(N_{1/12} = y) = e^{-3.75}\,\frac{3.75^y}{y!},\ y = 0, 1, 2, \ldots$

[19] Adaptado do Teste B de 10 de Novembro de 2007.

- **Probabilidade pedida**

  A probabilidade de o tempo entre duas passagens consecutivas de dois autocarros por esse local exceder 5 minutos é exactamente igual a

  $$P(N_{1/12} = 0) = e^{-3.75}\,\frac{3.75^0}{0!} = e^{-3.75} \simeq 0.0235.$$

Ao admitir que as passagens de autocarros são regidas por um processo de Poisson, somos capazes de adiantar uma resolução alternativa recorrendo à Proposição 4.30 que traduz a relação entre as distribuições exponencial e de Poisson.

- **Nova v.a.**

  X = tempo (em horas) entre duas passagens consecutivas de autocarros

- **Distribuição de X**

  Uma vez que $N_1 \sim \text{Poisson}(\lambda)$ pode concluir-se que

  $$X \sim \text{exponencial}(\lambda = 45).$$

  F.d.p. de X

  $$f_X(x) \;=\; \begin{cases} 0, & x < 0 \\ 45\,e^{-45x}, & x \geq 0. \end{cases}$$

- **Probabilidade pedida**

  $$\begin{aligned} P(X > 1/12h = 5min) &= \int_{1/12}^{+\infty} f_X(x)dx \\ &= \int_{1/12}^{+\infty} 45\,e^{-45x}dx \\ &= -e^{-45x}\Big|_{1/12}^{+\infty} \\ &= e^{-45/12} \\ &\simeq 0.0235. \end{aligned}$$

•

Capítulo 5

Distribuições conjuntas de probabilidade e complementos

Motivação 5.1 — Pares aleatórios
Ao realizar-se uma experiência aleatória é comum estarmos interessados em estudar mais do que uma v.a., nomeadamente, pares de v.a. ou pares aleatórios.

É também frequente estarmos preocupados em saber que relação existe entre esse par de v.a., em particular de que modo o conhecimento de uma v.a. pode influenciar o comportamento probabilístico da outra v.a. do par aleatório. •

Nota 5.2 — Pares aleatórios
É usual representar um par aleatório — ou v.a. bidimensional — por (X, Y). •

Exemplo 5.3 — Pares aleatórios

- O número de sinais emitidos (X) por um aparelho e o número desses sinais que foram recebidos (Y) por um aparelho receptor — par de v.a. discretas.

- A classificação do sinal transmitido (X) pelo aparelho emissor e a do sinal recebido (Y) pelo receptor, quanto à sua intensidade (alta, média, baixa) — par de v.a. discretas (qualitativas ordinais).

- A resistência de uma peça (X) e o número de manchas à sua superfície (Y) — par com uma v.a. contínua e uma v.a. discreta.

- A distância (X) de um ponto de impacto ao centro de um alvo e o ângulo (Y) com o eixo das abcissas — par de v.a. contínuas.

- O número de clientes (X) que encontramos à chegada a um sistema e o nosso tempo de permanência em fila de espera (Y) — par com uma v.a. discreta e uma v.a. com carácter misto. •

A definição de par aleatório ou v.a. bidimensional é análoga à de v.a. unidimensional: estamos a lidar mais uma vez com uma função — com características especiais — que transforma, neste caso, eventos em pares ordenados (vectores bidimensionais).

Definição 5.4 — Par aleatório
A função (X, Y) diz-se um par aleatório se possuir

- domínio Ω,
- contradomínio $I\!R_{X,Y}$ contido em $I\!R^2$

e tiver a particularidade de verificar uma condição de mensurabilidade explicitada já a seguir na Nota 5.5. Assim,

$$(X, Y) : \Omega \to I\!R_{X,Y} \subset I\!R^2, \tag{5.1}$$

i.e., a imagem de um evento $\omega \in \Omega$ é um vector bidimensional $(X, Y)(\omega) = (X(\omega), Y(\omega)) \in I\!R_{X,Y} \subset I\!R^n$. •

Nota 5.5 — Mensurabilidade
Considere-se uma e.a., que possui espaço de resultados Ω associado à $\sigma-$álgebra $\mathcal{A}$, e ainda um par aleatório (X, Y).

A condição de mensurabilidade prende-se, no caso bidimensional, com a existência de imagem inversa — segundo (X, Y) — de qualquer região do tipo $(-\infty, x] \times (-\infty, y]$ na $\sigma-$álgebra $\mathcal{A}$. Deste modo,

$$\forall (x, y) \in I\!R^2, (X, Y)^{-1}((-\infty, x] \times (-\infty, y]) \in \mathcal{A}.$$

De notar que, ao considerar-se $(X, Y)(\omega) = (X(\omega), Y(\omega))$, a imagem inversa a que nos referimos é dada por

$$(X, Y)^{-1}((-\infty, x] \times (-\infty, y]) = \{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\},$$

i.e., trata-se da colecção de todos os eventos de Ω aos quais a primeira componente do par aleatório, X, atribui valores no intervalo $(-\infty, x]$ e, simultaneamente, a segunda componente do par aleatório, Y, atribui valores no intervalo $(-\infty, y]$. •

5.1 Duas variáveis aleatórias discretas. Distribuições conjuntas, marginais e condicionais. Independência.

Definição 5.6 — Par aleatório discreto
A par aleatório (X,Y) diz-se discreto, caso tome valores exclusivamente num conjunto de valores finito ou infinito numerável, $\mathbb{R}_{X,Y} = \{(x_i,y_j)\}_{i=1,2,\dots;\, j=1,2,\dots}$, tal que

$$P[(X,Y) \in \mathbb{R}_{X,Y}] = \sum_i \sum_j P(X = x_i, Y = y_j) = 1 \tag{5.2}$$

$$P(X = x_i, Y = y_j) > 0,\ \forall (x_i,y_j) \in \mathbb{R}_{X,Y}. \tag{5.3}$$

•

Motivação 5.7 — Função de probabilidade conjunta
A caracterização probabilística de um par aleatório (X,Y) é usualmente feita à custa da respectiva função de probabilidade conjunta, tal como a de uma v.a. discreta é feita à custa da função de probabilidade. •

Definição 5.8 — Função de probabilidade conjunta
Seja (X,Y) um par aleatório discreto com contradomínio $\mathbb{R}_{X,Y}$ (naturalmente finito ou infinito numerável). Então a f.p. conjunta de (X,Y) é definida por

$$P(X = x, Y = y) = \begin{cases} P(X = x_i, Y = y_j), & \text{se } (x,y) = (x_i,y_j) \in \mathbb{R}_{X,Y} \\ 0, & \text{c.c.} \end{cases} \tag{5.4}$$

e[1] satisfaz:

- $P(X = x, Y = y) > 0,\ \forall (x,y) \in \mathbb{R}_{X,Y}$
- $P(X = x, Y = y) \geq 0,\ \forall (x,y) \in \mathbb{R}^2$
- $\sum_{(x,y)\in\mathbb{R}^2} P(X = x, Y = y) = \sum_{(x,y)\in\mathbb{R}_{X,Y}} P(X = x, Y = y) = 1$
- $P[(X,Y) \in A] = \sum_{(x,y)\in A\cap\mathbb{R}_{X,Y}} P(X = x, Y = y),\ \forall A \subset \mathbb{R}^2$. •

Nota 5.9 — F.p. conjunta
A f.p. conjunta de um par aleatório é raramente representada graficamente. Quando o contradomínio do par aleatório (X,Y) é finito e não muito numeroso — e.g. com $n \times m$ valores — é costume organizar a f.p. conjunta numa tabela do tipo

[1] Note-se que $P(X = x, Y = y) = P(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\})$.

X	Y				
	y_1	$\dots$	y_j	$\dots$	y_m
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1m}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
x_i	$\vdots$	$\dots$	p_{ij}	$\dots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
x_n	p_{n1}	$\dots$	p_{nj}	$\dots$	p_{nm}

onde $p_{ij} = P(X = x_i, Y = y_j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, tal como teremos oportunidade de ver no exemplo que se seguirá. •

Exemplo 5.10 — F.p. conjunta

Volte a considerar um teste americano com 3 questões, cujas respostas são dadas de forma independente. A resposta a cada questão pode estar correcta (C) com probabilidade $P(C) = 0.5$, ou incorrecta ($\overline{C}$) com probabilidade $P(\overline{C}) = 0.5$. Considere o par aleatório (X, Y), onde X e Y representam o número de respostas correctas e incorrectas no teste, respectivamente.

(a) Confirme que o conjunto de valores possíveis de (X, Y) é $\{(0,3), (1,2), (2,1), (3,0)\}$.

- **Par aleatório** (X, Y)

  X = número de respostas correctas neste teste americano

  Y = número de respostas incorrectas no mesmo

- **Contradomínio de** (X, Y)

  O conjunto de valores possíveis do par aleatório (X, Y) obtém-se identificando os valores de X e Y associados a cada evento elementar:

Eventos elementares	X	Y
$\overline{C}\overline{C}\overline{C} = \overline{C}_1 \cap \overline{C}_2 \cap \overline{C}_3$	0	3
$C\overline{C}\overline{C}$	1	2
$\overline{C}C\overline{C}$	1	2
$\overline{C}\overline{C}C$	1	2
$CC\overline{C}$	2	1
$C\overline{C}C$	2	1
$\overline{C}CC$	2	1
CCC	3	0

  onde C_i = resposta correcta à questão i, $i = 1, 2, 3$.

  Assim, confirma-se que $I\!R_{X,Y} = \{(0,3), (1,2), (2,1), (3,0)\}$.

(b) Defina a f.p. conjunta do par aleatório (X, Y).

- **F.p. conjunta** de (X, Y)

  Notando que a v.a. Y é uma função afim de X ($Y = 3 - X$) depressa se conclui que

$$
\begin{aligned}
P(X=0, Y=3) &= P(\overline{C}\,\overline{C}\,\overline{C}) \\
&= P(X=0) \\
&= \frac{1}{8} \\
P(X=1, Y=2) &= P(C\overline{C}\,\overline{C}) + P(\overline{C}C\overline{C}) + P(\overline{C}\,\overline{C}C) \\
&= P(X=1) \\
&= \frac{3}{8} \\
P(X=2, Y=1) &= P(CC\overline{C}) + P(C\overline{C}C) + P(\overline{C}CC) \\
&= P(X=2) \\
&= \frac{3}{8} \\
P(X=3, Y=0) &= P(CCC) \\
&= P(X=3) \\
&= \frac{1}{8},
\end{aligned}
$$

  ou de forma resumida

$$
P(X=x, Y=y) = \begin{cases} 1/8, & (x,y) = (0,3), (3,0) \\ 3/8, & (x,y) = (1,2), (2,1) \\ 0, & \text{outros pares de valores de } (x,y), \end{cases}
$$

  ou ainda sob a forma de uma tabela

X	Y			
	0	1	2	3
0	0	0	0	$\frac{1}{8}$
1	0	0	$\frac{3}{8}$	0
2	0	$\frac{3}{8}$	0	0
3	$\frac{1}{8}$	0	0	0

•

Definição 5.11 — Função de distribuição conjunta (caso discreto)
A f.d. conjunta do par aleatório discreto (X, Y) pode ser obtida à custa da f.p. conjunta:[2]

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) \\ &= \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j),\ (x,y) \in I\!R^2. \end{aligned} \tag{5.5}$$

•

Depois de um exemplo intencionalmente trivial considere-se um par aleatório um pouco mais rebuscado para ilustrar o cálculo da f.d. conjunta.

Exemplo 5.12 — F.d. conjunta (caso discreto)
Na transmissão de informação digital por certo equipamento, a probabilidade de um "bit" possuir distorção alta, moderada e baixa é de 0.01, 0.04 e 0.95, respectivamente. Suponha que são transmitidos 3 "bits" de modo independente e que X e Y representam o número de "bits" com distorção alta e moderada, respectivamente.[3]

(a) Complete, justificando, as entradas assinaladas com a, b e c, na seguinte tabela da f.p. conjunta do par aleatório (X, Y).

X	Y			
	0	1	2	3
0	a	0.108300	0.004560	0.000064
1	0.027075	0.002280	0.000048	c
2	0.000285	b	0	0
3	0.000001	0	0	0

- **Par aleatório** (X, Y)

  X = número de "bits" com distorção alta, em 3 "bits" emitidos

  Y = número de "bits" com distorção moderada, em 3 "bits" emitidos

- **Obtenção das constantes** a, b, c

  Para obter as constantes a e b é necessário ter em conta que por um lado $X \sim \text{binomial}(3, 0.01)$ — i.e.,

  $$P(X = x) = \binom{3}{x} 0.01^x\,(1 - 0.01)^{3-x},\ x = 0, 1, 2, 3$$

  — e que por outro $P(X = x) = \sum_{y=0}^{3} P(X = x, Y = y)$, $x = 0, 1, 2, 3$. Deste modo:

[2] As propriedades desta função serão descritas na secção seguinte.
[3] Adaptado do Exame de 13 de Julho de 2002.

$$
\begin{aligned}
a &= P(X=0,Y=0)\\
&= P(X=0) - \sum_{y=1}^{3} P(X=0,Y=y)\\
&= (1-0.01)^3 - (0.108300 + 0.004560 + 0.000064)\\
&= 0.857375
\end{aligned}
$$

$$
\begin{aligned}
b &= P(X=2,Y=1)\\
&= P(X=2) - \left[\sum_{y=0}^{1} P(X=2,Y=y) + P(X=2,Y=3)\right]\\
&= \frac{3!}{2!(3-2)!}\, 0.01^2 \times (1-0.01)^1 - (0.000285 + 0 + 0)\\
&= 0.000012
\end{aligned}
$$

$$
\begin{aligned}
c &= P(X=1,Y=3)\\
&= P(\emptyset)\\
&= 0.
\end{aligned}
$$

- **Nota**

  Comece-se por considerar que a v.a. $Z = 3 - X - Y$ representa o número "bits" com distorção baixa, em 3 "bits" emitidos e refira-se que $Z \sim \text{binomial}(3, 0.95)$. Então importa notar que, embora os eventos $\{X = 0, Y = 0\}$ e $\{Z = 3\}$ sejam equivalentes, os eventos $\{X = 2, Y = 1\}$ e $\{Z = 0\}$ não o são, pelo que

$$
\begin{aligned}
a &= P(X=0,Y=0)\\
&= P(3 \text{ "bits" com distorção baixa})\\
&= P(Z=3)\\
&= 0.95^3\\
&= 0.857375
\end{aligned}
$$

  e, no entanto,

$$
\begin{aligned}
b &= P(X=2,Y=1)\\
&\neq P(0 \text{ "bits" com distorção baixa})\\
&= P(Z=0)\\
&= (1-0.95)^3\\
&= 0.000125.
\end{aligned}
$$

Aliás, há quatro situações em que ocorre $\{Z = 0\}$, são elas $\{X = 0, Y = 3\}$, $\{X = 1, Y = 2\}$, $\{X = 2, Y = 1\}$ e $\{X = 3, Y = 0\}$.

$$\begin{aligned} P(Z=0) &= P(0 \text{ "bits" com distorção baixa}) \\ &= (1-0.95)^3 \\ &= 0.000125 \\ &= P(X=0,Y=3)+P(X=1,Y=2) \\ &\quad +P(X=2,Y=1)+P(X=3,Y=0) \\ &= 0.000064+0.000048+0.000012+0.000001 \\ &= 0.000125. \end{aligned}$$

(b) Determine a probabilidade de registar-se a transmissão de não mais de dois "bits" com distorção alta e quando muito um "bit" com distorção moderada.

- **Probabilidade pedida**

$$\begin{aligned} P(X \leq 2, Y \leq 1) &= F_{X,Y}(2,1) \\ &= \sum_{x=0}^{2}\sum_{y=0}^{1} P(X=x, Y=y) \\ &= 0.857375+0.108300+0.027075+0.002280 \\ &\quad +0.000285+0.000012 \\ &= 0.995327. \end{aligned}$$

•

Exemplo 5.13 — **F.p. conjunta** (e algo mais)
Um painel metálico é classificado de acordo com:

- X = número de defeitos que possui (0,1,2,3 defeitos)
- Y = número da fábrica que o produz (Fab.1 e Fab.2).

Estudos levados a cabo levam a crer que a f.p. conjunta de (X, Y) é dada pela seguinte tabela:

X	Y = 1	Y = 2
0	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{16}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{4}$

Calcule a probabilidade de um painel metático seleccionado ao acaso:

(a) possuir 2 defeitos e ter sido produzido pela Fab.1;

- **Par aleatório** (X, Y)

  $X =$ número de defeitos que o painel possui

  $Y =$ número da fábrica que o produz

- **Probabilidade pedida**

  $P(X = 2, Y = 1) = \frac{3}{16}$.

(b) possuir somente 2 defeitos;

- **Probabilidade pedida**

$$\begin{aligned} P(X = 2) &= P(X = 2, Y = 1) + P(X = 2, Y = 2) \\ &= \frac{3}{16} + \frac{1}{8} \\ &= \frac{5}{16}. \end{aligned}$$

(c) ser produzido pela Fab.1.

- **Probabilidade pedida**

$$\begin{aligned} P(Y = 1) &= \sum_{x=0}^{3} P(X = x, Y = 1) \\ &= \frac{1}{8} + \frac{1}{16} + \frac{3}{16} + \frac{1}{8} \\ &= \frac{1}{2}. \end{aligned}$$

•

As alíneas (b) e (c) do Exemplo 5.13 sugerem a definição das f.p. de X e de Y, que se dirão f.p. marginais.

Definição 5.14 — F.p. marginais de X e de Y

A f.p. marginal de X define-se à custa da f.p. conjunta do par aleatório (X, Y):[4]

$$P(X = x) = \sum_{y} P(X = x, Y = y). \tag{5.6}$$

A f.p. marginal de Y obtém-se de modo análogo:[5]

$$P(Y = y) = \sum_{x} P(X = x, Y = y). \tag{5.7}$$

•

[4]Caso a f.p. conjunta esteja organizada numa tabela como as duas descritas acima, a f.p. marginal de X obtém-se somando linha a linha.

[5]Do mesmo modo a f.p. marginal de Y obtém-se somando coluna a coluna ao lidar-se com tabelas como as duas anteriores.

Exercício 5.15 — **F.p. marginais**
Retome o Exemplo 5.13 e obtenha as f.p. marginais de X e de Y. •

Nota 5.16 — É recorrendo às f.p. marginais da v.a. X e a da v.a. Y que se obtém

- $F_X(x)$, $E(X)$, $V(X)$, etc.
- $F_Y(y)$, $E(Y)$, $V(Y)$, etc. •

Definição 5.17 — **F.d. marginais de X e de Y (caso discreto)**
Pode obter-se as f.d. marginais de X e de Y quer à custa das f.p. destas v.a. unidimensionais, quer recorrendo à f.p. conjunta:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{x_i \leq x} P(X = x_i) \\ &= \sum_{x_i \leq x} \sum_{y} P(X = x_i, Y = y), \; x \in I\!R \end{aligned} \tag{5.8}$$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= \sum_{y_j \leq y} P(Y = y_j) \\ &= \sum_{y_j \leq y} \sum_{x} P(X = x, Y = y_j), \; y \in I\!R. \end{aligned} \tag{5.9}$$

•

Motivação 5.18 — **F.p. condicionais**
É importante averiguar de que modo o registo de uma observação de uma das v.a. do par aleatório poderá vir a influenciar o comportamento probabilístico da outra v.a. Para o efeito será necessário definir f.p. condicionais. •

Definição 5.19 — **F.p. condicionais**

- F.p. de X condicional a $Y = y$

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \text{ se } P(Y = y) > 0. \tag{5.10}$$

- F.p. de Y condicional a $X = x$

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}, \text{ se } P(X = x) > 0. \tag{5.11}$$

•

Nota 5.20 — **V.a. condicionais**
Ao efectuar semelhantes condicionamentos passamos a lidar com duas v.a. unidimensionais — $X|Y = y$ e $Y|X = x$ —, pelo que faz sentido calcular f.d., valores esperados, variâncias, etc., em todo o caso ditas/os condicionais. •

Definição 5.21 — **F.d. condicionais (caso discreto)**

- F.d. de X condicional a $Y = y$ [6]

$$\begin{aligned} F_{X|Y=y}(x) &= P(X \leq x|Y = y) \\ &= \sum_{x_i \leq x} P(X = x_i|Y = y) \\ &= \sum_{x_i \leq x} \frac{P(X = x_i, Y = y)}{P(Y = y)}, \ x \in I\!R \end{aligned} \tag{5.12}$$

- F.d. de Y condicional a $X = x$ [7]

$$\begin{aligned} F_{Y|X=x}(y) &= P(Y \leq y|X = x) \\ &= \sum_{y_j \leq y} P(Y = y_j|X = x) \\ &= \sum_{y_j \leq y} \frac{P(X = x, Y = y_j)}{P(X = x)}, \ y \in I\!R. \end{aligned} \tag{5.13}$$

•

Exemplo 5.22 — **F.p. e f.d. condicionais (caso discreto)**
Responda às seguintes questões que dizem respeito ao Exemplo 5.13.

(a) Determine a probabilidade de um painel metálico possuir um defeito sabendo que foi produzido pela Fab.2.

- **Par aleatório** (X, Y)

  X = número de defeitos que o painel possui

  Y = número da fábrica que o produz

- **F.p. conjunta de** (X, Y)

  Convém recordar que esta função pode ser condensada na seguinte tabela:

[6] É necessário que $P(Y = y) > 0$ para definir esta função.
[7] Analogamente, esta f.d. só faz sentido se $P(X = x) > 0$.

X	Y	
	1	2
0	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{16}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{4}$

- **Probabilidade pedida**

$$\begin{aligned} P(X=1|Y=2) &= \frac{P(X=1,Y=2)}{P(Y=2)} \\ &= \frac{1/16}{1/16+1/16+1/8+1/4} \\ &= \frac{1/16}{1/2} \\ &= \frac{1}{8}. \end{aligned}$$

(b) Calcule a probabilidade de um painel metálico possuir quanto muito um defeito, quando se sabe de antemão que foi produzido pela Fab.2.

- **Probabilidade pedida**

$$\begin{aligned} P(X\leq 1|Y=2) &= F_{X|Y=2}(1) \\ &= \sum_{x=0}^{1} P(X=x|Y=2) = \sum_{x=0}^{1} \frac{P(X=x,Y=2)}{P(Y=2)} \\ &= \frac{1/16}{1/2}+\frac{1/16}{1/2} \\ &= \frac{1}{4}. \end{aligned}$$

•

Definição 5.23 — Valores esperados condicionais (caso discreto)

- Valor esperado de X condicional a $Y=y$

$$E(X|Y=y) = \sum_x x\, P(X=x|Y=y) \tag{5.14}$$

- Valor esperado de Y condicional a $X=x$

$$E(Y|X=x) = \sum_y y\, P(Y=y|X=x). \tag{5.15}$$

•

Definição 5.24 — Variâncias condicionais (caso discreto)

- Variância de X condicional a $Y = y$

$$
\begin{aligned}
V(X|Y=y) &= E(X^2|Y=y) - E^2(X|Y=y) \\
&= \sum_x x^2\, P(X=x|Y=y) - \left[\sum_x x\, P(X=x|Y=y)\right]^2 \qquad (5.16)
\end{aligned}
$$

- Variância de Y condicional a $X = x$

$$
\begin{aligned}
V(Y|X=x) &= E(Y^2|X=x) - E^2(Y|X=x) \\
&= \sum_y y^2\, P(Y=y|X=x) - \left[\sum_y y\, P(Y=y|X=x)\right]^2. \qquad (5.17)
\end{aligned}
$$

•

Exemplo 5.25 — Valor esperado condicional (e não só)
Uma firma envia dois tipos de faxes. O primeiro tipo de fax requer 40 segundos para a transmissão de cada página ao passo que o segundo tipo requer 60 segundos para essa transmissão.

Considere que a v.a. X representa o número de páginas de cada fax e Y a duração da transmissão de todo o fax. Após um estudo detalhado, baseado em centenas de transmissões de fax, a firma é da opinião que a função de probabilidade conjunta do par aleatório (X, Y) é igual a:

X	Y				
	40	60	80	120	180
1	0.15	0.10	0	0	0
2	0	0	0.30	0.35	0
3	0	0	0	0	0.10

Compare o valor esperado da duração da transmissão de um fax e a duração esperada da transmissão de um fax sabendo que tem 2 páginas. O que conclui?[8]

- **Par aleatório** (X, Y)

  $X =$ número de páginas de cada fax

  $Y =$ duração (em segundos) da transmissão do fax

[8] Adaptado do Exame de 17 de Janeiro de 2004.

- **F.p. marginal de Y**

  Obtém-se somando as entradas da tabela da f.p. conjunta de (X, Y), coluna a coluna:

$$
\begin{aligned}
P(Y = y) &= \sum_{x=1}^{3} P(X = x, Y = y) \\
&= \begin{cases} 0.15, & y = 40 \\ 0.10, & y = 60 \\ 0.30, & y = 80 \\ 0.35, & y = 120 \\ 0.10, & y = 180 \\ 0, & \text{outros valores de } y \end{cases}
\end{aligned}
$$

- **Valor esperado de Y**

$$
\begin{aligned}
E(Y) &= \sum_y y \times P(Y = y) \\
&= 40 \times 0.15 + 60 \times 0.10 + 80 \times 0.30 + 120 \times 0.35 + 160 \times 0.10 \\
&= 96 \text{ segundos}
\end{aligned}
$$

- **Nova v.a.**

  $Y|X = 2 =$ duração da transmissão do fax sabendo que tem 2 páginas

- **F.p. de Y condicional a $\{X = 2\}$**

  Ao notar que $P(X = 2) = \sum_y P(X = 2, Y = y) = 0 + 0 + 0.30 + 0.35 + 0 = 0.65$ e que os valores possíveis para a duração da transmissão de fax com 2 páginas são 80 e 120 segundos, segue-se:

$$
\begin{aligned}
P(Y = y|X = 2) &= \frac{P(X = 2, Y = y)}{P(X = 2)} \\
&= \begin{cases} \frac{0.30}{0.65} = \frac{6}{13}, & y = 80 \\ \frac{0.35}{0.65} = \frac{7}{13}, & y = 120 \\ 0, & \text{outros valores de } y \end{cases}
\end{aligned}
$$

- **Valor esperado de Y condicional a $\{X = 2\}$**

$$
\begin{aligned}
E(Y|X = 2) &= \sum_y y \times P(Y = y|X = 2) \\
&= 80 \times \frac{6}{13} + 120 \times \frac{7}{13} \\
&= \frac{1320}{13} \\
&\simeq 101.54 \text{ segundos}
\end{aligned}
$$

- **Conclusão** — Como $E(Y) \neq E(Y|X=2)$ pode adiantar-se que o conhecimento do número de páginas do fax altera a duração esperada da transmissão do mesmo. Logo as v.a. X e Y influenciam-se uma à outra; ou por outras palavras: X e Y NÃO SÃO V.A. INDEPENDENTES. •

Exercício 5.26 — Retomemos os dados do Exemplo 5.13.

(a) Prove que o número esperado de defeitos de um painel metálico que se sabe ter sido produzido pela Fab.2, $E(X|Y=2)$, é igual a $\frac{17}{8}$.

(b) Compare o valor esperado condicional obtido em (a) com $E(X)$ (por sinal igual a $\frac{15}{8}$). Comente. •

Motivação 5.27 — **(In)Dependência**
De um modo geral as v.a. que constituem um par aleatório influenciam-se mutuamente como no exemplo anterior, ou seja, são dependentes. Outras há em que tal não acontece. Com efeito o conceito de independência entre (dois) eventos pode ser estendido a (pares de) v.a. •

Definição 5.28 — **Independência (caso discreto)**
As v.a. discretas X e Y dizem-se independentes — escrevendo-se neste caso $X \perp\!\!\!\perp Y$ — se

$$P(X=x, Y=y) = P(X=x) \times P(Y=y), \ \forall (x,y) \in I\!R^2, \tag{5.18}$$

i.e., caso seja possível escrever a f.p. conjunta do par aleatório discreto (X,Y) à custa do produto das f.p. marginais de X e de Y. •

Nota 5.29 — **Independência (caso discreto)**
A Equação (5.18), que define independência entre as v.a. discretas X e Y, é equivalente a

$$F_{X,Y}(x,y) = F_X(x) \times F_Y(y), \ \forall (x,y) \in I\!R^2. \tag{5.19}$$

Ou seja, se $X \perp\!\!\!\perp Y$ é também possível escrever a f.d. conjunta do par aleatório discreto (X,Y) à custa do produto das f.d. marginais de X e de Y. Para além disso, caso $X \perp\!\!\!\perp Y$, segue-se:

- $P(X=x|Y=y) = P(X=x), \ \forall (x,y) \in I\!R^2$
- $E(X|Y=y) = E(X), \ \forall y \in I\!R$
- $V(X|Y=y) = V(X), \ \forall y \in I\!R$, etc.,

como tal o conhecimento de Y não vem influenciar o comportamento probabilístico de X. Analogamente, $P(Y=y|X=x) = P(Y=y), \ \forall (x,y) \in I\!R^2$, etc., caso $X \perp\!\!\!\perp Y$. •

Exemplo/Exercício 5.30 — Independência (caso discreto)

O número de portáteis (X) e PCs (Y) vendidos diariamente numa loja de material informático têm função de probabilidade conjunta dada parcialmente pela tabela abaixo:[9]

X	Y		
	0	1	2
0	0.1	0.1	0.3
1	0.2	0.1	0.1
2	0	0.1	a

(a) Complete a tabela e prove que X e Y são v.a. dependentes.

- **Par aleatório** (X, Y)

  $X =$ número de portáteis vendidos diariamente

  $Y =$ número de PCs vendidos diariamente

- **Obtenção de** a

$$a \quad : \quad \sum_{x=0}^{2}\sum_{y=0}^{2} P(X = x, Y = y) = 1$$

$$0.1 + 0.1 + 0.3 + 0.2 + 0.1 + 0.1 + 0 + 0.1 + a = 1$$

$$a = 0$$

- **Dependência entre** X **e** Y

  Se por um lado

$$P(X = 0, Y = 0) = 0.1,$$

  por outro

$$\begin{aligned} P(X = 0) \times P(Y = 0) &= \sum_{y=0}^{2} P(X = 0, Y = y) \\ &\quad \times \sum_{x=0}^{2} P(X = x, Y = 0) \\ &= (0.1 + 0.1 + 0.3) \times (0.1 + 0.2 + 0) \\ &= 0.15. \end{aligned}$$

  Deste modo conclui-se que

$$P(X = 0, Y = 0) \neq P(X = 0) \times P(Y = 0),$$

  pelo que X e Y são v.a. DEPENDENTES.

[9] Adaptado do Teste B de 22 de Abril de 2006.

(b) Calcule $V(X|Y \geq 1)$.

- **Nova v.a.**

  $X|Y \geq 1$ = número de portáteis vendidos em certo dia, sabendo que ocorreu venda de PCs

- **F.p. de X condicional a $\{Y \geq 1\}$**

  Uma vez que

$$\begin{aligned}
P(Y \geq 1) &= P(Y=1) + P(Y=2) \\
&= \sum_{x=0}^{2} P(X=x, Y=1) + \sum_{x=0}^{2} P(X=x, Y=2) \\
&= (0.1 + 0.1 + 0.1) + (0.3 + 0.1 + 0) \\
&= 0.7,
\end{aligned}$$

  tem-se

$$\begin{aligned}
P(X = x|Y \geq 1) &= \frac{P(X = x, Y \geq 1)}{P(Y \geq 1)} \\
&= \frac{P(X = x, Y = 1) + P(X = x, Y = 2)}{P(Y \geq 1)} \\
&= \begin{cases} \frac{0.1+0.3}{0.7} = \frac{4}{7}, & x = 0 \\ \frac{0.1+0.1}{0.7} = \frac{2}{7}, & x = 1 \\ \frac{0.1+0}{0.7} = \frac{1}{7}, & x = 2 \\ 0, & \text{restantes valores de } x \end{cases}
\end{aligned}$$

- **Valor esperado de $X|Y \geq 1$**

$$\begin{aligned}
E(X|Y \geq 1) &= \sum_{x=0}^{2} x \times P(X = x|Y \geq 1) \\
&= 0 \times \frac{4}{7} + 1 \times \frac{2}{7} + 2 \times \frac{1}{7} \\
&= \frac{4}{7}
\end{aligned}$$

- **2o. momento de $X|Y \geq 1$**

$$\begin{aligned}
E(X^2|Y \geq 1) &= \sum_{x=0}^{2} x^2 \times P(X = x|Y \geq 1) \\
&= 0^2 \times \frac{4}{7} + 1^2 \times \frac{2}{7} + 2^2 \times \frac{1}{7} \\
&= \frac{6}{7}
\end{aligned}$$

- **Variância de** $X|Y \geq 1$

$$\begin{aligned} V(X|Y \geq 1) &= E(X^2|Y \geq 1) - E^2(X|Y \geq 1) \\ &= \frac{6}{7} - \left(\frac{4}{7}\right)^2 \\ &= \frac{26}{49}. \end{aligned}$$

(c) Obtenha o valor esperado do número de portáteis vendidos diariamente, tirando partido do facto de

$$\begin{aligned} E(X) &= \sum_{x \in \mathbb{R}_X} x \times P(X = x) \\ &= \sum_{x \in \mathbb{R}_X} x \times \left[\sum_{y \in \mathbb{R}_Y} P(X = x, Y = y)\right] \\ &= \sum_{x \in \mathbb{R}_X} x \times \left[\sum_{y \in \mathbb{R}_Y} P(X = x|Y = y) \times P(Y = y)\right] \\ &= \sum_{y \in \mathbb{R}_Y} \left[\sum_{x \in \mathbb{R}_X} x \times P(X = x|Y = y)\right] \times P(Y = y) \\ &= \sum_{y \in \mathbb{R}_Y} E(X|Y = y) \times P(Y = y) \\ &= E[E(X|Y)]. \end{aligned}$$

Note que a v.a. $E(X|Y)$ toma os valores $E(X|Y = y)$ com probabilidade $P(Y = y)$, para $y \in \mathbb{R}_Y$. •

5.2 Duas variáveis aleatórias contínuas. Distribuições conjuntas, marginais e condicionais. Independência.

Motivação 5.31 — **F.d.p. conjunta**
A caracterização probabilística de uma v.a. unidimensional contínua faz-se recorrendo à f.d.p. Não surpreende pois que ao lidar com um par aleatório (X, Y) contínuo seja necessário definir f.d.p. conjunta. •

Definição 5.32 — **Par aleatório contínuo e f.d.p. conjunta**
A par aleatório (X, Y) diz-se contínuo, se tomar valores num conjunto de valores infinito não numerável, $I\!R_{X,Y} \subset I\!R^2$, e existir uma função denominada de f.d.p. conjunta, $f_{X,Y}(x, y)$, satisfazendo as propriedades seguintes:

- $f_{X,Y}(x, y) \geq 0, \forall (x, y) \in I\!R^2$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy dx = 1$
- $P[(X, Y) \in A] = \int \int_A f_{X,Y}(x, y) dy dx, \ \forall A \subset I\!R^2$. •

Exemplo 5.33 — **F.d.p. conjunta**
O par aleatório (X, Y) representa os tempos de vida, em milhares de horas, de duas componentes principais (A e B) de um sistema de controlo e possui f.d.p. conjunta dada por

$$f_{X,Y}(x, y) = \begin{cases} 1, & 0 < x < 1, \, 0 < y < 1 \\ 0, & \text{c.c.} \end{cases} \tag{5.20}$$

Calcule a probabilidade de X e de Y excederem 0.50 e 0.75, respectivamente.

- **Par aleatório**

  $X =$ tempo de vida da componente A (em $10^3 h$)

  $Y =$ tempo de vida da componente B (em $10^3 h$)

- **F.d.p. de** (X, Y)

$$f_{X,Y}(x, y) = \begin{cases} 1, & 0 < x < 1, \, 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$

- **Probabilidade pedida**

$$\begin{aligned}
P(X > 0.5, Y > 0.75) &= \int_{0.5}^{+\infty} \int_{0.75}^{+\infty} f_{X,Y}(x,y) dy\, dx \\
&= \int_{0.5}^{1} \int_{0.75}^{1} dy\, dx \\
&= \int_{0.5}^{1} y|_{0.75}^{1}\, dx \\
&= \int_{0.5}^{1} 0.25\, dx \\
&= 0.125
\end{aligned}$$

Importa notar que

$$P(X > 0.5, Y > 0.75) \neq 1 - P(X \leq 0.5, Y \leq 0.75).$$

•

Definição 5.34 — F.d. conjunta (caso contínuo)
A f.d. conjunta do par aleatório contínuo (X, Y) define-se à custa da sua f.d.p. conjunta:

$$\begin{aligned}
F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) \\
&= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v) dv du, \; \forall (x,y) \in I\!R^2, \qquad (5.21)
\end{aligned}$$

que corresponde, obviamente, ao volume sob a superfície da f.d.p. conjunta na região $(-\infty, x] \times (-\infty, y]$. •

Proposição 5.35 — Propriedades da f.d. conjunta de par aleatório (discreto ou contínuo)
A f.d. conjunta de um par aleatório (X, Y) satisfaz as propriedades quer no caso discreto, quer no caso contínuo. É uma:

1. função contínua à direita (no caso discreto) e contínua à direita e à esquerda (no caso contínuo)

2. função monótona não decrescente em qualquer das variáveis x e y.

Para além disso verifica:

3. $0 \leq F_{X,Y}(x,y) \leq 1$

4. $F_{X,Y}(-\infty, -\infty) = 0$

   $F_{X,Y}(+\infty, +\infty) = 1$

5. $F_{X,Y}(x, +\infty) = F_X(x)$, $\forall x \in I\!R$

   $F_{X,Y}(+\infty, y) = F_Y(y)$, $\forall y \in I\!R$.

Por seu lado, ao considerar-se a região

$$A = \{(x, y) \in I\!R^2 : a < x \leq b, c < y \leq d\}$$

(onde $a < b$ e $c < d$), tem-se:

6. $P[(X, Y) \in A] = F_{X,Y}(b, d) - F_{X,Y}(b, c) - F_{X,Y}(a, d) + F_{X,Y}(a, c)$. •

Exercício 5.36 — Demonstre a propriedade 6. da Proposição 5.35 recorrendo a gráficos e à avaliação sucessiva da f.d. conjunta. •

Exemplo/Exercício 5.37 — **F.d. conjunta (caso contínuo)**
Sejam X e Y v.a. que representam, respectivamente, a largura (em dm) e o comprimento (em dm) de uma peça rectangular. Admita que a f.d.p. conjunta de (X, Y) é dada por

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{c.c.} \end{cases} \tag{5.22}$$

(a) Represente graficamente o contradomínio e a f.d.p. conjunta do par aleatório.

- **Par aleatório**

  X = largura da peça rectangular (em dm)

  Y = comprimento da peça rectangular (em dm)

- **Representação gráfica do contradomínio de** (X, Y)

- **Gráfico da f.d.p. de** (X, Y)

(b) Obtenha a probabilidade de a largura e o comprimento da peça não excederem $5cm$ e $7.5cm$, respectivamente.

- **Probabilidade pedida**

$$\begin{aligned} P(X \leq 0.5, Y \leq 0.75) &= F_{X,Y}(0.5, 0.75) \\ &= \int_{-\infty}^{0.5} \int_{-\infty}^{0.75} f_{X,Y}(u,v)dv\,du \\ &= \int_{0}^{0.5} \int_{u}^{0.75} 2\,dv\,du \\ &= \int_{0}^{0.5} 2v|_{u}^{0.75}\,du \\ &= 2\int_{0}^{0.5} (0.75-u)\,du \\ &= 2\left(0.75u - \frac{u^2}{2}\right)\Big|_0^{0.5} \\ &= 0.5. \end{aligned}$$

(c) Verifique que a f.d. conjunta deste par aleatório é dada por

$$F_{X,Y}(x,y) = \begin{cases} 0, & x < 0 \text{ ou } y < 0 \\ 2x\left(y - \frac{x}{2}\right), & 0 < x < y < 1 \\ 2x\left(1 - \frac{x}{2}\right), & 0 < x < 1,\, y > 1 \\ y^2, & 0 < y < x < 1 \text{ ou } 0 < y < 1 < x \\ 1, & x > 1,\, y > 1. \end{cases}$$

•

Como seria de esperar a apresentação que se segue muito se assemelha à da secção anterior. Há uma diferença óbvia: ao invés de se somar, integrar-se-á, afinal estamos a lidar com pares aleatórios contínuos.

Definição 5.38 — F.d.p. marginais de X e de Y
As f.d.p. marginais de X e de Y obtêm-se à custa da f.d.p. conjunta do par aleatório contínuo (X, Y):

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \tag{5.23}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx. \tag{5.24}$$

•

Definição 5.39 — F.d. marginais de X e de Y (caso contínuo)
As f.d. marginais de X e de Y calculam-se quer à custa das f.d.p. marginais destas v.a. unidimensionais, quer por recurso à f.d.p. conjunta ou à f.d. conjunta:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(u)du \\ &= \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f_{X,Y}(u,y)dydu = F_{X,Y}(x,+\infty),\ x \in \mathbb{R} \end{aligned} \tag{5.25}$$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= \int_{-\infty}^{y} f_Y(v)dv \\ &= \int_{-\infty}^{y}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdu = F_{X,Y}(+\infty,y),\ y \in \mathbb{R}. \end{aligned} \tag{5.26}$$

•

Exemplo/Exercício 5.40 — F.d.p. marginais de X e de Y
Retome o Exemplo/Exercício 5.37.

(a) Verifique que

$$f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{c.c.} \end{cases} \tag{5.27}$$

$$f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases} \tag{5.28}$$

- **Par aleatório**

  X = largura da peça rectangular (em dm)

  Y = comprimento da peça rectangular (em dm)

- **F.d.p. conjunta de (X, Y)**

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{c.c.,} \end{cases}$$

- **F.d.p. marginal de X**

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \\ &= \begin{cases} \int_x^1 2\,dy = 2(1-x), & 0 < x < 1 \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

- **F.d.p. marginal de Y**

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \\ &= \begin{cases} \int_0^y 2\,dx = 2y, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

(b) Certifique-se que a probabilidade da largura da peça exceder $8cm$ é igual a 0.04. •

Definição 5.41 — F.d.p., f.d., valores esperados e variâncias condicionais (caso contínuo)

- F.d.p. de X condicional a $Y = y$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{ se } f_Y(y) > 0 \tag{5.29}$$

- F.d. de X condicional a $Y = y$

$$\begin{aligned} F_{X|Y=y}(x) &= P(X \le x|Y = y) \\ &= \int_{-\infty}^{x} f_{X|Y=y}(u)du, \ x \in I\!R \end{aligned} \tag{5.30}$$

- Valor esperado de X condicional a $Y = y$

$$E(X|Y = y) = \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)dx \tag{5.31}$$

- Variância de X condicional a $Y = y$

$$\begin{aligned} V(X|Y = y) &= E(X^2|Y = y) - E^2(X|Y = y) \\ &= \int_{-\infty}^{+\infty} x^2\, f_{X|Y=y}(x)dx - \left[\int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)dx\right]^2. \end{aligned} \tag{5.32}$$

Estas funções e parâmetros definem-se do mesmo modo para Y condicional a $X = x$. •

Exemplo 5.42 — Valores esperados condicionais (caso contínuo)
Responda às seguintes questões referentes ao Exemplo/Exercício 5.37.

(a) Verificou-se que o comprimento da peça é igual a $7.5cm$. Obtenha a probabilidade da largura exceder $5cm$.

- **Par aleatório**

  X = largura da peça rectangular (em dm)

  Y = comprimento da peça rectangular (em dm)

- **F.d.p. conjunta de** (X, Y)

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{c.c.,} \end{cases}$$

- **F.d.p. marginal de** Y

$$f_Y(y) \quad = \quad \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$

- **Nova v.a.**

  $X|Y = 0.75$

- **F.d.p. de** X **condicional a** $\{Y = 0.75\}$

$$\begin{aligned} f_{X|Y=0.75}(x) &= \frac{f_{X,Y}(x, 0.75)}{f_Y(0.75)} \\ &= \begin{cases} \frac{2}{2\times y} = \frac{1}{0.75}, & 0 < x < y = 0.75 \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

- **Probabilidade pedida**

$$\begin{aligned} P(X > 0.5|Y = 0.75) &= 1 - P(X \leq 0.5|Y = 0.75) \\ &= 1 - F_{X|Y=0.75}(0.5) \\ &= 1 - \int_{-\infty}^{0.5} f_{X|Y=0.75}(u)\, du \\ &= 1 - \int_{0}^{0.5} \frac{1}{0.75}\, du \\ &= \frac{1}{3}. \end{aligned}$$

(b) Calcule a largura esperada sabendo que o registo do comprimento foi de $8cm$. Compare o valor obtido com a largura esperada e comente.

- **Nova v.a.**

  $X|Y = 0.8$

- **F.d.p. de** X **condicional a** $\{Y = 0.8\}$

$$f_{X|Y=0.8}(x) \quad = \quad \begin{cases} \frac{1}{0.8}, & 0 < x < y = 0.8 \\ 0, & \text{c.c.} \end{cases}$$

- **Valor esperado de X condicional a $\{Y = 0.8\}$**

$$\begin{aligned}
E(X|Y=0.8) &= \int_{-\infty}^{+\infty} x \times f_{X|Y=0.8}(x)\,dx \\
&= \int_0^{0.8} x \times \frac{1}{0.8}\,dx \\
&= \left.\frac{x^2}{0.8\times 2}\right|_0^{0.8} \\
&= 0.4
\end{aligned}$$

- **Valor esperado de X**

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x \times f_X(x)\,dx \\
&= \int_0^1 x \times 2(1-x)\,dx \\
&= 2\left.\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\right|_0^1 \\
&= \frac{1}{3} \\
&\neq E(X|Y=0.8)
\end{aligned}$$

- **Comentário**

  O conhecimento do comprimento da peça influencia claramente o valor esperado da respectiva largura. Pode concluir-se que X e Y SÃO V.A. DEPENDENTES. •

Definição 5.43 — Independência (caso contínuo)
As v.a. contínuas X e Y dizem-se independentes, caso

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y),\ \forall (x,y) \in \mathbb{R}^2, \tag{5.33}$$

i.e., se formos capazes de escrever a f.d.p. conjunta do par aleatório contínuo (X,Y) como o produto das f.d.p. marginais de X e de Y. Equivalentemente, $X \perp\!\!\!\perp Y$ se pudermos factorizar a f.d. conjunta de (X,Y) no produto das f.d. marginais de X e de Y:

$$F_{X,Y}(x,y) = F_X(x) \times F_Y(y),\ \forall (x,y) \in \mathbb{R}^2. \tag{5.34}$$

•

Exemplo/Exercício 5.44 — (In)dependência (caso contínuo)

(a) Prove que as v.a. comprimento e a largura definidas no Exemplo/Exercício 5.37 são dependentes recorrendo para o efeito à equação (5.33).

- **Averiguação da (in)dependência entre X e Y**

  Se por um lado

  $$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{c.c.}, \end{cases}$$

  por outro lado

  $$f_X(x) \times f_Y(y) = \begin{cases} 2(1-x) \times 2y, & 0 < x < 1,\, 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$

  Logo

  $$f_{X,Y}(x,y) \neq f_X(x) \times f_Y(y)),$$

  i.e., X e Y são v.a. DEPENDENTES.

(b) Reescreva a Nota 5.29[10] para o caso contínuo. •

[10]Nesta nota podem encontrar-se algumas consequências da independência entre as v.a. X e Y.

5.3 Covariância e correlação. Propriedades.

Motivação 5.45 — Covariância e correlação
É crucial obter medidas que avaliem a associação entre duas v.a. Esta avaliação pode efectuar-se quer em termos

- absolutos, calculando a covariância entre X e Y, quer em termos
- relativos, determinando a correlação entre X e Y. •

Definição 5.46 — Covariância
Representa-se usualmente por $cov(X,Y)$ [11] e é definida por

$$\begin{aligned} cov(X,Y) &= E\{[X-E(X)]\times[Y-E(Y)]\} \\ &= E(XY)-E(X)E(Y), \end{aligned} \tag{5.35}$$

que no caso discreto se escreve

$$\begin{aligned} cov(X,Y) &= \sum_x\sum_y x\,y\,P(X=x,Y=y) \\ &\qquad -\sum_x x\,P(X=x)\times\sum_y y\,P(Y=y) \end{aligned} \tag{5.36}$$

e no caso contínuo

$$\begin{aligned} cov(X,Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,y\,f_{X,Y}(x,y)\,dydx \\ &\qquad -\int_{-\infty}^{+\infty} x\,f_X(x)\,dx\times\int_{-\infty}^{+\infty} y\,f_Y(y)\,dy. \end{aligned} \tag{5.37}$$

•

Proposição 5.47 — Propriedades da covariância
Se X e Y forem v.a. independentes então a covariância entre X e Y é nula. No entanto, a implicação no sentido inverso não é necessariamente verdadeira. Para além disso, se a covariância entre X e Y for não nula pode concluir-se imediatamente que X e Y são v.a. dependentes. Estes resultados são algumas das propriedades da covariância enunciadas já de seguida:

1. $X\perp\!\!\!\perp Y \Rightarrow cov(X,Y)=0$
2. $cov(X,Y)=0 \not\Rightarrow X\perp\!\!\!\perp Y$

[11] Ou por σ_{XY}.

3. $cov(X,Y) \neq 0 \Rightarrow X \not\perp\!\!\!\perp Y$

4. $cov(X,Y) = cov(Y,X)$

5. $cov(X,X) = V(X)$

6. $cov(aX+b,Y) = a\, cov(X,Y)$

7. $cov(X+Z,Y) = cov(X,Y) + cov(Z,Y)$

8. $cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_j\right) = \sum_{i=1}^n \sum_{j=1}^n cov(X_i,Y_j)$

9. $cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n V(X_i) + 2 \times \sum_{i=1}^n \sum_{j=i+1}^n cov(X_i,X_j)$. •

Exercício 5.48 — Obtenha a covariância entre as v.a. que constituem os pares aleatórios definidos nos Exercícios 5.13 e 5.37. •

Motivação 5.49 — Correlação
A covariância entre X e Y é uma medida absoluta de associação entre estas duas v.a., como tal com algumas desvantagens:

- não é adimensional; e
- o seu valor absoluto deve ser interpretado com alguma cautela já que a covariância pode ser alterada de modo arbitrário ao efectuar-se, por exemplo, uma mudança de escala como se pôde ver na propriedade 6 da Proposição 5.47.

Há pois a necessidade de quantificar em termos relativos[12] a associação entre X e Y à custa de uma medida adimensional e passível de interpretação. •

Definição 5.50 — Correlação
Será representada por $corr(X,Y)$ [13] e é igual a

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}}. \tag{5.38}$$

Caso $corr(X,Y) \neq 0$ (resp. $corr(X,Y) = 0$) as v.a. dizem-se correlacionadas (resp. não correlacionadas). •

[12]Por exemplo, pondo de parte a dispersão de X e Y.
[13]Ou por $\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Proposição 5.51 — **Propriedades da correlação**

Dada a forma como a correlação foi definida à custa da covariância não surpreende que aquela medida de associação relativa "herde" algumas propriedades da covariância já enunciadas:

1. $X \perp\!\!\!\perp Y \Rightarrow corr(X, Y) = 0$

2. $corr(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$

3. $corr(X, Y) \neq 0 \Rightarrow X \not\perp\!\!\!\perp Y$

4. $corr(X, Y) = corr(Y, X)$.

Importa ainda assinalar que:

5. $corr(X, X) = 1$

6. $corr(aX + b, Y) = corr(X, Y)$

7. $-1 \leq corr(X, Y) \leq 1$, para qualquer par de v.a.

8. $corr(X, Y) = -1 \Leftrightarrow Y = aX + b$, $a < 0$

9. $corr(X, Y) = 1 \Leftrightarrow Y = aX + b$, $a > 0$. •

Nota 5.52 — **Interpretação (do sinal) da correlação**

As propriedades 6 e 7 da Proposição 5.51 sugerem que afirmemos que

- a correlação quantifica a associação linear entre X e Y.

Assim, se o valor absoluto de $corr(X, Y)$ estiver próximo de 1 pode dizer-se que a associação entre X e Y é praticamente linear.

Por seu lado, o sinal da correlação entre X e Y deve ser interpretado do seguinte modo:

- caso $corr(X, Y) > 0$ (resp. $corr(X, Y) < 0$), pode afirmar-se que

  se X cresce então Y tem tendência a crescer (resp. decrescer).

São de evitar, no entanto, interpretações abusivas desta medida relativa de associação pois a existência de correlação entre duas v.a. não implica necessariamente uma relação de causa e efeito entre as mesmas. Basta pensar no número de ninhos de cegonhas e o número de nascimentos em cidades escandinavas, ou o número de refrigerantes vendidos e o números de internamentos por desidratação, etc. •

Exemplo 5.53 — Correlação (caso discreto)

Retome o Exemplo 5.30 e determine o coeficiente de correlação entre X e Y. Comente.

- **Par aleatório** (X, Y)

  $X =$ número de portáteis vendidos diariamente

  $Y =$ número de PCs vendidos diariamente

Uma vez que se pretende calcular

$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{V(X)\,V(Y)}},$$

onde $cov(X, Y) = E(XY) - E(X)E(Y)$, serão necessários alguns cálculos auxiliares.

- **F.p. marginal de** X

$$\begin{aligned} P(X = x) &= \sum_{y=0}^{2} P(X = x, Y = y) \\ &= \begin{cases} (0.1 + 0.1 + 0.3) = 0.5, & x = 0 \\ (0.2 + 0.1 + 0.1) = 0.4, & x = 1 \\ (0 + 0.1 + 0) = 0.1, & x = 2 \\ 0, & \text{outros valores de } x \end{cases} \end{aligned}$$

- **Valor esperado de** X

$$\begin{aligned} E(X) &= \sum_{x=0}^{2} x \times P(X = x) \\ &= 0 \times 0.5 + 1 \times 0.4 + 2 \times 0.1 \\ &= 0.6 \end{aligned}$$

- **2o. momento de** X

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{2} x^2 \times P(X = x) \\ &= 0^2 \times 0.5 + 1^2 \times 0.4 + 2^2 \times 0.1 \\ &= 0.8 \end{aligned}$$

- **Variância de** X

$$\begin{aligned} V(X) &= E(X^2) - E^2(X) \\ &= 0.8 - 0.6^2 \\ &= 0.44 \end{aligned}$$

- **F.p. marginal de** Y

$$
\begin{aligned}
P(Y=y) &= \sum_{x=0}^{2} P(X=x, Y=y) \\
&= \begin{cases} (0.1+0.2+0) = 0.3, & y=0 \\ (0.1+0.1+0.1) = 0.3, & y=1 \\ (0.3+0.1+0) = 0.4, & y=2 \\ 0, & \text{outros valores de } y \end{cases}
\end{aligned}
$$

- **Valor esperado de** Y

$$
\begin{aligned}
E(Y) &= \sum_{y=0}^{2} y \times P(Y=y) \\
&= 0 \times 0.3 + 1 \times 0.3 + 2 \times 0.4 \\
&= 1.1
\end{aligned}
$$

- **2o. momento de** Y

$$
\begin{aligned}
E(Y^2) &= \sum_{y=0}^{2} y^2 \times P(Y=y) \\
&= 0^2 \times 0.3 + 1^2 \times 0.3 + 2^2 \times 0.4 \\
&= 1.9
\end{aligned}
$$

- **Variância de** Y

$$
\begin{aligned}
V(Y) &= E(Y^2) - E^2(Y) \\
&= 1.9 - 1.1^2 \\
&= 0.69
\end{aligned}
$$

- **F.p. conjunta de** (X, Y)

X	Y		
	0	1	2
0	0.1	0.1	0.3
1	0.2	0.1	0.1
2	0	0.1	0

- **Momento cruzado de X e Y de ordem** $(1,1)$

$$
\begin{aligned}
E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xy \times P(X=x, Y=y) \\
&= 0 \times 0 \times 0.1 + \ldots + 2 \times 2 \times 0 \\
&= 1 \times 1 \times 0.1 + 1 \times 2 \times 0.1 + 2 \times 1 \times 0.1 + 2 \times 2 \times 0 \\
&= 0.5
\end{aligned}
$$

- **Covariância entre X e Y**

$$
\begin{aligned}
cov(X,Y) &= E(XY) - E(X)E(Y) \\
&= 0.5 - 0.6 \times 1.1 \\
&= -0.16
\end{aligned}
$$

- **Correlação entre X e Y**

$$
\begin{aligned}
corr(X,Y) &= \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} \\
&= \frac{-0.16}{\sqrt{0.44 \times 0.69}} \\
&\simeq -0.29
\end{aligned}
$$

- **Comentário** — Visto que $corr(X,Y) \simeq -0.29 \neq 0$ pode concluir-se que as v.a. são dependentes.

  Para além disso importa notar que $corr(X,Y) < 0$, pelo que as v.a. estão negativamente correlacionadas e apresentam tendência para variarem em sentidos opostos.

  Estando o valor de $|corr(X,Y)|$ afastado de 1, pode adiantar-se que as v.a. não estão linearmente correlacionadas. •

Exemplo 5.54 — Correlação (caso contínuo)

Na afinação de um aparelho é possível controlar duas variáveis contínuas X e Y. No entanto, estas variáveis só podem ser controladas em simultâneo e não individualmente. Com efeito, quando o aparelho está afinado, o par aleatório contínuo (X, Y) tem função densidade de probabilidade conjunta constante no triângulo $0 < x < 1$, $0 < y < x$ e nula no resto do plano, i.e.,

$$f_{X,Y}(x,y) = \begin{cases} k, & 0 < x < 1, \quad 0 < y < x \\ 0, & \text{outros pares } (x,y). \end{cases}$$

Após ter mostrado que $k = 2$, obtenha o coeficiente de correlação entre X e Y e comente o valor obtido.[14]

- **Obtenção de** k

$$\begin{aligned} k \quad : \quad & \begin{cases} f_{X,Y}(x,y) \geq 0 \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy\, dx = 1 \end{cases} \\ & \begin{cases} k \geq 0 \\ \int_0^1 \int_0^x k dy\, dx = 1 \end{cases} \\ & \ldots \\ & k = 2 \end{aligned}$$

- **F.d.p. marginal de** X

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy \\ &= \begin{cases} \int_0^x 2\, dy = 2x, & 0 < x < 1 \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

- **Valor esperado de** X

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \times f_X(x)\, dx \\ &= \int_0^1 x \times 2x\, dx \\ &= \left. \frac{2x^3}{3} \right|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

[14] Adaptado do Exame de 24 de Junho de 2006.

- **2o. momento de X**

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 \times f_X(x)\,dx \\
&= \int_0^1 x^2 \times 2x\,dx \\
&= \left.\frac{2x^4}{4}\right|_0^1 \\
&= \frac{1}{2}
\end{aligned}
$$

- **Variância de X**

$$
\begin{aligned}
V(X) &= E(X^2) - E^2(X) \\
&= \frac{1}{2} - \left(\frac{2}{3}\right)^2 \\
&= \frac{1}{18}
\end{aligned}
$$

- **F.d.p. marginal de Y**

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \\
&= \begin{cases} \int_y^1 2\,dx = 2(1-y), & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}
\end{aligned}
$$

- **Valor esperado de Y**

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty} y \times f_Y(y)\,dy \\
&= \int_0^1 y \times 2(1-y)\,dy \\
&= \left.\left(y^2 - \frac{2y^3}{3}\right)\right|_0^1 \\
&= \frac{1}{3}
\end{aligned}
$$

- **2o. momento de Y**

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{+\infty} y^2 \times f_Y(y)\,dy \\
&= \int_0^1 y^2 \times 2(1-y)\,dy \\
&= \left.\left(\frac{2y^3}{3} - \frac{2y^4}{4}\right)\right|_0^1 \\
&= \frac{1}{6}
\end{aligned}
$$

- **Variância de Y**

$$\begin{aligned} V(Y) &= E(Y^2) - E^2(Y) \\ &= \frac{1}{6} - \left(\frac{1}{3}\right)^2 \\ &= \frac{1}{18} \end{aligned}$$

- **F.d.p. conjunta de (X, Y)**

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < 1, \quad 0 < y < x \\ 0, & \text{outros pares } (x,y). \end{cases}$$

- **Momento cruzado de X e Y de ordem $(1,1)$**

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \times f_{X,Y}(x,y)\, dydx \\ &= \int_0^1 \int_0^x xy \times 2\, dy\, dx \\ &= \int_0^1 xy^2 \Big|_0^x\, dx \\ &= \int_0^1 x^3\, dx \\ &= \frac{1}{4} \end{aligned}$$

- **Covariância entre X e Y**

$$\begin{aligned} cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= \frac{1}{4} - \frac{1}{2} \times \frac{1}{3} \\ &= \frac{1}{36} \end{aligned}$$

- **Correlação entre X e Y**

$$\begin{aligned} corr(X,Y) &= \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} \\ &= \frac{1/36}{\sqrt{1/18 \times 1/18}} \\ &= \frac{1}{2} \end{aligned}$$

- **Comentário** — Dado que $corr(X, Y) = \frac{1}{2} > 0$ pode concluir-se que as v.a. são dependentes e estão positivamente correlacionadas pelo que apresentam tendência para variarem no mesmo sentido.

  Mais, como o valor de $corr(X, Y)$ não está próximo de 1, as v.a. não estão linearmente correlacionadas. •

Exercício 5.55 — Determine e interprete a correlação entre as v.a. que constituem os pares aleatórios definidos no Exemplo 5.13 e no Exemplo/Exercício 5.37. •

5.4 Combinações lineares de variáveis aleatórias.

Motivação 5.56 — Combinações lineares de v.a.
É frequente estarmos interessados em estudar o comportamento probabilístico de

- somas (totais)
- médias (eventualmente ponderadas)

envolvendo diversas v.a. Tratam-se de exemplos daquilo que usualmente se designa de combinações lineares de v.a. •

Definição 5.57 — Combinações lineares de v.a.
Sejam $X_1, \ldots, X_n$ v.a. e $c_1, \ldots, c_n$ constantes reais. Então a v.a.

$$Y = \sum_{i=1}^{n} c_i X_i \tag{5.39}$$

diz-se uma combinação linear das v.a. $X_1, \ldots, X_n$. •

Exemplo 5.58 — Combinações lineares de v.a.

- X_i = número de peças defeituosas em n_i seleccionadas no turno horário i $(i = 1, \ldots, 24)$

  $\sum_{i=1}^{24} X_i$ = número de peças defeituosas em $\sum_{i=1}^{24} n_i$ seleccionadas em 24 turnos horários (um dia)

- X_i = consumo de electricidade no dia i $(i = 1, \ldots, 30)$

  $\sum_{i=1}^{30} X_i$ = consumo mensal de electricidade

  $\frac{1}{30} \sum_{i=1}^{30} X_i$ = consumo médio diário de electricidade. •

A obtenção da f.(d.)p. de uma combinação linear nem sempre é tarefa fácil. Pode, no entanto, adiantar-se o valor esperado de qualquer combinação linear de v.a. e a variância de combinações lineares em situações particulares.

Proposição 5.59 — Valor esperado de combinação linear de v.a.

$$E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i). \tag{5.40}$$

Em particular, se $c_i = 1$, para $i = 1, \ldots, n$, tem-se

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i). \tag{5.41}$$

•

Embora o valor esperado de uma combinação linear de v.a. seja igual à combinação linear dos valores esperados o mesmo está longe de acontecer com a variância de uma combinação linear.

Proposição 5.60 — **Variância de algumas combinações lineares de v.a.**

$$V\left(c_1 X_1 + c_2 X_2\right) = c_1^2 V(X_1) + c_2^2 V(X_2) + 2c_1c_2 cov(X_1, X_2) \tag{5.42}$$

$$V\left(X_1 + X_2\right) = V(X_1) + V(X_2) + 2cov(X_1, X_2) \tag{5.43}$$

$$V\left(X_1 - X_2\right) = V(X_1) + V(X_2) - 2cov(X_1, X_2) \tag{5.44}$$

$$V\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 V(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} c_i\, c_j\, cov(X_i, X_j). \tag{5.45}$$

Ao lidar-se com v.a. não correlacionadas duas a duas — i.e., se $cov(X_i, X_j) = 0,\ \forall i \neq j$ — ou com v.a. independentes duas a duas — ou seja, $X_i \perp\!\!\!\perp X_j,\ \forall i \neq j$ —, tem-se:

$$V\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 V(X_i). \tag{5.46}$$

E se, para além de não correlacionadas ou independentes duas a duas, tivermos $c_i = 1$, para $i = 1, \ldots, n$, segue-se

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i). \tag{5.47}$$

•

Exemplo 5.61 — **Combinações lineares de v.a.** *(e não só)*
Uma componente mecânica é constituída por uma biela e uma manivela cujas massas são normalmente distribuídas com parâmetros na tabela seguinte:

Peça	Massas	
	Valor esperado	Desvio-Padrão
biela	1	0.02
manivela	0.8	0.05

Calcule a covariância entre as massas da biela e da manivela, sabendo que a variância da massa total da componente mecânica é de $0.0049 Kg^2$.[15]

- **Par aleatório** (X, Y)

  X = massa da biela

  Y = massa da manivela

[15] Adaptado do Teste B de 11 de Maio de 2002.

- **Distribuições de X e Y**

  $X \sim \text{normal}(\mu_X, \sigma_X^2)$

  $Y \sim \text{normal}(\mu_Y, \sigma_Y^2)$

- **Parâmetros**

  $\mu_X = 1$, $\sigma_X^2 = 0.02^2$

  $\mu_Y = 0.8$, $\sigma_Y^2 = 0.05^2$

- **Nova v.a.**

  $Z = X + Y =$ massa total da componente mecânica

- **Variância de Z**

  $V(Z) = 0.0049 Kg^2$

- **Covariância entre X e Y**

$$\begin{aligned} V(Z) &= V(X) + V(Y) + 2 \times cov(X,Y) \\ 0.0049 &= 0.02^2 + 0.05^2 + 2 \times cov(X,Y) \\ cov(X,Y) &= \frac{0.0049 - 0.02^2 - 0.05^2}{2} \end{aligned}$$

  Assim, conclui-se que $cov(X,Y) = 0.0001$. •

Exercício 5.62 — Combinações lineares de v.a.
Sejam $X_1, \ldots, X_n$ v.a. independentes e identicamente distribuídas (i.i.d.) à v.a. X — i.e., $X_i \sim_{i.i.d.} X$, $i = 1, \ldots, n$ — tais que

$$E(X_i) = E(X) = \mu \tag{5.48}$$

$$V(X_i) = V(X) = \sigma^2, \tag{5.49}$$

para $i = 1, \ldots, n$. Demonstre que

Combinação linear		Valor esperado	Variância
Total	$\sum_{i=1}^n X_i$	$n\mu$	$n\sigma^2$
Média aritmética	$\frac{1}{n}\sum_{i=1}^n X_i$	μ	$\frac{\sigma^2}{n}$

•

Motivação 5.63 — **Casos especiais de somas de v.a. independentes**
Vimos até agora que é possível adiantar o valor esperado e a variância de somas de v.a. Resta saber se somos capazes de identificar a distribuição dessas mesmas somas. A resposta é afirmativa nalguns casos especiais de somas de v.a. independentes que passamos a descrever na proposição seguinte. •

Proposição 5.64 — **Casos especiais de somas de v.a. independentes; mudança de escala**

V.a.	Combinação linear	Obs.
$X_i \sim_{indep}$ binomial(n_i, p), $i = 1, \dots, k$	$\sum_{i=1}^{k} X_i \sim$ binomial $\left(\sum_{i=1}^{k} n_i, p\right)$	(1)
$X_i \sim_{indep}$ Poisson(λ_i), $i = 1, \dots, n$	$\sum_{i=1}^{n} X_i \sim$ Poisson $\left(\sum_{i=1}^{n} \lambda_i\right)$	(2)
$X_i \sim_{indep.}$ normal(μ_i, σ_i^2), $i = 1, \dots, n$	$\sum_{i=1}^{n} X_i \sim$ normal $\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$	(3)
	$\sum_{i=1}^{n} c_i\, X_i \sim$ normal $\left(\sum_{i=1}^{n} c_i\, \mu_i, \sum_{i=1}^{n} c_i^2\, \sigma_i^2\right)$	(4)
$X \sim$ exponencial(λ)	$cX \sim$ exponencial $\left(\frac{\lambda}{c}\right)$	(5)

A tabela traduz o fecho de algumas famílias de distribuições para operações como a soma, as combinações lineares ou a mudança de escala:

(1) Propriedade reprodutiva da distribuição binomial, ao lidar-se com v.a. binomiais independentes com probabilidade de sucesso comum;

(2) Propriedade reprodutiva da distribuição de Poisson, ao lidar-se com v.a. de Poisson independentes;

(3) Propriedade reprodutiva da distribuição normal, ao lidar-se com v.a. normais independentes;

(4) A combinação linear de normais — independentes mas não necessariamente identicamente distribuídas — ainda tem distribuição normal;

(5) Mudança de escala da distribuição exponencial. •

Exemplo 5.65 — Casos especiais de somas de v.a. discretas independentes
Os números de "kits" de teste vendidos semanalmente por duas sucursais de uma empresa de biotecnologia são duas v.a. independentes com distribuição de Poisson cujas variâncias são iguais a 10 e 15, respectivamente.[16]

(a) Obtenha o valor exacto para a probabilidade de o número total de "kits"de teste vendidos semanalmente pelas duas sucursais da empresa exceder 25 unidades.

- **V.a.**
  $X_i =$ número de "kits"vendidos semanalmente pela sucursal i $(i = 1, 2)$
- **Distribuição de X_i**
  $X_i \overset{indep}{\sim} \text{Poisson}(\lambda_i), i = 1, 2$
- **Parâmetros**
  $\lambda_1 = E(X_1) = V(X_1) = 10$
  $\lambda_2 = E(X_2) = V(X_2) = 15$
- **Nova v.a.**
  $Y = X_1 + X_2 =$ número total de "kits" de teste vendidos semanalmente pelas duas sucursais
- **Distribuição de Y**
  Tratando-se da soma de duas v.a. independentes com distribuição de Poisson, pode afirmar-se que
  $$Y \sim \text{Poisson}(E(Y)).$$
- **Parâmetro**
  $E(Y) = E(X_1) + E(X_2) = 10 + 15 = 25$
- **Probabilidade pedida**
  $$\begin{aligned} P(Y > 25) &= 1 - P(Y \leq 25) \\ &= 1 - F_{Poisson(25)}(25) \\ &\overset{tabela}{=} 1 - 0.5529 \\ &= 0.4471. \end{aligned}$$

[16] Adaptado do Exame de 4 de Fevereiro de 2003.

(b) Admita que as sucursais facturam 200 e 225 Euros (respectivamente) por cada "kit" de teste vendido e determine o valor esperado e o desvio-padrão da facturação total semanal das duas sucursais desta empresa.

- **Nova v.a.**
  $F = 200 \times X_1 + 225 \times X_2$ =facturação total semanal das duas sucursais
- **Valor esperado de F**
  $E(F) = 200 \times E(X_1) + 225 \times E(X_2) = 200 \times 10 + 225 \times 15 = 5375$
- **Variância de F**
  $V(F) = 200^2 \times V(X_1) + 225^2 \times V(X_2) = 200^2 \times 10 + 225^2 \times 15 = 1159375$
- **Desvio-padrão de F**
  $DP(F) = +\sqrt{V(F)} = +\sqrt{1159375} \simeq 1076.74.$ •

Exemplo 5.66 — Casos especiais de somas de v.a. contínuas independentes
Engenheiros acreditam que um troço de uma ponte é capaz de suportar $W = 200$ toneladas sem que sejam causados danos estruturais. Para além disso, estudos levados a cabo apontam para que a massa (em tonelada) dos carros que por ela circulam seja bem modelada por uma v.a. com distribuição normal de valor esperado 1.5 e desvio-padrão 0.15.[17]

(a) Qual a probabilidade de ocorrerem danos estruturais caso estejam 130 carros nesse troço?

- **V.a.**
  $X_i =$ massa do $i-$ésimo carro, $i = 1, \ldots, 130$
- **Distribuição de X_i**
  $X_i \stackrel{i.i.d.}{\sim}$ normal(μ, σ^2), $i = 1, \ldots, 130$
- **Parâmetros**
  $\mu = E(X_i) = 1.5$
  $\sigma^2 = V(X_i) = 0.15^2$
- **Nova v.a.**
  $Y = \sum_{i=1}^{130} X_i =$ massa total dos 130 carros

[17] Adaptado do Exame de 13 de Julho de 2002.

- **Distribuição de Y**

  Tratando-se Y de uma soma de v.a. i.i.d. com distribuição normal, tem-se

  $$Y \sim \text{normal}(E(Y), V(Y)).$$

- **Parâmetros**

  $E(Y) = E\left(\sum_{i=1}^{130} X_i\right) = 130\mu = 195$

  $V(Y) = V\left(\sum_{i=1}^{130} X_i\right) = 130\sigma^2 = 2.925$

- **Probabilidade pedida**

$$\begin{aligned} P(Y > W = 200) &= 1 - P(Y \leq 200) \\ &= 1 - \Phi\left(\frac{200 - E(Y)}{\sqrt{V(Y)}}\right) \\ &\simeq 1 - \Phi(2.92) \\ &\stackrel{tabela}{=} 1 - 0.9982 \\ &= 0.0018. \end{aligned}$$

(b) Assuma agora que W é também uma v.a. com distribuição normal com valor esperado 200 e desvio-padrão 20. Demonstre que, caso o número de carros que estão nesse troço da ponte seja igual a 140, a probabilidade de ocorrência de danos estruturais é de pelo menos 0.1?

- **Novas v.a.**

  $T = \sum_{i=1}^{140} X_i =$ massa total dos 140 carros

  $W =$ carga máxima

  Importa notar que a probabilidade pedida é $P(T > W) = P(T - W > 0)$ pelo que é conveniente lidar com a v.a. $T - W$, que não passa de uma combinação linear de duas v.a. independentes com distribuição normal logo também normalmente distribuída.

- **Distribuições de T, W e $T - W$**

  $T \sim \text{normal}(E(T), V(T))$

  $W \sim \text{normal}(E(W), V(W))$

  $T - W \sim \text{normal}(E(T - W), V(T - W))$

- **Parâmetros**

  $E(T) = E\left(\sum_{i=1}^{140} X_i\right) = 140\mu = 210$

  $V(T) = V\left(\sum_{i=1}^{140} X_i\right) = 140\sigma^2 = 3.15$

$E(W) = 200$

$V(W) = 20^2$

$E(T - W) = E(T) - E(W) = 210 - 200 = 10$

$V(T - W) = V(T) + V(W) = 3.15 + 20^2 = 403.15$

- **Probabilidade pedida**

$$\begin{aligned} P(T > W) &= 1 - P(T - W \leq 0) \\ &= 1 - \Phi\left(\frac{0 - E(T-W)}{\sqrt{V(T-W)}}\right) \\ &\simeq 1 - \Phi(-0.50) \\ &= \Phi(0.50) \\ &\overset{tabela}{=} 0.6915. \end{aligned}$$

•

Exemplo 5.67 — A carga máxima de um elevador é de $350Kg$. Admita que a massa de um passageiro é uma v.a. $X \sim \text{normal}(75, 10^2)$. Obtenha a probabilidade de a massa total de 5 passageiros exceder a carga máxima.

- **V.a.**

  $X_i =$ massa do $i-$ésimo passageiro, $i = 1, \ldots, 5$

- **Distribuição de X_i**

  $X_i \overset{i.i.d.}{\sim} \text{normal}(\mu, \sigma^2)$, $i = 1, \ldots, 5$

- **Parâmetros**

  $\mu = E(X_i) = 75$

  $\sigma^2 = V(X_i) = 10^2$

- **Nova v.a.**

  $Y = \sum_{i=1}^{5} X_i =$ massa total dos 5 passageiros

- **Distribuição de Y**

  Já que Y é uma soma de v.a. i.i.d. com distribuição normal, $Y \sim \text{normal}(E(Y), V(Y))$.

- **Parâmetros**

  $E(Y) = E\left(\sum_{i=1}^{5} X_i\right) = 5\mu = 375$

  $V(Y) \stackrel{X_i\ indep}{=} V\left(\sum_{i=1}^{5} X_i\right) = 5\sigma^2 = 500$

- **Probabilidade pedida**

$$
\begin{aligned}
P(Y > 350) &= 1 - \Phi\left(\frac{350 - E(Y)}{\sqrt{V(Y)}}\right) \\
&\simeq 1 - \Phi(-1.12) \\
&\simeq \Phi(1.12) \\
&\stackrel{tabela}{=} 0.8686.
\end{aligned}
$$

- **Comentário**

  Esta probabilidade é elevadíssima, pelo que não é recomendável que o elevador transporte 5 passageiros. •

5.5 Desigualdade de Chebychev.

Motivação 5.68 — Desigualdade de Chebychev
A desigualdade de Chebychev[18] permite adiantar limites superiores ou inferiores para probabilidades de eventos que digam respeito a uma v.a. discreta ou contínua X, bastando para tal conhecer o seu valor esperado e a sua variância. Este resultado é particularmente conveniente quando não somos capazes de identificar a distribuição, por exemplo, de uma soma de v.a. •

Proposição 5.69 — Desigualdade de Chebychev
Seja X uma v.a. (discreta ou contínua) tal que

$$\begin{aligned} E(X) &= \mu \\ V(X) &= \sigma^2. \end{aligned}$$

Então

$$P(|X-\mu| \geq c\,\sigma) \leq \frac{1}{c^2}. \tag{5.50}$$

•

Nota 5.70 — Desigualdade de Chebychev
Esta desigualdade só faz sentido para $c > 0$ e só é útil caso $c > 1$. •

Exemplo 5.71 — Desigualdade de Chebychev
Retome o Exercício 5.67.

(a) Determine um limite para $P(45 < X < 105)$ admitindo que desconhece a distribuição da massa X de um passageiro mas que o valor esperado e o desvio-padrão são conhecidos e iguais a $75Kg$ e $10Kg$, respectivamente.

- **V.a.**
  X =massa do passageiro
- **Distribuição de X**
  DESCONHECIDA
- **Valor esperado**
  $\mu = E(X) = 75Kg$

[18] Pafnuty Lvovich Chebychev (1821–1894).

- **Desvio-padrão**
  $\sigma = \sqrt{V(X)} = 10Kg$
- **Limite para** $P(45 < X < 105)$
  Comece-se por notar que

$$\begin{aligned} P(45 < X < 105) &= P(|X - 75| < 30) \\ &= 1 - P(|X - \mu| \geq 3\sigma) \end{aligned}$$

  Ora recorrendo à desigualdade de Chebychev segue-se

$$\begin{aligned} P(|X - \mu| \geq 3\sigma) &\leq \frac{1}{3^2} \\ 1 - P(|X - \mu| \geq 3\sigma) &\geq 1 - \frac{1}{3^2} \\ P(45 < X < 105) &\geq \frac{8}{9}. \end{aligned}$$

(b) Compare o limite que obteve em (a) com o verdadeiro valor da probabilidade, caso tivesse assumido que a massa de um passageiro é normalmente distribuída.

- **Distribuição de** X
  $X \sim \text{normal}(\mu = 75, \sigma^2 = 10^2)$
- **Valor exacto de** $P(45 < X < 105)$

$$\begin{aligned} P(45 < X < 105) &= \Phi\left(\frac{105 - 75}{10}\right) - \Phi\left(\frac{45 - 75}{10}\right) \\ &= \Phi(3) - \Phi(-3) \\ &= 2 \times \Phi(3) - 1 \\ &= 0.9973 \\ &> \frac{8}{9} \end{aligned}$$

- **Comentário** — O limite inferior obtido usando a desigualdade de Chebychev subestima o valor da probabilidade em

$$\frac{|8/9 - 0.9973|}{0.9973} \times 100\% \simeq 10.87\%,$$

  caso a verdadeira distribuição da massa de um passageiro seja normal$(75, 10^2)$. •

5.6 Teorema do Limite Central. Aplicações às distribuições binomial e de Poisson.

Motivação 5.72 — **Teorema do Limite Central**
Nem todas as distribuições gozam da propriedade reprodutiva das distribuições binomial, de Poisson e normal, pelo que é crucial arranjar formas de obter valores aproximados de probabilidades de eventos envolvendo

- Somas $(\sum_{i=1}^{n} X_i)$
- Médias aritméticas $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$

de v.a. independentes e identicamente distribuídas (i.i.d.). A resposta pode encontrar-se no Teorema do Limite Central (TLC). •

Teorema 5.73 — **Teorema do Limite Central**
Considere-se que, para todo o inteiro positivo n, as v.a. $X_1, \ldots, X_n$ são i.i.d. com valor esperado μ e variância finita σ^2. Considere-se ainda

- $S_n = \sum_{i=1}^{n} X_i$ (soma)
- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (média aritmética).

Então

$$\begin{aligned}\lim_{n\to+\infty} P\left[\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leq z\right] &= \lim_{n\to+\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) \\ &= \Phi(z). \end{aligned} \tag{5.51}$$

Equivalentemente,

$$\begin{aligned}\lim_{n\to+\infty} P\left[\frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}} \leq z\right] &= \lim_{n\to+\infty} P\left(\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq z\right) \\ &= \Phi(z). \end{aligned} \tag{5.52}$$

•

Nota 5.74 — Interesse prático do Teorema do Limite Central
Ao lidarmos com

- n v.a. i.i.d. ($E(X_i) = \mu$, $V(X_i) = \sigma^2 < +\infty$) e
- n suficientemente grande[19]

podemos

- aproximar a f.d. da soma ou da média aritmética dos X_i's (S_n e $\overline{X}_n$, resp.) devidamente padronizadas

pela

- f.d. de v.a. normal padrão.

Com efeito, pelo facto de o TLC permitir afirmar que

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \stackrel{a}{\sim} \text{normal}(0,1) \tag{5.53}$$

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \stackrel{a}{\sim} \text{normal}(0,1) \tag{5.54}$$

(onde "$\stackrel{a}{\sim}$" se lê "tem distribuição aproximadamente..."), tem-se

$$\begin{aligned} P(S_n \leq s) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{s - n\mu}{\sqrt{n\sigma^2}}\right) \\ &\stackrel{TLC}{\simeq} \Phi\left(\frac{s - n\mu}{\sqrt{n\sigma^2}}\right) \end{aligned} \tag{5.55}$$

$$\begin{aligned} P(\overline{X}_n \leq x) &= P\left(\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq \frac{x - \mu}{\sqrt{\sigma^2/n}}\right) \\ &\stackrel{TLC}{\simeq} \Phi\left(\frac{x - \mu}{\sqrt{\sigma^2/n}}\right). \end{aligned} \tag{5.56}$$

•

Nota 5.75 — Teorema do Limite Central
O TLC é válido para somas e médias aritméticas de v.a. quer discretas, quer contínuas.

Este teorema justifica também as aproximações normais das distribuições binomial e de Poisson que descreveremos ainda nesta secção. Para tal basta considerar:

- $X_i \sim_{i.i.d.} \text{Bernoulli}(p), i = 1, \ldots, n \quad \rightarrow \quad S_n = \sum_{i=1}^n X_i \sim \text{binomial}(n, p)$
- $X_i \sim_{i.i.d.} \text{Poisson}(\lambda/n), i = 1, \ldots, n \quad \rightarrow \quad S_n = \sum_{i=1}^n X_i \sim \text{Poisson}(\lambda)$. •

[19] n suficientemente grande significa, de um modo geral, $n \geq 30$.

Exemplo 5.76 — Teorema do Limite Central (v.a. discretas)
Considere um modelo simples para o tráfego na *world wide web* em que o número de pacotes necessários para a transmissão de uma página na *net* distribui-se uniformemente em $\{1,\ldots,50\}$. Calcule um valor aproximado para a probabilidade do número de pacotes necessários à transmissão de 100 páginas exceder 3000.[20]
Nota: $\sum_{x=1}^{n} x = \frac{n(n+1)}{2}$ e $\sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}$.

- **V.a.**

  $X_i =$ número de pacotes necessários para a transmissão da página i, $i = 1,\ldots,100$

- **Distribuição de X_i**

  $X_i \stackrel{i.i.d.}{\sim}$ uniforme discreta$(\{1,2,\ldots,50\}), i = 1,\ldots,100$

- **F.p. de X_i**

$$P(X_i = x) = \begin{cases} \frac{1}{50}, & x = 1,2,\ldots,50 \\ 0, & \text{c.c.} \end{cases}$$

- **Valor esperado de X_i**

$$\begin{aligned} \mu &= E(X_i) \\ &= \sum_{x=1}^{50} x \times \frac{1}{50} \\ &= \frac{1}{50} \times \frac{50 \times (50+1)}{2} \\ &= 25.5 \end{aligned}$$

- **2o. momento de X_i**

$$\begin{aligned} E(X_i^2) &= \sum_{x=1}^{50} x^2 \times \frac{1}{50} \\ &= \frac{1}{50} \times \frac{50 \times (50+1)(2 \times 50+1)}{6} \\ &= 858.5 \end{aligned}$$

- **Variância de X_i**

$$\begin{aligned} \sigma^2 &= V(X_i) \\ &= E(X_i^2) - E^2(X_i) \\ &= 858.5 - 25.5^2 \\ &= 208.25 \end{aligned}$$

[20] Adaptado do Exame de 17 de Janeiro de 2004.

- **Nova v.a.**

  $Y = \sum_{i=1}^{100} X_i =$ número total de pacotes necessários para a transmissão 100 páginas

- **Valor esperado de Y**

  $E(Y) = 100 \times \mu = 2550$

- **Variância de Y**

  $V(Y) = 100 \times \sigma^2 = 20825$

- **Distribuição aproximada de Y**

  De acordo com o TLC pode adiantar-se que

$$\frac{Y - E(Y)}{\sqrt{V(Y)}} \stackrel{a}{\sim} \text{normal}(0,1).$$

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned} P(Y > 3000) &= 1 - P(Y \leq 3000) \\ &\stackrel{TLC}{\simeq} 1 - \Phi\left(\frac{3000 - 2550}{\sqrt{20825}}\right) \\ &\simeq 1 - \Phi(3.12) \\ &\stackrel{tabela}{=} 1 - 0.9991 \\ &= 0.0009. \end{aligned}$$

•

Exemplo/Exercício 5.77 — Teorema do Limite Central (v.a. contínuas)
Suponha-se que, ao adicionar números reais, cada número é arredondado previamente para o inteiro mais próximo. Admita-se que os erros de arredondamento são v.a. i.i.d. com distribuição comum

$$\text{uniforme}(-0.5, 0.5). \tag{5.57}$$

(a) Obtenha um valor aproximado para a probabilidade de o valor absoluto do erro da soma exceder 15 unidades ao adicionarmos 1500 números.

- **V.a.**

  $X_i =$ erro de arredondamento da parcela i, $i = 1, \ldots, 1500$

- **Distribuição de** X_i

  $X_i \stackrel{i.i.d.}{\sim}$ uniforme$(-0.5, 0.5)$, $i = 1, \ldots, 1500$

- **F.d.p. de** X_i

$$f_{X_i}(x) \stackrel{form}{=} \begin{cases} \frac{1}{0.5-(-0.5)} = 1, & -0.5 \leq x \leq 0.5 \\ 0, & \text{c.c.} \end{cases}$$

- **Valor esperado de** X_i

$$\begin{aligned} \mu &= E(X_i) \\ &\stackrel{form}{=} \frac{0.5 + (-0.5)}{2} \\ &= 0 \end{aligned}$$

- **Variância de** X_i

$$\begin{aligned} \sigma^2 &= V(X_i) \\ &\stackrel{form}{=} \frac{[0.5 - (-0.5)]^2}{12} \\ &= \frac{1}{12} \end{aligned}$$

- **Nova v.a.**

  $Y = \sum_{i=1}^{1500} X_i$ =erro de arredondamento da soma de 1500 parcelas

- **Valor esperado de** Y

  $E(Y) = 1500 \times \mu = 0$

- **Variância de** Y

  $V(Y) = 1500 \times \sigma^2 = \frac{1500}{12}$

- **Distribuição aproximada de** Y

  Pelo TLC pode escrever-se

$$\frac{Y - E(Y)}{\sqrt{V(Y)}} \stackrel{a}{\sim} \text{normal}(0, 1).$$

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned} P(|Y| > 15) &= 1 - P(-15 \leq Y \leq 15) \\ &\stackrel{TLC}{\simeq} 1 - \left[\Phi\left(\frac{15-0}{\sqrt{\frac{1500}{12}}}\right) - \Phi\left(\frac{-15-0}{\sqrt{\frac{1500}{12}}}\right)\right] \\ &\simeq 1 - [\Phi(1.34) - \Phi(-1.34)] \\ &= 2 \times [1 - \Phi(1.34)] \\ &\stackrel{tabela}{=} 2 \times (1 - 0.9099) \\ &= 0.1802. \end{aligned}$$

(b) Compare este valor aproximado com o limite inferior obtido por recurso à desigualdade de Chebychev. •

Motivação 5.78 — **Aproximações de distribuições**
Nem sempre é possível calcular a probabilidade de certos eventos recorrendo às tabelas disponíveis ou de uma outra forma expedita sem recurso a uma calculadora ou a um computador. Basta considerar-se o cálculo de $P(X \leq 40)$ onde $X \sim \text{binomial}(100, 0.1)$.

Não só as tabelas disponíveis não contemplam valores de n superiores a 20 como a obtenção de

$$P(X \leq 40) = \sum_{x=0}^{40} \frac{100!}{x!\,(100-x)!} 0.1^x (1-0.1)^{100-x}$$

pressupõe o cálculo de 41 parcelas.

Não surpreende que nestas e noutras situações similares seja frequente recorrer ao que usualmente se denomina de

- aproximações distribucionais

entre elas as duas que constam do título a secção. •

Importa referir o carácter empírico das aproximações das distribuições que se seguem e o facto de as condições em que se deve efectuar qualquer das 4 aproximações que veremos já de seguida não serem rígidas e variarem de livro de texto para livro de texto.

Nota 5.79 — 4 aproximações de distribuições

V.a. original	V.a. aproximativa	Condições de aproximação	Obs.
1. $X \sim$ hipergeométrica(N, M, n)	$\tilde{X} \sim$ binomial$(n, M/N)$	$n < 0.1N$	(1)
2. $X \sim$ binomial(n, p)	$\tilde{X} \sim$ Poisson(np)	$n > 20$ e $p < 0.1$	(2)
3. $X \sim$ binomial(n, p)	$\tilde{X} \sim$ normal$(np, np(1-p))$	$np > 5$ e $n(1-p) > 5$	(3)
4. $X \sim$ Poisson(λ)	$\tilde{X} \sim$ normal(λ, λ)	$\lambda > 5$	(4)

(1) Fazer ou não reposição deixa de ser relevante quando a dimensão da amostra é muito pequena quando comparada com a dimensão da população.

(2) Não estão disponíveis tabelas para a f.d. da binomial com $n > 20$.

(3) Por estarmos a aproximar uma v.a. discreta por uma contínua e por forma a melhorar a aproximação recomenda-se a efectuação do que se designa por *correcção de continuidade* que descreveremos daqui a pouco.[21]

(4) Pelos mesmos motivos apontados acima é recomendável efectuar a correcção de continuidade. Atente-se que existem tabelas da f.d. da v.a. $X \sim Poisson(\lambda)$ para diversos valores de $\lambda > 5$, e nesses casos não há qualquer necessidade de efectuar a aproximação.[22] •

Nota 5.80 — 4 aproximações de distribuições

As 4 aproximações de distribuições encontram justificação em resultados de convergência:

1. $X \sim$ hipergeométrica(N, M, n) aproximada por $\tilde{X} \sim$ binomial$(n, p = M/N)$.

   Para n e $M/N = p$ fixos, tem-se:

$$\lim_{\substack{N \to +\infty \\ n \text{ fixo} \\ M/N = p \text{ fixo}}} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}. \tag{5.58}$$

2. $X \sim$ binomial(n, p) aproximada por $\tilde{X} \sim$ Poisson$(\lambda = np)$.

   Para $np = \lambda$ fixo, segue-se:

$$\lim_{\substack{n \to +\infty \\ p \to 0 \\ np = \lambda \text{ fixo}}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}. \tag{5.59}$$

[21] A aplicação de correcção de continuidade não é obrigatória a menos que a solicitemos.

[22] E muito menos a correcção de continuidade.

3./4. $X \sim \text{binomial}(n,p)$ aproximada por $\tilde{X} \sim \text{normal}(\mu = np, \sigma^2 = np(1-p))$.

$X \sim \text{Poisson}(\lambda)$ aproximada por $\tilde{X} \sim \text{normal}(\mu = \lambda, \sigma^2 = \lambda)$.

Estas duas aproximações encontram justificação no Teorema do Limite Central.

Refira-se ainda que as 4 aproximações de distribuições preservam o valor esperado da v.a. original e a variância nos casos 3. e 4. Com efeito temos: $E(X) = E(\tilde{X})$, em qualquer dos 4 casos; $V(X) \simeq V(\tilde{X})$, para as duas primeiras 4 aproximações de distribuições (1. e 2.); e $V(X) = V(\tilde{X})$, para as restantes duas aproximações (3. e 4.). •

Exemplo 5.81 — A aproximação binomial da distribuição hipergeométrica
Para avaliar a correcção da composição anunciada de certo produto alimentar foram analisadas 10000 embalagens desse produto. Verificou-se que em 100 delas o produto não correspondia ao anunciado.

Indique uma expressão e um valor aproximado para a probabilidade de pelo menos duas embalagens do produto não corresponderem ao anunciado, caso tenham sido seleccionadas 500 embalagens ao acaso sem reposição daquelas 10000 embalagens do produto.[23]

- **V.a. original**

  X = número de embalagens que não correspondem ao anunciado, numa amostra de 500 embalagens seleccionadas ao acaso sem reposição de entre 10000 embalagens do produto das quais 100 não correspondem ao anunciado

- **Distribuição de X**

  $X \sim \text{hipergeométrica}(N, M, n)$

- **Parâmetros**

  $N = 10000$ embalagens do produto

  $M = 100$ embalagens do produto que não correspondem ao anunciado

  $n = 500$ embalagens seleccionadas

- **F.p. de X**

$$P(X = x) = \begin{cases} \frac{\binom{100}{x}\binom{10000-100}{500-x}}{\binom{10000}{500}}, & x = 0, \ldots, 100 \\ 0, & \text{c.c.} \end{cases}$$

[23] Adaptado do Exame de 7 de Fevereiro de 2004.

- **Expressão da probabilidade pedida**

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - \sum_{x=0}^{1} \frac{\binom{100}{x}\binom{10000-100}{500-x}}{\binom{10000}{500}}. \end{aligned}$$

- **V.a. aproximativa $\tilde{X}$**

  Dado que $n = 500 < 0.1N = 1000$ pode aproximar-se a f.p. da v.a. $X \sim \text{hipergeométrica}(N = 10000, M = 100, n = 500)$ pela f.p. da v.a.

$$\tilde{X} \sim \text{binomial}(n = 500, p = M/N = 0.01).$$

- **F.p. de $\tilde{X}$**

$$P(\tilde{X} = x) = \binom{500}{x} 0.01^x \, (1 - 0.01)^{500-x}, \; x = 0, 1, \ldots, 500$$

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &\simeq 1 - P(\tilde{X} \leq 1) \\ &= 1 - \sum_{x=0}^{1} \binom{500}{x} 0.01^x \, (1 - 0.01)^{500-x} \\ &= 1 - \left(0.99^{500} + 500 \times 0.01 \times 0.99^{499}\right) \\ &= 0.9602. \end{aligned}$$

•

Exemplo/Exercício 5.82 — A aproximação de Poisson da distribuição binomial

Um fabricante de computadores garante a substituição de todos os computadores que avariem durante o primeiro ano após a data de compra ou da substituição. Admite-se que os tempos de vida destes computadores são v.a. independentes e que seguem uma distribuição normal com valor esperado 3.526 anos e desvio padrão 0.85 anos.[24]

(a) Calcule a proporção de computadores que o fabricante pode ter que substituir.

  - **V.a.**

    X = tempo de vida de computador...

[24] Adaptado do Exame de 5 de Fevereiro de 2002.

- **Distribuição de X**

  $X \sim \text{normal}(\mu, \sigma^2)$

- **Parâmetros**

  $\mu = E(X) = 3.526$ anos

  $\sigma^2 = V(X) = 0.85^2$ anos^2

- **Probabilidade pedida**

  Seja S o evento que representa a substituição devido a avaria durante o primeiro ano. Então

$$
\begin{aligned}
P(S) &= P(X < 1) \\
&= P\left[\frac{X - E(X)}{\sqrt{V(X)}} < \frac{1 - E(X)}{\sqrt{V(X)}}\right] \\
&= \Phi\left[\frac{1 - E(X)}{\sqrt{V(X)}}\right] \\
&= \Phi\left(\frac{1 - 3.516}{0.85}\right) \\
&\simeq \Phi(-2.96) \\
&= 1 - \Phi(2.96) \\
&\stackrel{tabela}{=} 0.0015.
\end{aligned}
$$

(b) Se uma firma adquirir 100 computadores àquele fabricante, qual a probabilidade de 3 ou mais desses computadores serem substituídos?

- **Nova v.a. original**

  $Y =$ número de computadores que serão substituídos de entre 100

- **Distribuição de Y**

  $Y \sim \text{binomial}(n, p)$

- **Parâmetros**

  $n = 100$

  $p = P(S) = 0.0015$

- **F.p. de Y**

$$P(Y = y) = \binom{100}{y}\, 0.0015^y\, (1 - 0.0015)^{100-y},\ y = 0, 1, \ldots, 100$$

- **V.a. aproximativa $\tilde{Y}$**

  Dado que $n > 20$ e $p < 0.1$ pode aproximar-se a f.p. da v.a. $Y \sim \text{binomial}(n = 100, p = 0.0015)$ pela f.p. da v.a.

  $$\tilde{Y} \sim \text{Poisson}(\lambda = np = 0.15).$$

- **F.p. de $\tilde{Y}$**

  $$P(\tilde{Y} = y) = e^{-0.15}\,\frac{0.15^y}{y!},\ y = 0, 1, 2, \ldots$$

- **Valor aproximado da probabilidade pedida**

  $$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) \\ &\simeq 1 - P(\tilde{Y} \leq 2) \\ &= 1 - F_{Poisson(0.15)}(2) \\ &\stackrel{tabela}{=} 1 - 0.9995 \\ &= 0.0005. \end{aligned}$$

(c) Determine o valor exacto de $P(Y \geq 3)$ e calcule o erro relativo associado ao valor aproximado calculado na alínea anterior. •

Motivação 5.83 — Correcção de continuidade
A correcção de continuidade deve ser efectuada uma vez que, de um modo geral, vem melhorar a qualidade da aproximação de uma v.a. discreta (neste caso com distribuição binomial ou de Poisson) por uma v.a. contínua com distribuição normal.

Para além disso, não faria sentido aproximar $P(X = x)$ por $P(\tilde{X} = x)$, uma vez que $P(\tilde{X} = x) = 0$ pois $\tilde{X}$ é uma v.a. contínua. •

Nota 5.84 — Correcção de continuidade
Seja X uma v.a. discreta que toma valores de Δ em Δ.[25] Seja ainda $\tilde{X}$ a v.a. aproximativa contínua.[26] Então tem-se, para $\Delta = 1$ e quaisquer valores admissíveis x, a, b da v.a. X:

- $P(X = x) \simeq P(x - 1/2 < \tilde{X} \leq x + 1/2)$;
- $P(a < X \leq b) \simeq P(a + 1/2 < \tilde{X} \leq b + 1/2)$;
- $P(X < x) = P(X \leq x - 1) \simeq P[\tilde{X} \leq (x - 1) + 1/2]$. •

[25] No caso das distribuições binomial e de Poisson tem-se $\Delta = 1$, que é, aliás, o valor mais comum de Δ.
[26] V.a. esta com distribuição normal.

Exemplo 5.85 — Correcção de continuidade

Considere-se $X \sim \text{binomial}(30, 0.4)$. Uma vez que

- $n = 30 > 20$
- $np = 30 \times 0.4 = 12 > 5$ e
- $n(1-p) = 30 \times (1-0.4) = 18 > 5$,

pode aproximar-se a f.p. $P(X = x)$ por $P(x - 1/2 < \tilde{X} \leq x + 1/2)$ onde $\tilde{X} \sim \text{normal}(\mu = np = 30 \times 0.4 = 12,\ \sigma^2 = np(1-p) = 30 \times 0.4 \times (1-0.4) = 7.2)$.

x	valor exacto $P(X = x)$	valor aproximado $P(x - 1/2 < \tilde{X} \leq x + 1/2)$	erro relativo
0	0.000000	0.000008	-3297.714%
1	0.000004	0.000036	-724.420%
2	0.000043	0.000154	-260.716%
3	0.000266	0.000568	-113.712%
4	0.001197	0.001826	-52.599%
5	0.004149	0.005115	-23.285%
6	0.011524	0.012486	-8.349%
7	0.026341	0.026571	-0.872%
8	0.050487	0.049287	2.377%
9	0.082275	0.079694	3.138%
10	0.115185	0.112328	2.481%
11	0.139619	0.138014	1.149%
12	0.147375	0.147821	-0.303%
13	0.136039	0.138014	-1.452%
14	0.110127	0.112328	-1.999%
15	0.078312	0.079694	-1.764%
16	0.048945	0.049287	-0.698%
17	0.026872	0.026571	1.120%
18	0.012938	0.012486	3.493%
19	0.005448	0.005115	6.112%
20	0.002000	0.001826	8.574%
21	0.000634	0.000568	10.372%
22	0.000173	0.000154	10.853%
23	0.000040	0.000036	9.105%
24	0.000008	0.000007	3.671%
25	0.000001	0.000001	-8.127%
26	0.000000	0.000000	-32.009%
27	0.000000	0.000000	-82.321%
28	0.000000	0.000000	-203.425%
29	0.000000	0.000000	-583.648%
30	0.000000	0.000000	-2677.075%

Ao fazê-lo obtemos a tabela anterior,[27] onde pode encontrar-se também o valor do erro relativo da aproximação,

$$\left[1 - \frac{P(x - 1/2 < \tilde{X} \leq x + 1/2)}{P(X = x)}\right] \times 100\%.$$

De notar que a qualidade da aproximação é tanto melhor quanto mais próximo estiver o valor de x de $E(X) = np = 12$. •

Exercício 5.86 — Repita o Exemplo 5.85 considerando desta feita $X \sim \text{Poisson}(22.5)$ e confrontando os valores exactos e aproximados da f.d. desta v.a. discreta. •

Exemplo 5.87 — A aproximação normal da distribuição binomial
Um professor tenta lançar o sumário da sua aula teórica no FENIX no próprio dia e nos dias seguintes, fazendo uma única tentativa por dia. Contudo, a probabilidade de numa tentativa conseguir aceder ao sistema é de apenas 0.6, independentemente do dia considerado.

Durante um semestre (52 aulas) quantos sumários é de esperar que possam ser lançados no próprio dia? Indique um valor aproximado (com correcção de continuidade) para a probabilidade desse número exceder 30.[28]

- **V.a.**

  X = número de sumários lançados no próprio dia, em 52 sumários

- **Distribuição de X**

  $X \sim \text{binomial}(n, p)$

- **Parâmetros**

  $n = 52$

  $p = 0.6$

- **F.p. de X**

$$P(X = x) = \binom{52}{x} 0.6^x\, (1 - 0.6)^{52-x},\ x = 0, 1, \ldots, 52$$

[27] Os valores exactos e aproximados foram aproximados até à 6a. casa decimal e obtidos recorrendo ao *Mathematica*.

[28] Adaptado do Exame de 24 de Junho de 2006.

- **Valor esperado de X**

  $E(X) = n\,p = 52 \times 0.6 = 31.2$

- **Variância de X**

  $V(X) = n\,p\,(1-p) = 52 \times 0.6 \times (1-0.6) = 12.48$

- **V.a. aproximativa $\tilde{X}$**

  Dado que $np = 31.2 > 5$ e $n(1-p) = 20.8 > 5$ pode aproximar-se a f.d. da v.a. $X \sim \text{binomial}(n = 52, p = 0.6)$ pela f.d. da v.a.

  $$\tilde{X} \sim \text{normal}(\mu = np = 31.2, \sigma^2 = np(1-p) = 12.48).$$

  com uma correcção de continuidade.

- **Valor aproximado da probabilidade pedida**

  $$\begin{aligned}
  P(X > 30) &= 1 - P(X \leq 30) \\
  &\simeq 1 - P(\tilde{X} \leq 30 + 1/2) \\
  &= 1 - \Phi\left(\frac{30 + 1/2 - E(\tilde{X})}{\sqrt{V(\tilde{X})}}\right) \\
  &\simeq 1 - \Phi(-0.03) \\
  &\simeq \Phi(0.03) \\
  &\stackrel{tabela}{=} 0.5120.
  \end{aligned}$$

•

Exemplo 5.88 — A aproximação normal da distribuição de Poisson
O número de neutrinos registados em intervalos de 12 segundos, aquando da primeira observação da supernova S1987a por astrónomos, é bem modelado por uma distribuição de Poisson com variância igual a 0.8.

Confronte o valor exacto e o aproximado (com correcção de continuidade) da probabilidade de o número de neutrinos registados em 10 minutos exceder 40.[29]

- **V.a.**

  $X =$ número de neutrinos registados em intervalos de 12 segundos

- **Distribuição de X**

  $X \sim \text{Poisson}(\lambda)$

[29] Adaptado do Teste A de 22 de Abril de 2006.

- **Parâmetro**

  $\lambda = E(X) = V(X) = 0.8$ neutrinos (em intervalos de 12 segundos)

- **F.p. de X**

$$P(X = x) = \begin{cases} \frac{e^{-0.8} \times 0.8^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{outros valores de } x. \end{cases}$$

- **Nova v.a.**

  $Y =$ de neutrinos registados em intervalos de 10 minutos

- **Distribuição de Y**

  $Y \sim \text{Poisson}(\lambda')$[30]

- **Parâmetros**

  $\lambda' = \frac{10 \times 60}{12} \times 0.8 = 40$ (em intervalos de 10 minutos)

- **F.p. de Y**

$$P(Y = y) = \begin{cases} \frac{e^{-40} \times 40^y}{y!}, & y = 0, 1, 2, \ldots \\ 0, & \text{outros valores de } y. \end{cases}$$

- **Valor exacto da probabilidade pedida**

$$\begin{aligned} P(Y > 40) &= 1 - P(Y \leq 40) \\ &= 1 - F_{Poisson(40)}(40) \\ &\stackrel{tabela}{=} 1 - 0.5419 \\ &= 0.4581 \end{aligned}$$

- **V.a. aproximativa $\tilde{Y}$**

  Dado que $\lambda' = 40 > 5$ pode aproximar-se a f.d. da v.a. $Y \sim \text{Poisson}(\lambda' = 40)$ pela f.d. da v.a.

$$\tilde{Y} \sim \text{normal}(\mu = \lambda' = 40, \sigma^2 = \lambda' = 40).$$

  com correcção de continuidade.

[30]Este resultado deve-se à propriedade reprodutiva da distribuição de Poisson.

- **Valor aproximado da probabilidade pedida**

$$
\begin{aligned}
P(Y > 40) &= 1 - P(Y \leq 40) \\
&\simeq 1 - P(\tilde{Y} \leq 40 + 1/2) \\
&= 1 - \Phi\left(\frac{40 + 1/2 - E(\tilde{Y})}{\sqrt{V(\tilde{Y})}}\right) \\
&\simeq 1 - \Phi(0.08) \\
&\stackrel{tabela}{=} 1 - 0.5319 \\
&= 0.4681,
\end{aligned}
$$

valor este associado a um erro relativo igual a

$$\frac{|0.4681 - 0.4581|}{0.4581} \times 100\% = 2.18\%.$$

Refira-se também que há vantagem em efectuar correcção de continuidade uma vez que ao não efectuá-la o valor aproximado da probabilidade pedida é igual a 0.5 e o erro relativo associado da ordem de 9.15%. •

Capítulo 6

Estimação pontual

A Teoria das Probabilidades compreende o estudo dos modelos matemáticos capazes de descrever o comportamento de fenómenos aleatórios, modelos esses que se dizem probabilísticos.

Foi sobre o estudo de tais modelos que nos debruçámos nos capítulos 2 a 5. É altura de falarmos sobre Estatística, ramo da Matemática Aplicada que compreende técnicas quantitativas para recolher, apresentar e interpretar dados relativos a fenómenos aleatórios visando a caracterização da variabilidade desses mesmos fenómenos.

6.1 Inferência Estatística. Amostragem aleatória.

O estudo da Estatística assenta em alguns conceitos básicos que introduziremos informalmente já de seguida.

Definição informal 6.1 — V.a. ou característica de interesse
Não passa de uma característica crucial para o conhecimento do fenónemo aleatório em estudo.
Exemplos:

- a resistência de certo tipo de mola;
- o tempo até falha de pá de certo motor a jacto;
- o número de colisões de detritos em satélite em MEO no espaço de um ano.

População e unidade estatística
Conjunto de todos os objectos/indivíduos/etc. que têm em comum pelo menos uma característica de interesse. A cada elemento da população dá-se o nome de unidade estatística.
Exemplos:

- todas as molas produzidas do referido tipo;
- todas as pás de tais motores a jacto;
- todos os satélites em MEO.

Amostra e dado estatístico
Dada a impossibilidade de observar toda uma população — ou devido ao facto de ser infinita, ou por implicar a sua destruição, ou por razões de economia, comodidade, ou tempo — é fundamental recolher um subconjunto que se pretende representativo da população; este subconjunto é denominado de amostra.
A cada resultado observado — relativo à característica de interesse e respeitante a cada unidade estatística pertencente à amostra — damos o nome de dado estatístico.
Exemplos:

- recolher a 2a., 12a., 22a., 32a. e 42a. mola da produção diária;
- seleccionar completamente ao acaso 5 pás da produção semanal;
- seleccionar ao acaso um satélite chinês em MEO, um russo e três americanos.

Em qualquer dos casos anteriores as amostras possuem dimensão 5.

Amostragem
Trata-se de um vasto conjunto de procedimentos estatísticos que encontra motivação na necessidade de obtenção de amostras, i.e.,"imagens à escala da população".[1]
Exemplos:

- amostragem sistemática;[2]

[1]E justifica só por si uma ou mais disciplinas dedicadas ao seu estudo em licenciaturas como a licenciatura em Matemática Aplicada e Computação no IST.

[2]Uma amostra sistemática de tamanho n de uma população (numerada) de N unidades obtém-se fixando (ou seleccionando aleatoriamente) um número k do conjunto $\{1, 2, \ldots, N\}$, extraindo aleatoriamente uma unidade das primeiras k, que designamos por j_k, e tomando por sistema as unidades $j_k + i\,k, i = 1, 2, 3, \ldots$, até se recolher um total de n elementos ou se percorrer toda a população. No exemplo $N = 50, k = 10, j_k = 2$

- amostragem aleatória simples;[3]

- amostragem estratificada.[4]

Estatística descritiva
Com a recolha da amostra obtém-se um conjunto de dados, com um aspecto caótico, cuja mera leitura depressa se reconhece nada contribuir para a compreensão do fenónemo aleatório em estudo. A Estatística Descritiva resolve (parcialmente) esta dificuldade ao consistir numa bateria de métodos gráficos e numéricos que permitem patentear de forma sumária a informação relevante contida nos dados. •

Definição informal 6.2 — Inferência estatística
A Inferência Estatística compreende um amplo conjunto de métodos que tem por objectivo usar a informação (dados/amostra) de modo a responder a questões específicas sobre a população — muito em especial sobre aspectos relativos ao carácter aleatório da(s) v.a.(s) de interesse sob estudo. Pretende-se, por exemplo:

- adiantar valores ou intervalos de valores razoáveis para parâmetros desconhecidos como μ, σ^2, etc. — estimação pontual (Cap. 6) ou estimação intervalar (Cap. 7);

- averiguar a razoabilidade de

  – conjecturas (hipóteses) sobre parâmetros desconhecidos ou de distribuições (ou famílias de distribuições) para explicar a variabilidade da v.a. de interesse — testes de hipóteses (Cap. 8) ou

  – modelos de regressão que expliquem a relação entre um par de variáveis — regressão linear simples (Cap. 9).

A Inferência Estatística parte, assim, do particular (amostra) para o geral (população), daí designar-se também de Inferência Indutiva. •

[3] O termo aleatório significa que a selecção é aleatória, pelo que todos os elementos da população têm a mesma probabilidade de serem escolhidos e de virem a fazer parte da amostra com dimensão previamente fixada.

[4] Este tipo de amostragem passa pela divisão da população em classes mais homogéneas (estratos) de cada uma das quais se extrai uma amostra aleatória de tamanho especificado.

Nota 6.3 — A forma como interagem aqueles conceitos e a Inferência Estatística pode ser representada no seguinte esquema:

$$
\begin{array}{ccccc}
 & & \boxed{\textbf{Amostragem}} & & \\
 & \nearrow & & \searrow & \\
\boxed{\begin{array}{c}\textbf{População}\\ \text{Característica de interesse}\\ f_X(x), \mu, \text{ etc.}\end{array}} & & & & \boxed{\begin{array}{c}\textbf{Amostra}\\ \text{Estatística descritiva}\\ \text{histograma}, \bar{x}, \text{ etc.}\end{array}} \\
 & \nwarrow & & \swarrow & \\
 & & \boxed{\begin{array}{c}\textbf{Inferência Estatística}\\ \text{Estimação, testes, etc.}\end{array}} & &
\end{array}
$$

•

Motivação 6.4 — **Amostragem aleatória**
Para que as inferências sejam rigorosas[5] é natural exigir que o processo de recolha de informação seja baseado (total ou parcialmente) na intervenção do acaso. •

Definição 6.5 — **Amostra aleatória**
Sejam:

- X uma v.a. de interesse;
- $X_1, \ldots, X_n$ v.a. independentes e identicamente distribuídas (i.i.d.) a X, i.e., $X_i \sim_{i.i.d.} X, i = 1, \ldots, n$ $(n \in \mathbb{N})$.

Então o vector aleatório

- $\underline{X} = (X_1, \ldots, X_n)$ diz-se uma amostra aleatória (a.a.) de dimensão n proveniente da população X.[6] •

Definição 6.6 — **Amostra**
À observação particular da a.a. $\underline{X} = (X_1, \ldots, X_n)$ dá-se o nome de amostra e representa-se por $\underline{x} = (x_1, \ldots, x_n)$. •

[5]Ou por outra: para que às inferências esteja associado um "pequeno" grau de incerteza, incerteza esta que será quantificada probabilisticamente.

[6]Talvez fosse mais razoável designar-se por amostra aleatória (a.a.) de dimensão n respeitante à v.a. de interesse X. Há, no entanto, autores que designam X indistintamente de *população* e de *v.a. de interesse*.

Nota 6.7 — **Amostra aleatória e amostra**
Convém recordar que a a.a. $\underline{X} = (X_1, \ldots, X_n)$ é um vector aleatório $n-$dimensional e o que mesmo não acontece com a amostra $\underline{x} = (x_1, \ldots, x_n)$ que não passa de um vector de $\mathbb{R}^n$. •

Proposição 6.8 — **Caracterização da amostra aleatória**
Pelo facto de a a.a. ser constituída por n v.a. i.i.d. a X, a caracterização probabilística da a.a. $\underline{X} = (X_1, \ldots, X_n)$ faz-se sem grande dificuldade. Com efeito, tem-se para os casos:

- discreto — f.p. conjunta de $\underline{X}$

$$\begin{aligned} P(\underline{X} = \underline{x}) &= P(X_1 = x_1, \ldots, X_n = x_n) \\ &\stackrel{X_i\ indep.}{=} P(X_1 = x_1) \times \ldots \times P(X_n = x_n) \\ &= \prod_{i=1}^{n} P(X_i = x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^{n} P(X = x_i) \end{aligned} \tag{6.1}$$

- contínuo — f.d.p. conjunta de $\underline{X}$

$$\begin{aligned} f_{\underline{X}}(\underline{x}) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \\ &\stackrel{X_i\ indep.}{=} f_{X_1}(x_1) \times \ldots \times f_{X_n}(x_n) \\ &= \prod_{i=1}^{n} f_{X_i}(x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i). \end{aligned} \tag{6.2}$$

•

Motivação 6.9 — **Estatística**
É fundamental e conveniente condensar/sumariar a amostra (os dados) em medidas sumárias como a média, o desvio-padrão da amostra ou outras medidas já estudadas em Estatística Descritiva. Estas medidas mais não são que valores particulares de v.a., definidas à custa da a.a. e denominadas de estatísticas. •

Definição 6.10 — **Estatística**
Seja $\underline{X} = (X_1, \ldots, X_n)$ uma a.a. de dimensão n proveniente da população X. Neste caso

- T diz-se uma estatística se se tratar de uma função exclusiva da a.a., i.e., se $T = T(\underline{X})$. •

Nota 6.11 — **Estatística**

Uma estatística $T = T(\underline{X})$ não depende de qualquer parâmetro desconhecido. •

Exemplo 6.12 — **Estatísticas**

Estatística		Valor observado da estatística	
Mínimo da a.a.	$X_{(1)} = \min_{i=1,\dots,n} X_i$	mínimo da amostra	$x_{(1)} = \min_{i=1,\dots,n} x_i$
Máximo da a.a.	$X_{(n)} = \max_{i=1,\dots,n} X_i$	máximo da amostra	$x_{(n)} = \max_{i=1,\dots,n} x_i$
Amplitude da a.a.	$R = X_{(n)} - X_{(1)}$	amplitude da amostra	$r = x_{(n)} - x_{(1)}$
Média da a.a.	$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$	média da amostra	$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$
Var. corrigida da a.a.	$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$	var. corrigida da am.	$s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$
Var. não corrig. da a.a.	$(S')^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$	var. não corrig. da am.	$(s')^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$

Na tabela acima condensámos alguns exemplos de estatísticas, seus valores particulares e respectivas designações. •

Nota 6.13 — **Fórmula alternativa da variância corrigida (não corrigida) da amostra**

Uma vez que

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \left(\sum_{i=1}^n x_i^2\right) - n(\bar{x})^2, \tag{6.3}$$

a variância corrigida da amostra e a variância não corrigida da amostra podem escrever-se do seguinte modo:

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n x_i^2\right) - \frac{n}{n-1}(\bar{x})^2 \qquad (6.4) \\ (s')^2 &= \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n}\left(\sum_{i=1}^n x_i^2\right) - (\bar{x})^2 \qquad (6.5) \\ &= \frac{n-1}{n} s^2, \end{aligned}$$

respectivamente.

Escusado será dizer que estas fórmulas alternativas poupam operações aritméticas e poupam-nos a alguns erros de arredondamento. •

Exemplo/Exercício 6.14 — Fórmula alternativa da variância corrigida (não corrigida) da amostra

- Demonstre o resultado (6.3).

Desenvolvendo o quadrado e tirando partido do facto de $\sum_{i=1}^{n} x_i = n\bar{x}$, tem-se sucessivamente:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n(\bar{x})^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x} \times n\bar{x} + n(\bar{x})^2 \\
&= \left(\sum_{i=1}^{n} x_i^2\right) - n(\bar{x})^2.
\end{aligned}$$

•

6.2 Estimadores e suas propriedades.

O objectivo principal da Estatística é efectuar inferências sobre características da v.a. de interesse com base na amostra recolhida. Considera-se, em geral, que a distribuição de X é ou parcial, ou totalmente desconhecida.

- Parcialmente desconhecida, caso o tipo distribucional de X seja considerado conhecido (e.g., binomial, Poisson, exponencial, etc.) a menos de um ou mais parâmetros desconhecidos (e.g. μ, σ^2, etc.).[7] Nesta situação as inferências que possamos fazer dizem-se do tipo paramétrico.

- Totalmente desconhecida, se o tipo distribucional de X for especificado de modo vago (e.g., distribuição discreta, etc.). Neste caso as inferências dizem-se não paramétricas.

Nota 6.15 — Parâmetro desconhecido
Um parâmetro desconhecido unidimensional (resp.multidimensional) será de um modo geral representado por θ (resp. $\underline{\theta}$). •

Definição informal 6.16 — Espaço paramétrico
Corresponde ao conjunto de todos os valores possíveis para o parâmetro desconhecido θ e é frequentemente representado por Θ.

Modelo paramétrico
Família de distribuições possíveis para a v.a. de interesse X.[8] Esta família é usualmente representada por $\{\qquad : \theta \in \Theta\}$ onde no espaço em branco se colocará indistintamente a expressão geral da f.(d.)p. de X ou o nome da distribuição de X, dependentes em todo o caso do parâmetro θ. •

Motivação 6.17 — Estimadores
É fundamental adiantar valores razoáveis para parâmetros desconhecidos que caracterizem a distribuição da nossa v.a. de interesse. Para tal iremos recorrer a estatísticas com características especiais que denominaremos de estimadores. •

Definição 6.18 — Estimador
A estatística $T = T(\underline{X})$ diz-se um estimador do parâmetro desconhecido θ, caso $T = T(\underline{X})$ tome valores exclusivamente no espaço paramétrico Θ. •

[7] Considera-se em todo o caso que o número de parâmetros desconhecidos é finito.
[8] Paulino (1994) define modelo paramétrico à custa de $\underline{X}$.

Definição 6.19 — Estimativa
Ao valor observado do estimador $T = T(\underline{X})$ do parâmetro desconhecido θ, $t = T(\underline{x})$, damos o nome de estimativa de θ. Trata-se naturalmente de um valor razoável para θ já que $t = T(\underline{x}) \in \Theta$. •

Exemplo 6.20 — Modelo e espaço paramétricos; estimador e estimativa
Admita que vai inquirir n condutores/as quanto à sua preferência (ou não) por motores a gasóleo e que as respostas possíveis (admissíveis) neste inquérito são:

- Sim (1), prefiro motor a gasóleo;
- Não (0), prefiro motor a gasolina.

Procure identificar: a v.a. de interesse; a respectiva distribuição; o parâmetro desconhecido; o modelo e o espaço paramétricos; uma estimativa e um estimador do parâmetro desconhecido.

- **V.a. de interesse**

$$\begin{aligned} X &= \text{resposta de condutor/a inquirido/a} \\ &= \begin{cases} 1 \text{ (resposta afirmativa)}, & \text{com probabilidade } \theta \\ 0 \text{ (resposta negativa)}, & \text{com probabilidade } (1-\theta) \end{cases} \end{aligned}$$

- **Distribuição de X**

  $X \sim \text{Bernoulli}(\theta)$

- **Parâmetro desconhecido**

  $\theta = P(X = 1) = P(\text{resposta afirmativa})$

- **Espaço paramétrico**

  $\Theta = [0, 1]$

- **Modelo paramétrico**

  $\{\text{Bernoulli}(\theta), \theta \in \Theta\}$ ou alternativamente $\{\theta^x(1-\theta)^{1-x}, \theta \in \Theta\}$

- **A.a.**

  $\underline{X} = (X_1, \ldots, X_n)$ a.a. de dimensão n proveniente da população X

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$ onde $\underline{x} \in \{0, 1\}^n$

- **Estimativa de θ**

  Candidata: um valor razoável para θ é

$$\begin{aligned} T(\underline{x}) &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \bar{x} \\ &= \text{proporção observada de “SIM'S”} \end{aligned}$$

- **Estimador de θ**

  Candidato:

$$T(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  Verificações:

  1. $T(\underline{X})$ só depende de $\underline{X}$
  2. $T(\underline{X})$ toma valores em $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\} \subset \Theta = [0,1]$

  Conclusão:

  $T(\underline{X})$ é estimador de θ. •

Motivação 6.21 — Propriedades dos estimadores
Um estimador conduzirá a inferências/estimativas mais rigorosas se gozar de algumas das propriedades descritas já de seguida. •

Definição 6.22 — Estimador centrado
O estimador T diz-se um estimador centrado de θ [9] se

$$E[T(\underline{X})] = \theta, \forall \theta \in \Theta, \tag{6.6}$$

i.e., o centro de gravidade do estimador é igual a θ independentemente do valor que este parâmetro desconhecido possa assumir. •

Definição 6.23 — Estimador enviesado
O estimador T diz-se um estimador enviesado de θ [10] se

$$\exists \theta \in \Theta : E[T(\underline{X})] \neq \theta. \tag{6.7}$$

•

[9] Ou um “estimador não enviesado” de θ.
[10] Ou um “estimador não centrado” de θ.

Definição 6.24 — **Enviesamento de um estimador**
O estimador de θ, T, possui enviesamento[11] dado por

$$\text{bias}_\theta[T(\underline{X})] = E[T(\underline{X})] - \theta. \tag{6.8}$$

Como seria de esperar um estimador centrado (enviesado, resp.) de θ possui enviesamento nulo (não nulo, resp.). •

Nota 6.25 — **Enviesamento**
Escusado será dizer que há a possibilidade de adiantar mais que um estimador para um parâmetro desconhecido. Um estimador de θ será tanto "melhor" quanto menor for o seu enviesamento. •

Exemplo/Exercício 6.26 — **Estimadores centrados de μ e σ^2**
Considere que X é uma v.a. de interesse com distribuição arbitrária, valor esperado μ e variância σ^2.

(a) Prove que

- a média da a.a.
  $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$
- e a variância corrigida da a.a.
  $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^n X_i^2 - n\,(\bar{X})^2\right]$

são estimadores centrados de μ e σ^2, respectivamente.

- **V.a.**
  $X_i \overset{i.i.d.}{\sim} X$,
  $E(X_i) = E(X) = \mu$,
  $V(X_i) = V(X) = \sigma^2,\ i = 1, \ldots, n$
- **Estimador de μ**
  $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$

[11] O termo anglo-saxónico para enviesamento ou viés é "bias".

- **Estimador centrado de μ ?**

  Trata-se de facto de um estimador centrado de μ já que

$$\begin{aligned}
E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&\overset{X_i \sim X}{=} \frac{1}{n}\sum_{i=1}^{n} E(X) \\
&= \frac{1}{n} \times n\mu \\
&= \mu
\end{aligned}$$

- **Estimador de σ^2**

  $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} X_i^2\right) - n\,(\bar{X})^2\right]$

- **Estimador centrado de σ^2 ?**

  De facto, ao tirar-se partido da fórmula alternativa de S^2 e ao notar que $E(Z^2) = V(Z) + E^2(Z)$, $E(\bar{X}) = \mu$ e $V(\bar{X}) = \sigma^2/n$, segue-se

$$\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}\,E\left[\left(\sum_{i=1}^{n} X_i^2\right) - n(\bar{X})^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE[(\bar{X})^2]\right] \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}[V(X_i) + E^2(X_i)] - n \times [V(\bar{X}) + E^2(\bar{X})]\right\} \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n \times (\sigma^2/n + \mu^2)\right] \\
&= \frac{1}{n-1}\,(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) \\
&= \frac{1}{n-1}\,(n-1)\,\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

  Assim, conclui-se que S^2 é efectivamente um estimador centrado de σ^2.

(b) Demonstre também que a

- a variância não corrigida da a.a.
  $(S')^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2\right) - n(\bar{X})^2 = \frac{n-1}{n} S^2$

é um estimador enviesado de σ^2 e calcule o respectivo enviesamento.

- **Outro estimador de** σ^2
  $(S')^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2$
- **Estimador centrado de** σ^2 ?
  Tendo em conta que $(S')^2 = \frac{n-1}{n} S^2$ rapidamente se conclui que

$$
\begin{aligned}
E\left[(S')^2\right] &= E\left(\frac{n-1}{n} S^2\right) \\
&= \frac{n-1}{n}\, E\left(S^2\right) \\
&= \frac{n-1}{n}\, \sigma^2 \\
&\neq \sigma^2,
\end{aligned}
$$

  pelo que $(S')^2$ não é um estimador centrado de σ^2.

- **Enviesamento de** $(S')^2$

$$
\begin{aligned}
\text{bias}_{\sigma^2}[(S')^2] &= E[(S')^2] - \sigma^2 \\
&= \frac{n-1}{n}\, \sigma^2 - \sigma^2 \\
&= -\frac{1}{n}\, \sigma^2 \\
&< 0,
\end{aligned}
$$

  donde se possa concluir que $(S')^2$ subestima (em valor esperado) σ^2. •

Nota 6.27 — Variância (resp. não) corrigida
É pelo facto de S^2 (resp. $(S')^2$) ser um estimador centrado (resp. enviesado) de σ^2 que se denomina este estimador de "variância CORRIGIDA (resp. NÃO CORRIGIDA) da a.a.". •

Motivação 6.28 — Erro quadrático médio
Não basta que um estimador de θ seja centrado para garantir estimativas rigorosas. Estas serão tanto mais rigorosas quanto menos o estimador se dispersar em torno do verdadeiro valor do parâmetro desconhecido θ. •

Definição 6.29 — **Erro quadrático médio**
O erro quadrático médio (EQM)[12] do estimador de θ, $T = T(\underline{X})$, é dado por

$$\begin{aligned} EQM_\theta[T(\underline{X})] &= E\left\{[T(\underline{X}) - \theta]^2\right\} \\ &= V[T(\underline{X})] + \{E[T(\underline{X})] - \theta\}^2 \\ &= V[T(\underline{X})] + \{bias_\theta[T(\underline{X})]\}^2 . \end{aligned} \tag{6.9}$$

•

Exercício 6.30 — Demonstre o resultado (6.9). •

Nota 6.31 — **Erro quadrático médio**
Uma vez definido o erro quadrático médio, escusado será dizer que:

1. EQM quantifica a dispersão esperada do estimador em torno do verdadeiro valor do parâmetro desconhecido θ.

2. Um estimador será tanto "melhor" quanto menor for o seu EQM. Assim, ao lidarmos com dois estimadores de θ devemos optar por aquele que possuir o menor EQM, já que conduzirá a estimativas mais rigorosas de θ. Deste modo estaremos a optar pelo estimador mais "eficiente" de θ. •

Definição 6.32 — **Eficiência relativa de estimadores**
Sejam $T_1 = T_1(\underline{X})$ e $T_2 = T_2(\underline{X})$ dois estimadores do parâmetro desconhecido θ. Então a eficiência de T_1 — com respeito a T_2 na estimação de θ — é dada por

$$e_\theta[T_1(\underline{X}), T_2(\underline{X})] = \frac{EQM_\theta[T_2(\underline{X})]}{EQM_\theta[T_1(\underline{X})]}. \tag{6.10}$$

Assim sendo, se

$$e_\theta[T_1(\underline{X}), T_2(\underline{X})] > 1 \Leftrightarrow EQM_\theta[T_2(\underline{X})] > EQM_\theta[T_1(\underline{X})], \tag{6.11}$$

diremos que o estimador $T_1(\underline{X})$ é mais eficiente que $T_2(\underline{X})$ na estimação de θ. •

[12] A designação anglo-saxónica é "mean square error" (MSE).

Exemplo 6.33 — Eficiência relativa de estimadores
Num estudo prévio ao lançamento no mercado de uma nova pilha de "pacemaker" foram postas algumas questões acerca da sua duração (em milhares de dias) a um engenheiro. Estudos anteriores (embora com outros tipos de pilhas) levam a crer que tal v.a. possui distribuição uniforme$(0,\theta)$, onde o parâmetro θ é positivo, desconhecido e representa a idade máxima da pilha.

Calcule a eficiência relativa de $X_{(n)}$ com respeito a $2\bar{X}$ no que se refere à estimação do parâmetro θ. Para o efeito, atente que $E[X_{(n)}] = \frac{n}{n+1}\,\theta$ e $V[X_{(n)}] = \frac{n}{(n+2)(n+1)^2}\,\theta^2$. Diga qual dos dois estimadores é mais eficiente.

- **V.a.**

  $X_i =$ duração da pilha $i,\ i = 1,\ldots,n$

  $X_i \stackrel{i.i.d.}{\sim} X,\ i = 1,\ldots,n$

- **Distribuição**

  $X \sim$ uniforme$(0,\theta)$

- **Parâmetro**

  θ DESCONHECIDO $(\theta > 0)$

- **Estimador de θ**

  $X_{(n)} = \max_{i=1,\ldots,n} X_i$

- **Erro quadrático médio de $X_{(n)}$**

$$\begin{aligned}
EQM_\theta[X_{(n)}] &= V[X_{(n)}] + \left\{bias_\theta[X_{(n)}]\right\}^2 \\
&= V[X_{(n)}] + \left\{E[X_{(n)}] - \theta\right\}^2 \\
&= \frac{n}{(n+2)(n+1)^2}\,\theta^2 + \left(\frac{n}{n+1}\,\theta - \theta\right)^2 \\
&= \frac{2}{(n+2)(n+1)}\,\theta^2
\end{aligned}$$

- **Outro estimador de θ**

  $2\bar{X}$

- **Erro quadrático médio de $2\bar{X}$**

  Se se tiver em consideração que $E(\bar{X}) = E(X)$, $V(\bar{X}) = \frac{V(X)}{n}$, $E(X) \stackrel{form.}{=} \frac{\theta}{2}$ e $V(X) \stackrel{form.}{=} \frac{\theta^2}{12}$ tem-se

$$
\begin{aligned}
EQM_\theta(2\bar{X}) &= V(2\bar{X}) + \left[bias_\theta(2\bar{X})\right]^2 \\
&= V(2\bar{X}) + \left[E(2\bar{X}) - \theta\right]^2 \\
&= \frac{2^2}{n} V(X) + [2E(X) - \theta]^2 \\
&= \frac{2^2}{n} \frac{\theta^2}{12} + \left(2 \times \frac{\theta}{2} - \theta\right)^2 \\
&= \frac{1}{3n} \theta^2
\end{aligned}
$$

- **Eficiência relativa de $X_{(n)}$ com respeito a $2\bar{X}$**

$$
\begin{aligned}
e_\theta[X_{(n)}, 2\bar{X}] &= \frac{EQM_\theta(2\bar{X})}{EQM_\theta(X_{(n)})} \\
&= \frac{\frac{1}{3n}\theta^2}{\frac{2}{(n+2)(n+1)}\theta^2} \\
&= \frac{(n+2)(n+1)}{6n}
\end{aligned}
$$

  que constitui o termo geral de uma sucessão monótona não decrescente cujos dois primeiros termos são iguais a 1.

- **Comentário**

  Tendo em conta a expressão de $e_\theta[X_{(n)}, 2\bar{X}]$ pode afirmar-se que

  - $X_{(n)}$ e $2\bar{X}$ são igualmente eficientes, para $n = 1, 2$,

  no entanto,

  - $X_{(n)}$ é mais eficiente que $2\bar{X}$, para $n > 2$.

  Curiosamente, $X_{(n)}$ não é estimador centrado de θ ao contrário de $2\bar{X}$. •

6.3 Método da máxima verosimilhança.

Até ao momento introduzimos estimadores cujas concretizações constituem valores razoáveis para parâmetros desconhecidos. Apresentámos também propriedades desejáveis para esses mesmos estimadores por forma a que conduzam a estimativas rigorosas desses parâmetros. Resta adiantar um método de obtenção sistemática de estimadores de parâmetros desconhecidos e já agora averiguar se tais estimadores possuem boas propriedades.

Motivação 6.34 — Método da máxima verosimilhança
O método da máxima verosimilhança (MV) permite obter o valor mais plausível/ verosímil de um parâmetro desconhecido — de entre todos os valores possíveis para esse mesmo parâmetro —, tendo em conta a amostra $\underline{x} = (x_1, \ldots, x_n)$ de que dispomos. •

Por forma a descrever o método da MV é necessário definir a função de verosimilhança.

Definição 6.35 — Função de verosimilhança
A função de verosimilhança[13] é representada por $L(\theta|\underline{x})$, dá ideia de quão plausível é o valor θ para o parâmetro desconhecido, caso se tenha recolhido a amostra $\underline{x}$, e define-se do seguinte modo:

- Caso discreto

$$\begin{aligned} L(\theta|\underline{x}) &= P(\underline{X} = \underline{x}|\theta) \\ &= \prod_{i=1}^{n} P(X = x_i|\theta), \theta \in \Theta, \end{aligned} \tag{6.12}$$

- Caso contínuo

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}|\theta) \\ &= \prod_{i=1}^{n} f_X(x_i|\theta), \theta \in \Theta, \end{aligned} \tag{6.13}$$

onde $P(\bullet|\theta)$ e $f_X(\bullet|\theta)$ representam a f.p. e a f.d.p. (resp.) da v.a. de interesse X tendo em conta que θ é o verdadeiro valor do parâmetro desconhecido. •

[13] Na literatura anglo-saxónica “likelihood function”.

Nota 6.36 — Função de verosimilhança

1. Por tradição quer o parâmetro desconhecido, quer o valor que se lhe possa atribuir são representados por θ.

2. $L(\theta|\underline{x}) : \Theta \to I\!R$, i.e., a função de verosimilhança tem como argumento (exclusivo) θ, possui como domínio o espaço paramétrico Θ e toma valores em $I\!R$, para cada valor fixo da amostra $\underline{x}$.[14] •

Definição 6.37 — Estimativa de máxima verosimilhança
Obtida a amostra $\underline{x} = (x_1, \ldots, x_n)$, a estimativa de máxima verosimilhança do parâmetro desconhecido corresponde ao ponto de máximo da função de verosimilhança ou, equivalentemente, ao ponto de máximo do logaritmo da função de verosimilhança.[15] Esta estimativa é representada por $\hat{\theta}$ e verifica

$$L(\hat{\theta}|\underline{x}) = \max_{\theta \in \Theta} L(\theta|\underline{x}) \tag{6.14}$$

ou, equivalentemente,

$$\ln L(\hat{\theta}|\underline{x}) = \max_{\theta \in \Theta} \ln L(\theta|\underline{x}), \tag{6.15}$$

onde a função $\ln L(\hat{\theta}|\underline{x})$ é usualmente designada de log-verosimilhança. •

Nota 6.38 — Estimativa de máxima verosimilhança

1. É analiticamente mais conveniente obter o ponto de máximo da função log-verosimilhança (uma soma de logaritmos) que o ponto de máximo da função de verosimilhança (um produto).

2. Quando o espaço paramétrico é um conjunto discreto ($\Theta = \{\theta_1, \ldots, \theta_m\}$) o ponto de máximo da função de (log-)verosimilhança obtém-se por pesquisa ponto por ponto.

3. No caso em que o espaço paramétrico Θ é contínuo recorre-se ao procedimento usual de maximização — começa-se por obter o ponto de estacionaridade para de seguida averiguar se tal ponto é efectivamente um ponto de máximo.

[14] Na verdade $L(\theta|\underline{x})$ toma valores no intervalo $[0, 1]$, no caso discreto, e em $I\!R^+$, no caso contínuo.
[15] Na verdade seria mais correcto definir a estimativa de MV como um ponto de supremo.

Ao lidar-se com um único parâmetro desconhecido tem-se:[16]

$$\hat{\theta}: \quad \left.\frac{d\ln L(\theta|\underline{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0 \text{ (ponto de estacionaridade)} \tag{6.16}$$

$$\left.\frac{d^2\ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} < 0 \text{ (ponto de máximo).} \tag{6.17}$$

Ao lidar-se com um vector de p ($p > 1$) parâmetros desconhecidos, $\underline{\theta} = (\theta_1, \ldots, \theta_p)$, a estimativa de MV, $\hat{\underline{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)$ não só verifica

$$\left.\frac{\partial \ln L[(\theta_1, \ldots, \theta_p)|\underline{x}]}{\partial\theta_j}\right|_{\underline{\theta}=\hat{\underline{\theta}}} = 0, \ j = 1, \ldots, p \tag{6.18}$$

como a matriz hessiana —

$$\mathbf{H}(\theta) = \nabla^2 \ln L[(\theta_1, \ldots, \theta_p)|\underline{x}] = [h_{ij}(\underline{\theta})]_{i,j=1,\ldots,p} \tag{6.19}$$

onde $h_{ij}(\underline{\theta}) = \frac{\partial^2 \ln L[(\theta_1,\ldots,\theta_p)|\underline{x}]}{\partial\theta_i\partial\theta_j}$ — seja definida negativa quando avaliada em $\hat{\underline{\theta}}$, quando $p > 1$.

Caso $p = 2$, $\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$ satisfaz não só (6.18) mas também as duas seguintes condições (Casella e Berger, 2002, p. 322):

(a) $\begin{cases} \left.\frac{\partial^2 \ln L[(\theta_1,\theta_2)|\underline{x}]}{\partial\theta_1^2}\right|_{\underline{\theta}=\hat{\underline{\theta}}} < 0 \quad \text{ou} \\ \left.\frac{\partial^2 \ln L[(\theta_1,\theta_2)|\underline{x}]}{\partial\theta_2^2}\right|_{\underline{\theta}=\hat{\underline{\theta}}} < 0; \end{cases}$

(b) o determinante da matriz hessiana $[h_{ij}(\underline{\theta})]_{i,j=1,2}$ é positivo quando avaliado em $\underline{\theta} = \hat{\underline{\theta}}$, i.e.

$$h_{11}(\hat{\underline{\theta}}) \times h_{22}(\hat{\underline{\theta}}) - h_{12}(\hat{\underline{\theta}}) \times h_{21}(\hat{\underline{\theta}}) > 0.$$

4. As equações (6.18) denominam-se equações de verosimilhança.

5. A estimativa de MV é, naturalmente, uma função da amostra, i.e.,

$$\hat{\theta} = g(\underline{x}). \tag{6.20}$$

Para além disso não se trata de uma v.a. mas da concretização de uma v.a. com um nome particular: estimador. •

[16] A satisfação da equação (6.16) é condição necessária mas não suficiente para que se obtenha um ponto de máximo (eventualmente local). Para que tal aconteça é fundamental que se verifique também a continuidade das segundas derivadas numa vizinhança do ponto de máximo e que a segunda derivada seja negativa.

Definição 6.39 — Estimador de máxima verosimilhança
O estimador de MV de θ obtém-se por substituição de $\underline{x} = (x_1, \dots, x_n)$ por $\underline{X} = (X_1, \dots, X_n)$ na expressão geral da estimativa de MV, $\hat{\theta} = g(\underline{x})$, obtendo-se

$$\text{EMV}(\theta) = g(\underline{X}). \tag{6.21}$$

Trata-se de uma v.a. exclusivamente dependente da a.a. $\underline{X}$, logo uma estatística. •

Exemplo 6.40 — Estimador e estimativa de MV (caso discreto)
Um inquérito recente feito a 1000 habitantes de uma região rural revelou que 448 pessoas apoiam a aplicação de penas de prisão pesadas em crimes de fogo posto.

Deduza a estimativa de máxima verosimilhança da probabilidade (p) de uma pessoa escolhida ao acaso em tal região ser favorável à aplicação da referida pena. Verifique que o estimador associado é centrado.

- **V.a. de interesse**

  $X =$ resposta ao inquérito

- **Distribuição**

  $X \sim \text{Bernoulli}(p)$

- **Parâmetro** DESCONHECIDO

  $p = P(X = 1), \ 0 \leq p \leq 1$

- **F.p.**

  $$P(X = x) \stackrel{form}{=} p^x \, (1-p)^{1-x}, \ x = 0, 1$$

- **Amostra**

  $\underline{x} = (x_1, \dots, x_n)$ amostra de dimensão n proveniente da população X onde $x_i =$ resposta da $i -$ ésima pessoa

  $$\begin{aligned} \underline{x} \ : \ & n = 1000 \\ & \bar{x} = \frac{448}{1000} = 0.448 = 44.8\% \text{ respostas afirmativas} \end{aligned}$$

- **Obtenção da estimativa de MV de p**

Passo 1 — Função de verosimilhança

$$\begin{aligned}
L(p|\underline{x}) &= \prod_{i=1}^{n} P(X = x_i) \\
&= \prod_{i=1}^{n} \left[p^{x_i}\,(1-p)^{1-x_i}\right] \\
&= p^{\sum_{i=1}^{n} x_i}\,(1-p)^{n-\sum_{i=1}^{n} x_i}
\end{aligned}$$

Passo 2 — Função de log-verosimilhança

$$\begin{aligned}
\ln L(p|\underline{x}) &= \ln\left[p^{\sum_{i=1}^{n} x_i}\,(1-p)^{n-\sum_{i=1}^{n} x_i}\right] \\
&= \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\left(n - \sum_{i=1}^{n} x_i\right)
\end{aligned}$$

Passo 3 — Maximização

A estimativa de MV de p, $\hat{p}$, obtém-se resolvendo

$$\hat{p} \;:\; \begin{cases} \left.\frac{d \ln L(p|\underline{x})}{dp}\right|_{p=\hat{p}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(p|\underline{x})}{dp^2}\right|_{p=\hat{p}} < 0 & \text{(ponto de máximo)} \end{cases}$$

Tendo em conta a função log-verosimilhança e relembrando que $0 \leq p \leq 1$,[17] tem-se sucessivamente

$$\hat{p} \;:\; \begin{cases} \left.\frac{d\left[\ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\left(n-\sum_{i=1}^{n} x_i\right)\right]}{dp}\right|_{p=\hat{p}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\left[\ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\left(n-\sum_{i=1}^{n} x_i\right)\right]}{dp^2}\right|_{p=\hat{p}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \left.\left(\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p}\right)\right|_{p=\hat{p}} = 0 \\ \left.-\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{n-\sum_{i=1}^{n} x_i}{(1-p)^2}\right|_{p=\hat{p}} < 0 \end{cases}$$

$$\begin{cases} \frac{\sum_{i=1}^{n} x_i}{\hat{p}} - \frac{n-\sum_{i=1}^{n} x_i}{1-\hat{p}} = 0 \\ -\frac{\sum_{i=1}^{n} x_i}{\hat{p}^2} - \frac{n-\sum_{i=1}^{n} x_i}{(1-\hat{p})^2} < 0 \end{cases}$$

[17] Aqui e ali será necessário admitir que $p \neq 0, 1$.

$$\begin{cases} (1-\hat{p}) \sum_{i=1}^{n} x_i - \hat{p}\ (n - \sum_{i=1}^{n} x_i) = 0 \\ -\frac{\sum_{i=1}^{n} x_i}{\hat{p}^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-\hat{p})^2} < 0 \end{cases}$$

$$\begin{cases} \hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i \\ \text{Proposição verdadeira já que } 0 \leq p \leq 1 \end{cases}$$

Passo 4 — Concretização

Para este inquérito tem-se:

$$\begin{aligned} \hat{p} &= \frac{1}{n} \sum_{i=1}^{n} x_i \qquad\qquad (\hat{p} = \bar{x} = \text{média da amostra}) \\ &= \frac{\text{no. obs. de respostas afirmativas}}{\text{no. pessoas inquiridas}} \\ &= 0.448 \end{aligned}$$

- **Estimador de MV de p**

  Será representado pela v.a. $EMV(p) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$ (i.e., pela média da a.a.) e possui valor esperado igual a $E(\bar{X}) = E(X) = p$. Deste modo conclui-se que o estimador de MV de p é centrado. •

Exemplo 6.41 — Estimador e estimativa de MV (caso contínuo)
Os tempos observados (em anos) até à primeira colisão de detritos espaciais com diâmetro inferior a 1mm em 4 satélites em MEO foram de 1.2, 1.5, 1.8, 1.4.

Admita que tal tempo possui distribuição pertencente ao modelo exponencial de parâmetro λ. Obtenha o estimador e a estimativa de MV de λ.

- **V.a. de interesse**

  $X =$ tempo até primeira colisão de detritos espaciais (em anos)

- **Distribuição**

  $X \sim$ exponencial(λ)

- **Parâmetro** DESCONHECIDO

  λ $(\lambda > 0)$

- **F.d.p.**

$$f_X(x) \overset{form}{=} \begin{cases} \lambda\, e^{-\lambda\, x}, \ x \geq 0 \\ 0, \ \text{c.c.}, \end{cases}$$

- **Amostra**

 $\underline{x} = (x_1, \ldots, x_n)$ amostra de dimensão n proveniente da população X

$$\begin{aligned} \underline{x} \quad : \quad & n = 4 \\ & \bar{x} = \frac{1}{4}(1.2 + 1.5 + 1.8 + 1.4) = 1.475 \end{aligned}$$

- **Obtenção da estimativa de MV de λ**

 Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\lambda|\underline{x}) &= \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left(\lambda\, e^{-\lambda\, x_i}\right) \\ &= \lambda^n\, e^{-\lambda \sum_{i=1}^{n} x_i} \end{aligned}$$

 Passo 2 — Função de log-verosimilhança

$$\begin{aligned} \ln L(\lambda|\underline{x}) &= \ln\left(\lambda^n\, e^{-\lambda \sum_{i=1}^{n} x_i}\right) \\ &= n\, \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \end{aligned}$$

 Passo 3 — Maximização

 A estimativa de MV de λ é aqui representada por $\hat{\lambda}$ e

$$\hat{\lambda} \quad : \quad \begin{cases} \left.\frac{d \ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases}$$

 Substituindo a função log-verosimilhança nas expressões acima e tendo em conta que $\lambda > 0$, obtém-se

$$\hat{\lambda} \quad : \quad \begin{cases} \left.\left(\frac{n}{\lambda} - \sum_{i=1}^{n} x_i\right)\right|_{\lambda=\hat{\lambda}} = 0 \\ \left.-\frac{n}{\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 \end{cases}$$

$$\begin{cases} \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} x_i = 0 \\ -\frac{n}{\hat{\lambda}^2} < 0 \end{cases}$$

$$\begin{cases} \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} \\ \text{Proposição verdadeira já que } \lambda > 0 \end{cases}$$

Passo 4 — Concretização

Para esta amostra tem-se:

$$\begin{aligned} \hat{\lambda} &= (\bar{x})^{-1} \\ &= \text{inverso da média da amostra} \\ &= 1.475^{-1} \\ &\simeq 0.678 \end{aligned}$$

- **Estimador de MV de λ**

  Será representado pela v.a. $EMV(\lambda) = (\bar{X})^{-1}$ (i.e, inverso da média da a.a.). Por sinal não se trata de estimador centrado de λ. •

O estimador de MV nem sempre é único e nem sempre é centrado. No entanto, os estimadores de MV gozam de várias propriedades importantes, das quais destacamos três que enunciaremos informalmente já de seguida.

Nota 6.42 — Propriedades dos estimadores de MV

1. **Invariância**

   Sejam:

   - $\hat{\theta}$ a estimativa de MV de θ
   - EMV(θ) o estimador de MV de θ
   - $h(\theta)$ uma função bijectiva de θ.[18]

   Então a estimativa de MV de $h(\theta)$ é dada por

[18]Exigir que $h(\theta)$ seja uma função bijectiva de θ pode parecer demasiado restritivo. Com efeito, trata-se de uma condição suficiente mas não necessária para que seja satisfeita a invariância. De acordo com Rohatgi e Saleh (2001, p. 418) basta que a função $h(\theta) : \mathbb{R}^p \to \mathbb{R}^m$ $(m \leq p)$ transforme conjuntos abertos de $\mathbb{R}^p$ em abertos de $\mathbb{R}^m$.

$$\widehat{h(\theta)} = h(\hat{\theta}) \tag{6.22}$$

e o estimador de MV de $h(\theta)$ dado por

$$\text{EMV}(h(\theta)) = h[\text{EMV}(\theta)]. \tag{6.23}$$

2. **Suficiência**

   A suficiência pode ser descrita informalmente do seguinte modo: as estimativas de MV condensam em geral toda a informação relevante, contida na amostra, sobre o parâmetro desconhecido.

3. **Consistência**

   Esta propriedade, que de um modo geral os estimadores de MV possuem, pode ser informalmente traduzida no seguinte comportamento probabilístico: à medida que aumentamos a dimensão da a.a. (n), o EMV(θ) dispersa-se cada vez menos em torno do verdadeiro valor de θ (i.e., as inferências tornam-se cada vez mais rigorosas). •

Exemplo 6.43 — Propriedade da invariância dos estimadores de MV
Com o objectivo de estudar o tempo até falha de certo equipamento electrónico (em dezenas de milhar de horas), uma engenheira recolheu um total de 50 observações que conduziram à média geométrica amostral $m_g = \left(\prod_{i=1}^{50} x_i\right)^{1/50} = 4.2427$.

Confirmada a adequação do modelo $\{\text{Pareto}(2.5, \lambda), \lambda > 0\}$, cuja f.d.p. é dada por

$$f_X(x) = \begin{cases} \lambda\, 2.5^\lambda\, x^{-(\lambda+1)}, & x \geq 2.5 \\ 0, & \text{c.c.,} \end{cases}$$

aquela mesma engenheira passou para a fase de estimação pontual do parâmetro desconhecido e de uma sua função.[19]

(a) Prove que a estimativa de máxima verosimilhança de λ é igual a $\hat{\lambda} = [\ln(m_g) - ln(2.5)]^{-1}$.

- **V.a. de interesse**

  X = tempo até falha de certo equipamento electrónico (em 10^4 horas)

- **Distribuição**

  $X \sim \text{Pareto}(2.5, \lambda)$

[19] Adaptado do Exame de 4 de Fevereiro de 2003.

- **Parâmetro** DESCONHECIDO

  λ $(\lambda > 0)$

- **F.d.p.**

$$f_X(x) = \begin{cases} \lambda\, 2.5^{\lambda}\, x^{-(\lambda+1)}, x \geq 2.5 \\ 0, \text{c.c.}, \end{cases}$$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$ é uma amostra de dimensão n proveniente da população X tal que

$$\begin{aligned} n &= 50 \\ m_g &= \left(\prod_{i=1}^{50} x_i\right)^{1/50} = 4.2427 \end{aligned}$$

- **Obtenção da estimativa de MV de λ**

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\lambda|\underline{x}) &= \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left[\lambda\, 2.5^{\lambda}\, x_i^{-(\lambda+1)}\right] \\ &= \lambda^n \times 2.5^{n\,\lambda} \left(\prod_{i=1}^{n} x_i\right)^{-(\lambda+1)} \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\begin{aligned} \ln L(\lambda|\underline{x}) &= \ln\left[\lambda^n \times 2.5^{n\,\lambda} \left(\prod_{i=1}^{n} x_i\right)^{-(\lambda+1)}\right] \\ &= n\, \ln(\lambda) + n\,\lambda \ln(2.5) - (\lambda+1) \sum_{i=1}^{n} \ln(x_i) \\ &= n\, \ln(\lambda) + n\,\lambda \ln(2.5) - n\,(\lambda+1)\, \ln(m_g) \end{aligned}$$

  Passo 3 — Maximização

  Representar-se-á a estimativa de MV de λ por $\hat{\lambda}$ e é sabido que

$$\hat{\lambda} \;:\; \begin{cases} \left.\frac{d \ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo).} \end{cases}$$

Tirando partido da expressão da função log-verosimilhança e do facto de $\lambda > 0$, segue-se

$$\hat{\lambda} \ : \ \begin{cases} \frac{n}{\hat{\lambda}} + n\,\ln(2.5) - n\,\ln(m_g) = 0 \\ -\frac{n}{\hat{\lambda}^2} < 0 \end{cases}$$

$$\begin{cases} \hat{\lambda} = \frac{1}{\ln(m_g)-\ln(2.5)} \\ \text{Proposição verdadeira já que } \lambda > 0 \end{cases}$$

Passo 4 — Concretização

Particularizando para a amostra recolhida obtém-se:

$$\begin{aligned} \hat{\lambda} &= [\ln(m_g) - \ln(2.5)]^{-1} \\ &= [\ln(4.2427) - \ln(2.5)]^{-1} \\ &\simeq 1.8907. \end{aligned}$$

(b) Obtenha a estimativa de MV da probabilidade de a duração do equipamento exceder um período de 35.000 horas.

- **Outro parâmetro** DESCONHECIDO

$$\begin{aligned} h(\lambda) &= P(X > 3.5) \\ &= \int_{3.5}^{+\infty} \lambda\, 2.5^{\lambda}\, x^{-(\lambda+1)} dx \\ &= \lambda\, 2.5^{\lambda} \left. \frac{x^{-(\lambda+1)+1}}{-(\lambda+1)+1} \right|_{3.5}^{+\infty} \\ &= \left(\frac{2.5}{3.5}\right)^{\lambda} \end{aligned}$$

- **Estimativa de MV de** $h(\lambda)$

Uma vez que $h(\lambda)$ é uma função bijectiva de λ pode invocar-se a propriedade da invariância dos estimadores de MV e concluir que a estimativa de MV de $h(\lambda)$ é

$$\widehat{h(\lambda)} = h(\hat{\lambda}) = \left(\frac{2.5}{3.5}\right)^{\hat{\lambda}} = \left(\frac{2.5}{3.5}\right)^{1.8907} \simeq 0.5293.$$

•

6.4 Distribuições amostrais.

Motivação 6.44 — Distribuição amostral
A caracterização probabilística de estatísticas, de estimadores ou de suas funções revela-se crucial para

- avaliar as propriedades dos estimadores (enviesamento, EQM, eficiência relativa, etc.) e
- obter estimativas intervalares de parâmetros desconhecidos (intervalos de confiança — Cap. 7). •

Definição 6.45 — Distribuição amostral
A distribuição de uma estatística, estimador ou sua função é denominada de distribuição amostral (ou distribuição por amostragem). •

Proposição 6.46 — Duas distribuições amostrais
Seja $\underline{X} = (X_1, \ldots, X_n)$ uma a.a. de dimensão n proveniente da população X com f.d. $F_X(x)$. Então

Estatística	Distribuição amostral
$X_{(1)} = \min_{i=1,\ldots,n} X_i$	$F_{X_{(1)}}(x) = 1 - [1 - F_X(x)]^n$
$X_{(n)} = \max_{i=1,\ldots,n} X_i$	$F_{X_{(n)}}(x) = [F_X(x)]^n$

•

Nota 6.47 — Duas distribuições amostrais
Os resultados da Proposição 6.46 são válidos para qualquer v.a. de interesse, independentemente da sua distribuição ou do seu carácter ser discreto ou contínuo. Para além disso, caso X_i represente a duração da $i-$ésima componente de um sistema constituído por n componentes, tem-se que:

- $X_{(1)} = \min_{i=1,\ldots,n} X_i$ representa a duração de um sistema em série;
- $X_{(n)} = \max_{i=1,\ldots,n} X_i$ representa a duração de um sistema em paralelo. •

Exemplo/Exercício 6.48 — Duas distribuições amostrais

Demonstre a Proposição 6.46.

- **V.a.**

  $X_i \overset{i.i.d.}{\sim} X,\ i = 1, \ldots, n$

- **F.d. de X**

  $F_X(x) = P(X \leq x),\ -\infty < x < +\infty$

- **Nova v.a.**

  $X_{(1)} = \min_{i=1,\ldots,n} X_i$

- **Distribuição amostral de $X_{(1)}$**

$$
\begin{aligned}
F_{X_{(1)}}(x) &= P[X_{(1)} \leq x] \\
&= 1 - P[X_{(1)} > x] \\
&= 1 - P(X_i > x, i = 1, \ldots, n) \\
&\overset{X_i\ indep}{=} 1 - \prod_{i=1}^{n} P(X_i > x) \\
&\overset{X_i \sim X}{=} 1 - \prod_{i=1}^{n} P(X > x) \\
&= 1 - [P(X > x)]^n \\
&= 1 - [1 - F_X(x)]^n
\end{aligned}
$$

- **Outra v.a.**

  $X_{(n)} = \max_{i=1,\ldots,n} X_i$

- **Distribuição amostral de $X_{(n)}$**

$$
\begin{aligned}
F_{X_{(n)}}(x) &= P[X_{(n)} \leq x] \\
&= P(X_i \leq x, i = 1, \ldots, n) \\
&\overset{X_i\ indep}{=} \prod_{i=1}^{n} P(X_i \leq x) \\
&\overset{X_i \sim X}{=} \prod_{i=1}^{n} P(X \leq x) \\
&= [P(X \leq x)]^n \\
&= [F_X(x)]^n.
\end{aligned}
$$

•

Exercício 6.49 — Distribuições amostrais

Considere que um sistema mecânico é composto por 5 componentes cujos tempos até falha são v.a. i.i.d. a $X \sim$ exponencial(λ) e valor esperado comum igual 1000 horas.

(a) Calcule a probabilidade de o tempo até falha do sistema exceder 2500 horas ao admitir que as componentes foram colocadas em paralelo.

- **V.a.**

  $X_i =$ duração da componente i, $i = 1, \ldots, 5$

  $X_i \stackrel{i.i.d.}{\sim} X$, $i = 1, \ldots, 5$

- **Distribuição de** X

  $X \sim$ exponencial(λ)

- **Parâmetro**

$$
\begin{aligned}
\lambda \ : \ & E(X) = 1000 \\
& \frac{1}{\lambda} = 1000 \\
& \lambda = 0.001
\end{aligned}
$$

- **F.d. de** X

$$
F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-0.001x}, & x \geq 0 \end{cases}
$$

- **Duração do sistema em paralelo**

  $X_{(5)} = \max_{i=1,\ldots,5} X_i$

- **F.d. de** $X_{(5)}$

$$
\begin{aligned}
F_{X_{(5)}}(x) &= [F_X(x)]^5 \\
&= \left(1 - e^{-0.001x}\right)^5, x \geq 0
\end{aligned}
$$

- **Probabilidade pedida**

$$
\begin{aligned}
P[X_{(5)} > 2500] &= 1 - F_{X_{(5)}}(2500) \\
&= 1 - \left(1 - e^{-0.001 \times 2500}\right)^5 \\
&= 0.348353.
\end{aligned}
$$

(b) Volte a calcular a probabilidade solicitada em (a) admitindo que as componentes foram colocadas em série.

- **Duração do sistema em série**

  $X_{(1)} = \min_{i=1,\dots,5} X_i$

- **F.d. de** $X_{(1)}$

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - [1 - F_X(x)]^5 \\ &= 1 - [1 - (1 - e^{-0.001x})]^5 \\ &= 1 - e^{-5\times 0.001x} \\ &= F_{exp(5\times 0.001)}(x), \ x \geq 0 \end{aligned}$$

- **Nota**

  $X_i \overset{i.i.d.}{\sim} \text{exponencial}(\lambda), i = 1, \dots, n \Rightarrow$

  $X_{(1)} = \min_{i=1,\dots,n} X_i \sim \text{exponencial}(n \times \lambda)$.

- **Probabilidade pedida**

$$\begin{aligned} P[X_{(1)} > 2500] &= 1 - F_{X_{(1)}}(2500) \\ &= 1 - (1 - e^{-5\times 0.001 \times 2500}) \\ &\simeq 3.7 \times 10^{-6} \end{aligned}$$

(c) Comente os valores obtidos em (a) e (b).

- **Comentário** — Constata-se que

$$P[X_{(1)} > 2500] << P[X_{(5)} > 2500], \tag{6.24}$$

  confirmando um facto já bem conhecido: os sistemas em série têm duração (estocasticamente) menor que os sistemas em paralelo. •

6.5 Distribuições amostrais de médias.

Motivação 6.50 — Distribuições amostrais da média
A média é de um modo geral o estimador de MV do valor esperado de qualquer v.a. de interesse,[20] pelo que é fundamental saber qual a sua distribuição exacta que, refira-se, nem sempre é de fácil obtenção. •

Proposição 6.51 — Duas distribuições amostrais da média
Seja $\underline{X} = (X_1, \ldots, X_n)$ uma a.a. de dimensão n proveniente da população X. Então

População	Distribuição amostral da média
$X \sim \text{normal}(\mu, \sigma^2)$	$\bar{X} \sim \text{normal}(\mu, \sigma^2/n) \Leftrightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{normal}(0,1)$
X com distrib. arbitrária (não normal), $E(X) = \mu$, $V(X) = \sigma^2 < +\infty$, n grande	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim}_{TLC} \text{normal}(0,1)$

•

Nota 6.52 — Duas distribuições amostrais da média
O primeiro dos dois resultados da Proposição 6.51 é um resultado exacto e deve-se ao facto de a combinação linear de v.a. independentes e com distribuições normais ainda possuir distribuição normal.

O segundo resultado é aproximado, deve-se ao Teorema do Limite Central e só deve ser aplicado quando a v.a. de interesse não possui distribuição normal e a dimensão da amostra é suficientemente grande. •

Exemplo 6.53 — Uma distribuição amostral (aproximada) da média da a.a.
Admita que o desvio absoluto de uma medição instrumental em relação a uma norma é uma v.a. X com distribuição exponencial com parâmetro λ desconhecido.

Calcule $E(\bar{X})$ e $V(\bar{X})$, onde $\bar{X}$ representa, naturalmente, a média de uma amostra aleatória de dimensão n proveniente da população X. Tirando partido dos resultados anteriores mostre que, para n suficientemente grande, se tem[21]

$$Z = (\lambda \bar{X} - 1)\sqrt{n} \quad \stackrel{a}{\sim} \quad \text{normal}(0,1).$$

[20] Ou está de algum modo relacionada com esse estimador.
[21] Adaptado do Exame de 18 de Janeiro de 2003.

- **V.a.**

  X =desvio absoluto de uma medição instrumental em relação a uma norma

- **Distribuição**

  $X \sim$ exponencial(λ)

- **Parâmetro**

  λ DESCONHECIDO ($\lambda > 0$)

- **Nova v.a.**

  $Z = (\lambda\bar{X} - 1)\sqrt{n}$

- **Distribuição aproximada de Z**

  Comece-se por notar que neste caso $E(X) = 1/\lambda$ e $V(X) = 1/\lambda^2$, pelo que

$$\begin{aligned} E(\bar{X}) &= E(X) \\ &= \frac{1}{\lambda} \\ V(\bar{X}) &= \frac{V(X)}{n} \\ &= \frac{1}{n\,\lambda^2} \\ &< +\infty. \end{aligned}$$

  Então, de acordo com o TLC, pode afirmar-se que, para n suficientemente grande ($n > 30$),

$$\begin{aligned} Z &= \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} \\ &= \frac{\bar{X} - \frac{1}{\lambda}}{\sqrt{\frac{1}{n\,\lambda^2}}} \\ &= (\lambda\bar{X} - 1)\sqrt{n} \\ &\stackrel{a}{\sim} \text{normal}(0, 1). \end{aligned}$$

•

Teremos ocasião de estudar outras distribuições amostrais da média da a.a. que serão oportunamente introduzidas à medida que forem necessárias no capítulo seguinte.

Capítulo 7

Estimação por intervalos

7.1 Noções básicas.

Motivação 7.1 — **Intervalos de confiança**
Para além de uma estimativa pontual para o parâmetro desconhecido, é importante adiantar um intervalo que dê uma ideia da confiança que se pode depositar na estimativa pontual.

Este intervalo ou estimativa intervalar é, de um modo geral, denominado de intervalo de confiança. E os valores mais usuais para o grau de confiança são: 90%, 95% e 99%. •

Exemplo 7.2 — **Intervalo de confiança**
Admita que a resistência de uma componente electrónica (em *ohm*, Ω) é uma v.a. $X \sim \text{normal}(\mu, \sigma^2)$, onde μ é desconhecido e σ é conhecido e igual a 4Ω.

Para obter-se informação sobre o valor esperado da resistência da componente, μ, recolheu-se uma amostra de dimensão $n = 4$, tendo-se registado o seguinte conjunto de observações: $\underline{x} = (5.0, 8.5, 12.0, 15.0)$.

(a) Obtenha a estimativa de MV de μ.

$$\hat{\mu} = \bar{x} = \frac{1}{4} \times (5.0 + 8.5 + 12.0 + 15.0) = 10.125\,\Omega.$$

(b) Obtenha um intervalo de confiança a 95% para μ.

É sabido que $EMV(\mu) = \bar{X}$. Mais ainda, que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{normal}(0, 1). \tag{7.1}$$

Importa notar que a v.a. Z depende de

- $EMV(\mu) = \bar{X}$
- μ,

no entanto, possui distribuição independente de μ. Assim sendo, pode calcular-se, por exemplo, a seguinte probabilidade:

$$\begin{aligned} P(-1.96 \leq Z \leq +1.96) &= \Phi(1.96) - \Phi(-1.96) \\ &= \Phi(1.96) - [1 - \Phi(1.96)] \\ &= 0.9750 - 0.0250 \\ &= 0.9500. \end{aligned} \tag{7.2}$$

Equivalentemente,

$$\begin{aligned} &P\left(-1.96 \leq \tfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq +1.96\right) = 0.9500 \\ &P\left(-1.96 \times \tfrac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq +1.96 \times \tfrac{\sigma}{\sqrt{n}}\right) = 0.9500 \\ &P\left(\bar{X} - 1.96 \times \tfrac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \tfrac{\sigma}{\sqrt{n}}\right) = 0.9500. \end{aligned} \tag{7.3}$$

Importa notar que:

1. O intervalo de extremos aleatórios
$$\left[\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right] \tag{7.4}$$
contém $\{\mu\}$ com probabilidade 0.9500.

2. A concretização deste intervalo para a nossa amostra $\underline{x} = (5.0, 8.5, 12.0, 15.0)$ é dada por:
$$\begin{aligned} &\left[\bar{x} - 1.96 \times \tfrac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \times \tfrac{\sigma}{\sqrt{n}}\right] \\ &= \left[10.125 - 1.96 \times \tfrac{4}{\sqrt{4}}, 10.125 + 1.96 \times \tfrac{4}{\sqrt{4}}\right] \\ &= [6.205\Omega, \ 14.045\Omega]. \end{aligned} \tag{7.5}$$

3. μ pertence ao intervalo $[6.205\Omega, 14.045\Omega]$ ou com probabilidade 1 (um) ou com probabilidade 0 (zero), i.e., μ pertence ou não ao intervalo e nada mais podemos adiantar por desconhecimento do verdadeiro valor de μ.

4. Caso recolhêssemos um grande número de amostras de dimensão $n = 4$ e obtivéssemos os correspondentes intervalos

$$\left[\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right], \tag{7.6}$$

a proporção destes intervalos aos quais pertenceria o verdadeiro valor de μ seria de aproximadamente 95%. (Esquema...) É esta a interpretação frequencista da expressão *confiança a 95%.* •

Definição informal 7.3 — **Intervalo de confiança** (IC)
Um IC para o parâmetro θ é do tipo $[l, u]$, onde l e u representam os respectivos limites inferior ("lower") e superior ("upper"), respectivamente. Estes limites têm ainda a particularidade de serem funções:

- da amostra $\underline{x}$, em particular de uma estimativa pontual de θ;
- dos quantis de probabilidade respeitantes à distribuição de uma v.a. que, apesar de depender de θ e de um estimador de θ, possui distribuição não dependente de θ.[1]

A este intervalo está associado um grau de confiança, usualmente representado por $(1 - \alpha) \times 100\%$, cujos valores mais usuais são 90%, 95% e 99%, i.e., $\alpha = 0.1, 0.05, 0.01$. •

Motivação 7.4 — **Método de obtenção de intervalos de confiança**
É necessário adiantar um método de obtenção sistemática de intervalos de confiança para um parâmetro desconhecido θ. •

Definição 7.5 — **Método da v.a. fulcral**
Antes de mais é necessário descrever a situação com que lidamos, em particular convém identificar:

- a v.a. de interesse X e a respectiva distribuição;
- o parâmetro desconhecido θ para o qual se pretende obter um IC;
- outro eventual parâmetro (conhecido ou desconhecido) da distribuição de X.

[1]Esta definição de IC está claramente associada ao método de obtenção de intervalos de confiança descrito já de seguida. Para mais métodos de obtenção de IC, recomenda-se a leitura de Rohatgi e Saleh (2001, p. 532–559).

Posto isto, o método da v.a. fulcral compreende 4 passos:

- **Passo 1 — Selecção da v.a. fulcral para θ**

  É necessário identificar uma v.a. exclusivamente dependente da a.a. $\underline{X}$ e do parâmetro desconhecido θ, doravante representada por

$$Z = Z(\underline{X}, \theta), \tag{7.7}$$

  que, no entanto, possui distribuição exacta (ou aproximada) independente de θ. Esta v.a. é, por regra, uma função trivial do estimador de MV de θ, consta, de um modo geral, do formulário da disciplina e é usualmente designada de v.a. fulcral para θ.

  Escusado será dizer que a v.a. fulcral para θ, Z, não depende de mais nenhum parâmetro desconhecido senão de θ.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Identificada que foi a v.a. fulcral para θ e a respectiva distribuição exacta (ou aproximada), deve proceder-se à obtenção de um par de quantis dependentes do grau de confiança e, por isso mesmo, representados por a_α e b_α. Assim:

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases} \tag{7.8}$$

  (Esquema...) Deste modo, conclui-se que

$$\begin{cases} a_\alpha = F_Z^{-1}(\alpha/2) \\ b_\alpha = F_Z^{-1}(1 - \alpha/2), \end{cases} \tag{7.9}$$

  onde $F_Z^{-1}(p)$ representa o quantil de ordem p da distribuição exacta (ou aproximada) da v.a. fulcral para θ.[2]

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  De modo a obter um intervalo com extremos aleatórios que contenha $\{\theta\}$ com probabilidade $(1 - \alpha)$ é crucial inverter a dupla desigualdade $a_\alpha \leq Z \leq b_\alpha$ em ordem a θ:

$$\begin{aligned} P(a_\alpha \leq Z \leq b_\alpha) &= 1 - \alpha \\ &\vdots \\ P\left[T_1(\underline{X}) \leq \theta \leq T_2(\underline{X})\right] &= 1 - \alpha, \end{aligned} \tag{7.10}$$

[2] Pela forma como estes quantis estão definidos eles dizem quantis equilibrados. Refira-se ainda que a primeira equação do sistema (7.8) é redundante, pelo que será doravante omitida.

onde $T_1(\underline{X})$ e $T_2(\underline{X})$ são os extremos aleatórios há pouco referidos, não só dependentes da a.a. $\underline{X}$ mas também dos quantis de probabilidade a_α e b_α.

- **Passo 4 — Concretização**

  Neste passo limitamo-nos a substituir, nas expressões de $T_1(\underline{X})$ e $T_2(\underline{X})$,

  $$X_1, \dots, X_n \tag{7.11}$$

  pelas respectivas observações,

  $$x_1, \dots, x_n, \tag{7.12}$$

  obtendo-se deste modo o IC a $(1-\alpha) \times 100\%$ para θ:

  $$IC_{(1-\alpha)\times 100\%}(\theta) = [T_1(\underline{x}),\ T_2(\underline{x})]. \tag{7.13}$$

•

7.2 Intervalos de confiança para o valor esperado, variância conhecida.

Nesta secção, à semelhança do que acontecerá com as seguintes, serão apresentados intervalos de confiança para parâmetros de interesse. Esta apresentação obedecerá um figurino que passa pela identificação da distribuição da nossa v.a. de interesse (população) e dos parâmetros conhecidos e desconhecidos, muito em particular aquele para o qual pretendemos obter um IC. A esta identificação seguem-se a aplicação do método da v.a. fulcral e alguns reparos que entendamos pertinentes.

Na obtenção de intervalos de confiança para o valor esperado de uma população com variância conhecida, distinguiremos dois casos:

1. População normal
2. População com distribuição arbitrária e dimensão da amostra suficientemente grande.

CASO 1 — IC para o valor esperado de população normal com variância conhecida

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO[3]

  σ^2 conhecido.

- **Passo 1 — Selecção da v.a. fulcral para μ**

  Nunca será de mais salientar que a v.a. fulcral para μ encontra-se no formulário e será seleccionada tendo sempre presente que deve depender de μ (e de mais nenhum outro parâmetro desconhecido) e de um estimador de μ e deve possuir distribuição exacta com parâmetros conhecidos, logo independente de μ. Assim:

  $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{normal}(0, 1). \tag{7.14}$$

  (Ver formulário.)

[3] A palavra “desconhecido” está intencionalmente em letras maiúsculas para chamar atenção para o facto de ser este o parâmetro para o qual pretendemos calcular o IC.

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$(a_\alpha, b_\alpha): \quad \begin{cases} P(Z < a_\alpha) = \alpha/2 \\ P(Z > b_\alpha) = \alpha/2 \end{cases}$$

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \tag{7.15}$$

  Os quantis de probabilidade a_α e b_α são simétricos pois a f.d.p. da distribuição normal$(0,1)$ é simétrica em relação à origem.

  a_α e b_α obtêm-se por consulta da tabela de quantis da página 8 do conjunto de tabelas disponibilizadas para esta disciplina.

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$P\left(a_\alpha \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq b_\alpha\right) = 1-\alpha$$

$$P\left(\bar{X} - b_\alpha \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left[\bar{X} - \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}},\ \ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right]. \tag{7.16}$$

Exemplo 7.6 — IC para o valor esperado de população normal com variância conhecida

Retome o Exemplo 7.2 e obtenha agora um intervalo de confiança a 90% para o valor esperado da resistência da componente electrónica. Recorra ao método da v.a. fulcral, passo a passo.

- **V.a.**

  $X =$ resistência da componente electrónica (em Ω)

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ^2 conhecido.

- **Passo 1 — Selecção da v.a. fulcral para μ**

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{normal}(0, 1).$$

já que pretendemos um IC para o valor esperado de população normal com variância conhecida. (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

Dado que o nível de confiança é igual a $(1 - \alpha) \times 100\% = 90\%$ (i.e. $\alpha = 0.1$), lidaremos com os quantis

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.95) \stackrel{tabela}{=} -1.6449 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.95) = 1.6449. \end{cases}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

$$\begin{aligned} & P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ & \vdots \\ & P\left[\bar{X} - \Phi^{-1}(1 - \alpha/2) \times \tfrac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}(1 - \alpha/2) \times \tfrac{\sigma}{\sqrt{n}}\right] = 1 - \alpha. \end{aligned}$$

- **Passo 4 — Concretização**

Ao notar-se que

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} - \Phi^{-1}(1 - \alpha/2) \times \frac{\sigma}{\sqrt{n}},\ \bar{x} + \Phi^{-1}(1 - \alpha/2) \times \frac{\sigma}{\sqrt{n}}\right],$$

onde

$$\begin{aligned} & n = 4 \\ & \bar{x} = 10.125\Omega \\ & \Phi^{-1}(1 - \alpha/2) = 1.6449 \\ & \sigma = 4\Omega \end{aligned}$$

tem-se

$$\begin{aligned} IC_{90\%}(\mu) &= \left[10.125 - 1.6449 \times \frac{4}{\sqrt{4}},\ 10.125 + 1.6449 \times \frac{4}{\sqrt{4}}\right] \\ &= [6.8352\Omega, 13.4148\Omega]. \end{aligned}$$

•

CASO 2 — IC aproximado para o valor esperado de população arbitrária com variância conhecida

- **Situação**

  X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$

  μ DESCONHECIDO

  σ^2 conhecido

  n suficientemente grande para justificar o recurso ao Teorema do Limite Central (TLC).

- **Passo 1 — Selecção da v.a. fulcral para** μ

  Ao aplicar o TLC conclui-se que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} \text{normal}(0,1). \tag{7.17}$$

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Uma vez que só conhecemos a distribuição aproximada de v.a. fulcral para μ, Z, os quantis a_α e b_α definidos abaixo enquadram esta v.a. com probabilidade aproximadamente igual a $(1 - \alpha)$. Com efeito,

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \Rightarrow \begin{cases} P(Z < a_\alpha) \simeq \alpha/2 \\ P(Z > b_\alpha) \simeq \alpha/2 \end{cases} \tag{7.18}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

$$\begin{aligned} & P(a_\alpha \leq Z \leq b_\alpha) \simeq 1 - \alpha \\ & \vdots \\ & P\left[\bar{X} - \Phi^{-1}(1-\alpha/2) \times \tfrac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}(1-\alpha/2) \times \tfrac{\sigma}{\sqrt{n}}\right] \simeq 1 - \alpha \end{aligned}$$

- **Passo 4 — Concretização**

  O IC

$$IC(\mu) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}},\ \ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right] \tag{7.19}$$

  possui grau de confiança aproximadamente igual a $(1 - \alpha) \times 100\%$ já que só se conhece a distribuição aproximada da v.a. fulcral de μ.[4] Não surpreende pois que este intervalo seja denominado de IC aproximado para μ. •

[4] Daí ter-se retirado o índice $(1 - \alpha) \times 100\%$ de IC.

Exemplo 7.7 — IC aproximado para o valor esperado de população arbitrária com variância conhecida

Ao estudar-se a densidade de construção (X) num projecto de urbanização foi recolhida uma amostra de 50 lotes desse projecto, tendo-se obtido $\sum_{i=1}^{50} x_i = 227.2$.[5]

(a) Assumindo que o desvio-padrão de X é igual a 4, obtenha um intervalo de confiança aproximadamente igual a 95% para o valor esperado da densidade de construção.

- **V.a.**

  X = densidade de construção em projecto de urbanização

- **Situação**

  X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$

  μ DESCONHECIDO

  σ^2 conhecido

  $n = 50 > 30$, pelo que a dimensão da amostra é suficientemente grande.

- **Passo 1 — Selecção da v.a. fulcral para μ**

  Deve usar a seguinte v.a. fulcral para μ

  $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim}_{TLC} \text{normal}(0,1)$$

  dado que foi solicitado um IC aproximado para o valor esperado de população arbitrária com variância conhecida e a dimensão da amostra ($n = 50$) é suficientemente grande para justificar o recurso ao TLC.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Ao considerar-se um nível de confiança aproximadamente igual a $(1 - \alpha) \times 100\% = 95\%$ (i.e., $\alpha = 0.05$), faremos uso dos quantis

  $$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.975) \overset{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  $$P(a_\alpha \leq Z \leq b_\alpha) \simeq 1 - \alpha$$
  $$\vdots$$
  $$P\left[\bar{X} - \Phi^{-1}(1 - \alpha/2) \times \tfrac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}(1 - \alpha/2) \times \tfrac{\sigma}{\sqrt{n}}\right] \simeq 1 - \alpha$$

[5] Adaptado do Exame de 19 de Janeiro de 2002.

- **Passo 4 — Concretização**

  Tendo em conta que a expressão do IC aproximado a $(1-\alpha)\times 100\%$ é

$$IC(\mu) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2)\times\frac{\sigma}{\sqrt{n}},\ \ \bar{x} + \Phi^{-1}(1-\alpha/2)\times\frac{\sigma}{\sqrt{n}}\right],$$

  onde

$$\begin{array}{l} n = 50 \\ \bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 227.2/50 = 4.544 \\ \Phi^{-1}(1-\alpha/2) = 1.9600 \\ \sigma = 4, \end{array}$$

  segue-se

$$\begin{aligned} IC(\mu) &= \left[4.544 - 1.9600\times\frac{4}{\sqrt{50}},\ 4.544 + 1.9600\times\frac{4}{\sqrt{50}}\right] \\ &= [3.435, 5.653]. \end{aligned}$$

(b) Que dimensão deveria ter a amostra para que a amplitude do intervalo aproximado fosse reduzida para metade?

- **Obtenção da nova dimensão da amostra**

  Em primeiro lugar note-se que a amplitude de um intervalo de confiança aproximadamente igual a $(1-\alpha)\times 100\%$ é dada por

$$\left[\bar{x} + \Phi^{-1}(1-\alpha/2)\times\tfrac{\sigma}{\sqrt{n}}\right] - \left[\bar{x} - \Phi^{-1}(1-\alpha/2)\times\tfrac{\sigma}{\sqrt{n}}\right]$$

$$= 2\,\Phi^{-1}(1-\alpha/2)\,\tfrac{\sigma}{\sqrt{n}},$$

  ao recorrer-se a uma amostra de dimensão n.

  Seja n^* a nova dimensão de amostra que reduz tal amplitude para metade. Então

$$\begin{aligned} n^* \ &:\ 2\,\Phi^{-1}(1-\alpha/2)\,\frac{\sigma}{\sqrt{n^*}} = \frac{1}{2}\times\left[2\,\Phi^{-1}(1-\alpha/2)\,\frac{\sigma}{\sqrt{n}}\right] \\ &\quad \frac{1}{\sqrt{n^*}} = \frac{1}{2}\times\frac{1}{\sqrt{n}} \\ &\quad n^* = 4n \\ &\quad n^* = 200. \end{aligned}$$

  É necessário quadruplicar a dimensão da amostra para que a amplitude do intervalo aproximado seja reduzida para metade. •

Nota 7.8 — **Intervalos de confiança**

1. A amplitude do IC está relacionada com a precisão das estimativas pontual e intervalar. Com efeito, quanto menor a amplitude do IC (com $(1-\alpha)\times 100\%$ fixo) mais precisas são estas estimativas.

2. Ao aumentarmos a dimensão da amostra, mantendo o grau de confiança $(1-\alpha)\times 100\%$ fixo, diminuimos a amplitude do IC e aumentamos a precisão destas estimativas.

3. Ao aumentarmos $(1-\alpha)\times 100\%$, mantendo a dimensão da amostra n fixa, aumentamos a amplitude do IC, logo diminuimos a precisão das estimativas.

4. Não faz sentido considerar $(1-\alpha)\times 100\% = 100\%$ (i.e., $\alpha = 0$) por este valor conduzir a um intervalo de confiança absolutamente inútil: $IC = \Theta$. •

7.3 Intervalos de confiança para a diferença de dois valores esperados, variâncias conhecidas.

Motivação 7.9 — IC para a diferença de valores esperados
É frequente desejar confrontar os valores esperados dos desempenhos de duas máquinas ou de dois métodos, etc.[6] A obtenção de um IC para a diferença entre esses valores esperados revelar-se-á de extrema utilidade, nomeadamente, para inferir da superioridade (ou não) de desempenho de um/a método/máquina. •

Tal como aconteceu na secção anterior, distinguiremos dois casos, desta vez

1. Duas populações independentes, com distribuições normais

2. Duas populações independentes, com distribuições arbitrárias e dimensões das duas amostras suficientemente grandes,

sendo que as variâncias são, em qualquer dos dois casos, conhecidas.

CASO 1 — IC para a diferença de valores esperados de duas populações normais independentes com variâncias conhecidas

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 conhecidos.

- Obs. — Admita-se que $\bar{X}_i$ representa a média da a.a. i de dimensão n_i respeitante à v.a. de interesse/população X_i $(i = 1, 2)$.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  A v.a. fulcral para $(\mu_1 - \mu_2)$ encontra-se no formulário e tem a particularidade de depender de $(\mu_1 - \mu_2)$ (e de mais nenhum outro parâmetro desconhecido) e de um estimador de $(\mu_1 - \mu_2)$, $(\bar{X}_1 - \bar{X}_2)$. Para além disso, possui distribuição exacta conhecida e independente de $(\mu_1 - \mu_2)$. Assim, lidaremos com

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{normal}(0, 1) \qquad (7.20)$$

[6] Passamos pois a lidar com duas variáveis de interesse/populações.

já que $\bar{X}_i \sim_{indep.} \text{normal}(\mu_i, \sigma_i^2/n_i), i = 1, 2.$[7] (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \tag{7.21}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  A inversão faz-se de modo análogo...

$$P\left[a_\alpha \leq \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \leq b_\alpha\right] = 1-\alpha$$

$$\ldots$$

$$P\Bigg[(\bar{X}_1-\bar{X}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq \mu_1-\mu_2$$
$$\leq (\bar{X}_1-\bar{X}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\Bigg] = 1-\alpha$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha)\times 100\%}(\mu_1-\mu_2) = \Bigg[(\bar{x}_1-\bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},$$
$$(\bar{x}_1-\bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\Bigg]. \tag{7.22}$$

Nota 7.10 — IC para o valor esperado ou a diferença de valores esperados
Dado que a distribuição normal possui f.d.p. simétrica em relação à origem, minizaremos a amplitude esperada do IC — logo aumentaremos a precisão do IC — ao tomarmos quantis simétricos. Aliás, é por isso que temos considerado até ao momento quantis de probabilidade simétricos/equilibrados. •

[7] Para justificar estes resultados basta notar que: $\bar{X}_i$ é uma combinação linear de v.a. normais independentes, logo também com distribuição normal; $\bar{X}_1$ e $\bar{X}_2$ são independentes uma vez que assumimos que as duas a.a. também o são. Por fim, resta invocar o fecho da distribuição normal para somas/diferenças de v.a. independentes para concluir que $(\bar{X}_1 - \bar{X}_2)$ também tem distribuição normal, com valor esperado $E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$ e variância $V(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

Exemplo 7.11 — IC para a diferença de valores esperados de duas populações normais independentes com variâncias conhecidas

Uma amostra de 10 peixes foi recolhida de um lago e medidas as concentrações de PCB. Os resultados em *ppm* (partes por milhão) foram para o Lago 1: 11.5, 10.8, 11.6, 9.4, 12.4, 11.4, 12.2, 11.0, 10.6, 10.8.

Por outro lado, foi recolhida uma amostra de 8 peixes de outro lago e medidas também as concentrações de PCB, tendo-se obtido os valores seguintes para o Lago 2: 11.8, 12.6, 12.2, 12.5, 11.7, 12.1, 10.4, 12.6.

Considere que estes dois grupos de medições são concretizações de duas a.a. independentes provenientes de distribuições normais com valores esperados desconhecidos μ_1 e μ_2 e variâncias $\sigma_1^2 = 0.09$ e $\sigma_2^2 = 0.06$, respectivamente.

Obtenha um IC a 99% para a diferença dos valores esperados das concentrações de PCB nestes dois lagos. Comente.[8]

- **V.a.**

  $X_i =$ concentração de PCB no lago i $(i = 1, 2)$

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 conhecidos.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Uma vez que se tem em vista a obtenção de IC para a diferença de valores esperados de duas populações normais independentes com variâncias conhecidas, faremos uso da seguinte v.a. fulcral para $(\mu_1 - \mu_2)$:

  $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{normal}(0, 1).$$

  (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Dado que o nível de confiança é igual a $(1 - \alpha) \times 100\% = 99\%$ (i.e., $\alpha = 0.01$), utilizaremos os quantis

[8] Adaptado do Exame de 25 de Junho de 2002.

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.995) \stackrel{tabela}{=} -2.5758 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.995) = 2.5758. \end{cases}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omite-se este passo...

- **Passo 4 — Concretização**

  Ao ter em conta que a expressão geral deste IC para $(\mu_1 - \mu_2)$ é

$$IC_{(1-\alpha)\times 100\%}(\mu_1 - \mu_2) = \left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \right.$$
$$\left. (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right].$$

  e o facto de

$$n_1 = 10, \quad \bar{x}_1 = 11.17, \quad \sigma_1^2 = 0.09$$
$$n_2 = 8, \quad \bar{x}_2 = 11.9875, \quad \sigma_2^2 = 0.06$$
$$\Phi^{-1}(1-\alpha/2) = 2.5758$$

  conclui-se que

$$\begin{aligned} IC_{99\%}(\mu_1 - \mu_2) &= \left[(11.17 - 11.9875) - 2.5758 \times \sqrt{\frac{0.09}{10} + \frac{0.06}{8}}, \right. \\ &\quad \left. (11.17 - 11.9875) + 2.5758 \times \sqrt{\frac{0.09}{10} + \frac{0.06}{8}}\right] \\ &= [-1.1484, -0.4866]. \end{aligned}$$

- **Comentário** — A estimativa de MV de $(\mu_1 - \mu_2)$ é igual a

$$\bar{x}_1 - \bar{x}_2 = -0.8175 \neq 0.$$

  Mais,

$$IC_{99\%}(\mu_1 - \mu_2) \not\supset \{0\},$$

  pelo que o "zero"não é um dos valores razoáveis para $(\mu_1 - \mu_2)$. Ora, se tivermos em conta tudo isto e ainda o facto de

$$\mu_1 - \mu_2 \neq 0 \Leftrightarrow \mu_1 \neq \mu_2,$$

  não parece disparatado afirmar que as concentrações de PCB diferem de um lago para outro. •

CASO 2 — IC aproximado para a diferença de valores esperados de duas populações arbitrárias (não normais) independentes com variâncias conhecidas

- **Situação**

  $X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (não normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1, 2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 conhecidos

  n_1 e n_2 suficientemente grandes.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Recorrendo ao TLC obtém-se a v.a. fulcral para $(\mu_1 - \mu_2)$ que por sinal não se encontra no formulário por tratar-se de uma variante da v.a. fulcral usada no caso anterior. Escusado será dizer que depende de $(\mu_1 - \mu_2)$ e de um estimador de $(\mu_1 - \mu_2)$, para além de possuir distribuição aproximadamente normal padrão e como tal independente de $\mu_1 - \mu_2$.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \overset{a}{\sim}_{TLC} \text{normal}(0, 1). \tag{7.23}$$

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Os quantis

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) \end{cases} \tag{7.24}$$

  enquadram a v.a. fulcral para $(\mu_1 - \mu_2)$ com probabilidade aproximadamente igual a $(1 - \alpha)$.

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  A inversão da desigualdade é análoga...

- **Passo 4 — Concretização**

$$\begin{aligned} IC(\mu_1 - \mu_2) = \Bigg[& (\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \\ & (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Bigg] \end{aligned} \tag{7.25}$$

  possui grau de confiança aproximadamente igual a $(1 - \alpha) \times 100\%$.

Exemplo 7.12 — IC aproximado para a diferença de valores esperados de duas populações arbitrárias (não normais) independentes com variâncias conhecidas Foram efectuados estudos em Los Angeles e Nova York com o objectivo de determinar a concentração de monóxido de carbono (CO) perto de vias rápidas. Para o efeito foram recolhidas amostras de ar para as quais se determinou a respectiva concentração de CO em *ppm* (partes por milhão).

(a) Obtenha um IC aproximado a 95% para a diferença dos valores esperados das concentrações de CO nestas duas cidades tendo em conta a seguinte informação:

Los Angeles	$n_1 = 40$	$\bar{x}_1 = 117.77$	$\sigma_1^2 = 10$
Nova York	$n_2 = 50$	$\bar{x}_2 = 97.14$	$\sigma_2^2 = 32$

- **V.a.**

  $X_1 =$ concentração de CO perto de vias rápidas em Los Angeles

  $X_2 =$ concentração de CO perto de vias rápidas em Nova York

- **Situação**

  $X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (não normais):

  $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1, 2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 conhecidos

  $n_1 = 40 > 30$ e $n_2 = 50 > 30$ logo amostras são suficientemente grandes.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Já que se pretende IC aproximado para a diferença de valores esperados de duas populações arbitrárias (não normais) independentes com variâncias conhecidas e as dimensões das amostras justificam o recurso ao TLC, usar-se-á a seguinte v.a. fulcral:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \overset{a}{\sim} \text{normal}(0, 1).$$

- **Passo 2 — Obtenção dos quantis de probabilidade**

  O seguinte par de quantis enquadra a v.a. fulcral Z com probabilidade aproximadamente igual a $(1 - \alpha) = 0.95$ (i.e., $\alpha = 0.05$):

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.975) \overset{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omite-se este passo...

- **Passo 4 — Concretização**

  Tirando partido do facto de a expressão geral do IC aproximado a $(1-\alpha)\times 100\%$ ser

$$IC(\mu_1-\mu_2) = \left[(\bar{x}_1-\bar{x}_2) - \Phi^{-1}(1-\alpha/2)\times\sqrt{\tfrac{\sigma_1^2}{n_1}+\tfrac{\sigma_2^2}{n_2}},\right.$$
$$\left.(\bar{x}_1-\bar{x}_2) + \Phi^{-1}(1-\alpha/2)\times\sqrt{\tfrac{\sigma_1^2}{n_1}+\tfrac{\sigma_2^2}{n_2}}\right]$$

  e de

$$n_1 = 40, \quad \bar{x}_1 = 117.77, \quad \sigma_1^2 = 10$$
$$n_2 = 50, \quad \bar{x}_2 = 97.14, \quad \sigma_2^2 = 32$$
$$\Phi^{-1}(1-\alpha/2) = 1.9600,$$

  obtém-se

$$\begin{aligned} IC(\mu_1-\mu_2) &= \left[(117.77-97.14) - 1.9600\times\sqrt{\frac{10}{40}+\frac{32}{50}},\right. \\ &\qquad \left.(117.77-97.14) + 1.9600\times\sqrt{\frac{10}{40}+\frac{32}{50}}\right] \\ &= [18.7809, 22.4791]. \end{aligned}$$

(b) Num programa televisivo em que se debatiam questões ambientais, o presidente da câmara de Nova York afirmou que: “A concentração esperada de CO em Nova York não difere da concentração esperada em Los Angeles.”Comente esta afirmação.

- **Comentário** — A estimativa de MV de $(\mu_1-\mu_2)$ é superior a 0. Com efeito,

$$\bar{x}_1 - \bar{x}_2 = 20.63$$

  e para além disso

$$IC(\mu_1-\mu_2) \not\supset \{0\},$$

  pelo que o presidente da câmara de Nova York deveria ter afirmado que a concentração esperada de CO em Los Angeles é superior à concentração esperada em New York. •

7.4 Intervalos de confiança para o valor esperado, variância desconhecida.

Motivação 7.13 — IC para o valor esperado com variância desconhecida
Na maior parte das situações desconhecemos quer o valor esperado, quer a variância da nossa v.a. de interesse. Neste caso $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{normal}(0,1)$ deixa de ser uma v.a. fulcral para μ já que depende de σ e este parâmetro é desconhecido. •

Por forma a adiantar uma v.a. fulcral para μ quando lidamos com uma população normal com variância desconhecida, será preciso enunciar alguns resultados auxiliares que condensaremos na seguinte proposição.

Proposição 7.14 — Distribuições do qui-quadrado e de t-Student
Considere-se $X_i \sim_{i.i.d.} \text{normal}(\mu, \sigma^2), i = 1, \ldots, n$. Então:

1. $\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim_{i.i.d.} \chi^2_{(1)}, i = 1, \ldots, n$ [9]

2. $\sum_{i=1}^{n} \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$ [10]

3. $\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)}{\sigma^2} \times \frac{1}{n-1} \sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2_{(n-1)}$

4. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \perp\!\!\!\perp \frac{(n-1)S^2}{\sigma^2}$

Sejam $W \sim \text{normal}(0,1)$ e $Y \sim \chi^2_{(\nu)}$ duas v.a. independentes ($W \perp\!\!\!\perp Y$). Então:

5. $\frac{W}{\sqrt{Y/\nu}} \sim t_{(\nu)}$ [11]

6. $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} \stackrel{d}{=} \frac{\text{normal}(0,1)}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}} \sim t_{(n-1)}$, já que o numerador é uma v.a. independente do denominador. •

Nota 7.15 — Distribuições do qui-quadrado e de t-Student
Os quantis de probabilidade das distribuições do qui-quadrado e de t-Student constam das páginas 9 e 10 das tabelas disponibilizadas nesta disciplina. •

---

[9] Leia-se: "a v.a. ... possui distribuição do qui-quadrado com um grau de liberdade".
[10] Leia-se: "a v.a. ... possui distribuição do qui-quadrado com n grau de liberdade". Trata-se de uma consequência da propriedade reprodutiva da v.a. do qui-quadrado.
[11] Leia-se: "a v.a. ... possui distribuição de t-Student com ν grau de liberdade"

CASO 1 — IC para o valor esperado de população normal com variância desconhecida

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ^2 desconhecido.

- **Passo 1 — Selecção da v.a. fulcral para μ**

  Recorrendo à Proposição 7.14 podemos adiantar uma v.a. fulcral para μ:

  $$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}. \tag{7.26}$$

  (Ver formulário.) Com efeito, esta v.a. depende de μ mas não do parâmetro desconhecido σ, para além de depender de um estimador de μ, $\bar{X}$, e de um de σ, S. Mais, possui distribuição exacta conhecida e independente de μ e de σ^2.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Uma vez que a distribuição de t-Student com $(n-1)$ graus de liberdade é igualmente simétrica em relação à origem, minimizaremos a amplitude esperada do IC ao escolher quantis de probabilidade simétricos:

  $$\begin{cases} a_\alpha = -F^{-1}_{t_{(n-1)}}(1-\alpha/2) \\ b_\alpha = F^{-1}_{t_{(n-1)}}(1-\alpha/2). \end{cases} \tag{7.27}$$

  (Esquema e consulta de tabelas...)

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  Similar à da obtenção de IC para μ com σ conhecido.

- **Passo 4 — Concretização**

  $$\begin{aligned} & IC_{(1-\alpha)\times 100\%}(\mu) \\ & = \left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \tfrac{s}{\sqrt{n}},\ \ \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \tfrac{s}{\sqrt{n}}\right], \end{aligned} \tag{7.28}$$

  onde, recorde-se, $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \times [(\sum_{i=1}^n x_i^2) - n(\bar{x})^2]$.

Nota 7.16 — IC para o valor esperado de população normal, com variância desconhecida
O IC em (7.28) é similar ao obtido para μ com σ conhecido em (7.16). De facto ocorreram as seguintes substituições:

- $\Phi^{-1}(1-\alpha/2) \quad \longrightarrow \quad F^{-1}_{t_{(n-1)}}(1-\alpha/2)$;
- $\sigma \quad \longrightarrow \quad s$.

E mais uma vez temos um IC do tipo:

- estimativa pontual $\pm$ quantil $\times$ concretização do denominador da v.a. fulcral. •

Exemplo 7.17 — IC para o valor esperado de população normal com variância desconhecida
A duração (em horas) de certa marca de pilha para máquinas fotográficas digitais possui distribuição que se admite normal com valor esperado μ e variância σ^2 desconhecidos.

Um revendedor seleccionou ao acaso 10 pilhas de um lote, tendo obtido o seguinte conjunto de dados: $\underline{x} = (251, 238, 236, 229, 252, 253, 245, 242, 235, 230)$. Tire partido do facto de $\sum_{i=1}^{10} x_i = 2411$ e $\sum_{i=1}^{10} x_i^2 = 582009$ e obtenha um intervalo de confiança a 99% para μ.

- **V.a.**

  $X =$ duração (em horas) de certa marca de pilha

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ^2 desconhecido.

- **Passo 1 — Selecção da v.a. fulcral para μ**

  $$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

  dado que se pretende IC para o valor esperado de população normal com variância desconhecida. (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Como $n = 10$ e $(1-\alpha)\times 100\% = 99\%$ (ou seja, $\alpha = 0.01$), far-se-á uso dos quantis

$$\begin{cases} a_\alpha = -F^{-1}_{t_{(n-1)}}(1-\alpha/2) = -F^{-1}_{t_{(9)}}(0.995) \stackrel{tabela}{=} -3.250 \\ b_\alpha = F^{-1}_{t_{(n-1)}}(1-\alpha/2) = F^{-1}_{t_{(9)}}(0.995) = 3.250 \end{cases}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omite-se novamente este passo...

- **Passo 4 — Concretização**

  Ora, é sabido que neste caso

$$\begin{aligned} IC_{(1-\alpha)\times 100\%}(\mu) &= \left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}},\right. \\ &\qquad \left. \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right], \end{aligned}$$

  onde

$$n = 10$$

$$\bar{x} = \tfrac{1}{n}\sum_{i=1}^n x_i = \tfrac{2411}{10} = 241.1$$

$$\begin{aligned} s^2 &= \tfrac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \tfrac{1}{n-1} \times [(\textstyle\sum_{i=1}^n x_i^2) - n(\bar{x})^2] \\ &= \tfrac{1}{10-1}(582009 - 10 \times 241.1^2) \\ &= 79.66 \end{aligned}$$

$$F^{-1}_{t_{(n-1)}}(1-\alpha/2) = 3.250.$$

  Deste modo conclui-se que

$$\begin{aligned} IC_{99\%}(\mu) &= \left[241.1 - 3.250 \times \sqrt{\frac{79.66}{10}},\ \ 241.1 + 3.250 \times \sqrt{\frac{79.66}{10}}\right] \\ &= [231.927, 250.273]. \end{aligned}$$

•

CASO 2 — IC aproximado para o valor esperado de população arbitrária com variância desconhecida

- **Situação**

  X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$

  μ DESCONHECIDO

  σ^2 desconhecido

  n suficientemente grande.

- **Passo 1 — Selecção da v.a. fulcral para μ**

  A v.a. fulcral a utilizar nesta situação encontra-se no formulário e muito se assemelha a $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$, usada quando a v.a. de interesse possui distribuição normal com ambos os parâmetros desconhecidos. De facto têm a mesma expressão mas distribuições distintas:

  $$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} \text{normal}(0,1). \tag{7.29}$$

  Para a utilização desta v.a. é fundamental que a dimensão da amostra seja suficientemente grande já que este resultado distribucional assenta na aplicação do TLC e do Teorema de Slutsky (Rohatgi e Saleh (2001, p. 270)).

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Os quantis de probabilidade simétricos a utilizar dizem respeito à distribuição aproximada da v.a. fulcral para μ:

  $$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \tag{7.30}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  Análoga à obtenção de IC para o valor esperado de população normal com σ desconhecido.

- **Passo 4 — Concretização**

  O IC aproximado é, neste caso, dado por:

  $$IC(\mu) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}},\ \ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right]. \tag{7.31}$$

Exemplo 7.18 — IC aproximado para o valor esperado de população arbitrária com variância desconhecida

Uma máquina automática é usada para encher e selar latas de um litro de um produto líquido. Na sequência de algumas queixas acerca do facto de as latas se encontrarem demasiado cheias, foi medido o conteúdo de 100 latas seleccionadas ao acaso do fabrico diário, tendo-se obtido $\sum_{i=1}^{100} x_i = 108.983$ e $\sum_{i=1}^{100} x_i^2 = 120.238$.

Deduza um intervalo aproximado de confiança a 95% para o valor esperado μ do conteúdo das latas.[12]

- **V.a.**

  $X =$ conteúdo de lata de um litro de um produto líquido

- **Situação**

  X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$

  μ DESCONHECIDO

  σ^2 desconhecido

  $n = 100 > 30$ (suficientemente grande).

- **Passo 1 — Selecção da v.a. fulcral para μ**

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} \text{normal}(0,1)$$

  pois pretende-se IC aproximado para o valor esperado de população com distribuição arbitrária com variância desconhecida e estamos a lidar com amostra suficientemente grande que justifica invocarmos o TLC e o Teorema de Slutsky. (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Uma vez que o grau de confiança é aproximadamente de $(1-\alpha) \times 100\% = 95\%$ (ou seja, $\alpha = 0.05$) iremos usar os quantis simétricos seguintes:

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.975) \stackrel{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

  Estes quantis dizem respeito à distribuição aproximada da v.a. fulcral para μ e enquandram-na com probabilidade aproximadamente igual a $(1-\alpha) = 0.95$.

[12] Adaptado do Exame de 24 de Junho de 2006.

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Tal como nos exemplos anteriores omite-se este passo...

- **Passo 4 — Concretização**

  O IC aproximado a $(1-\alpha) \times 100\%$ para μ é dado por

$$IC(\mu) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}}, \;\; \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right],$$

  onde

$$n = 100$$

$$\bar{x} = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i = \tfrac{108.983}{100} = 1.08983$$

$$\begin{aligned} s^2 &= \tfrac{1}{n-1}\left[\left(\textstyle\sum_{i=1}^n x_i^2\right) - n(\bar{x})^2\right] \\ &= \tfrac{1}{100-1}(120.238 - 100 \times 1.08983^2) \\ &= 0.0147945 \end{aligned}$$

$$\Phi^{-1}(1-\alpha/2) = 1.9600.$$

  Assim sendo

$$\begin{aligned} IC(\mu) &= \left[1.08983 - 1.9600 \times \sqrt{\frac{0.0147945}{100}}, \right. \\ &\qquad \left. 1.08983 + 1.9600 \times \sqrt{\frac{0.0147945}{100}}\right] \\ &= [1.06599, \; 1.11367]. \end{aligned}$$

•

Antes de prosseguirmos para a secção seguinte convinha realçar que em certas situações lidamos com uma v.a. de interesse cuja distribuição depende de um único parâmetro desconhecido e que quer o valor esperado, quer a variância dependem desse mesmo parâmetro.

Entre essas situações, considere-se, para já, aquela em que, ao invocar o TLC, obtém-se uma v.a. fulcral para o valor esperado cuja expressão é de tal modo simplificada que não é sequer necessário estimar $V(X)$ à custa da variância da a.a. e como tal não é preciso invocar o Teorema de Slutsky.

Ilustraremos esta situação com um exemplo paradigmático em que $X \sim \text{exponencial}(\lambda = 1/\delta)$.

Exemplo 7.19 — IC aproximado para o valor esperado de população arbitrária com variância desconhecida dependente do valor esperado mas que não carece de estimação...

Admita que o tempo entre falhas consecutivas de certo equipamento eléctrico é caracterizado por ter a seguinte f.d.p.:

$$f_X(x) = \begin{cases} \frac{e^{-x/\delta}}{\delta}, & x \geq 0 \\ 0, & c.c., \end{cases}$$

onde $\delta > 0$. Admita ainda que recolheu uma amostra de dimensão $n = 35$ desta população, tendo obtido $\sum_{i=1}^{35} x_i = 77.349$ e $\sum_{i=1}^{35} x_i^2 = 293.452$.

Justifique que

$$\sqrt{n}\left(\frac{\bar{X}}{\delta} - 1\right) \stackrel{a}{\sim} \text{normal}(0,1),$$

onde $\bar{X}$ é o estimador de MV de δ. Tire partido deste resultado e da amostra para construir um intervalo de confiança aproximado a 90% para o valor esperado $E(X) = \delta$.[13]

- **V.a.**

  $X =$ tempo entre falhas consecutivas de certo equipamento eléctrico

- **Situação**

  X com distribuição não normal...

  $X \sim \text{exponencial}(\lambda = 1/\delta)$, $\mu = E(X) = \delta$, $\sigma^2 = V(X) = \delta^2$

  μ DESCONHECIDO

  σ^2 desconhecido

  $n = 35 > 30$ (suficientemente grande).

- **Passo 1 — Selecção da v.a. fulcral para δ**

  De acordo com o TLC, pode afirmar-se que:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - E(X)}{\sqrt{\frac{V(X)}{n}}} = \frac{\bar{X} - \delta}{\sqrt{\frac{\delta^2}{n}}} = \sqrt{n}\left(\frac{\bar{X}}{\delta} - 1\right) \stackrel{a}{\sim} \text{normal}(0,1).$$

[13] Adaptado do Exame de 23 de Junho de 2007.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Ao pretendermos IC aproximado a $(1-\alpha)\times 100\% = 90\%$ (ou seja, $\alpha = 0.1$) é suposto usar os quantis

  $$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.95) \stackrel{tabela}{=} -1.6449 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.95) = 1.6449, \end{cases}$$

  que enquadram a v.a. fulcral para δ com probabilidade aproximadamente igual a $(1-\alpha) = 0.90$.

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  Uma vez que a v.a. fulcral tem expressão em nada semelhantes às anteriores não se omitirá este passo:

  $$P(a_\alpha \leq Z \leq b_\alpha) \simeq 1-\alpha$$

  $$P\left[a_\alpha \leq \sqrt{n}\left(\frac{\bar{X}}{\delta}-1\right) \leq b_\alpha\right] \simeq 1-\alpha$$

  $$P\left(1+\frac{a_\alpha}{\sqrt{n}} \leq \frac{\bar{X}}{\delta} \leq 1+\frac{b_\alpha}{\sqrt{n}}\right) \simeq 1-\alpha$$

  $$P\left(\frac{\bar{X}}{1+\frac{b_\alpha}{\sqrt{n}}} \leq \delta \leq \frac{\bar{X}}{1+\frac{a_\alpha}{\sqrt{n}}}\right) \simeq 1-\alpha$$

  $$P\left[\frac{\bar{X}}{1+\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}} \leq \delta \leq \frac{\bar{X}}{1-\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}}\right] \simeq 1-\alpha.$$

- **Passo 4 — Concretização**

  O IC aproximado a $(1-\alpha)\times 100\%$ para δ possui expressão geral dada por

  $$IC(\delta) = \left[\frac{\bar{x}}{1+\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}},\ \frac{\bar{x}}{1-\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}}\right],$$

  com

  $$n = 35$$

  $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{77.349}{35} = 2.20997$$

  $$\Phi^{-1}(1-\alpha/2) = 1.6449.$$

  Deste modo o IC aproximado a 90% é igual a

  $$\begin{aligned} IC(\delta) &= \left[\frac{2.20997}{1+\frac{1.6449}{\sqrt{35}}},\ \frac{2.20997}{1-\frac{1.6449}{\sqrt{35}}}\right] \\ &= [1.72919,\ 3.06106]. \end{aligned}$$

•

7.5 Intervalos de confiança para a diferença de dois valores esperados, variâncias desconhecidas.

Escusado será dizer que na apresentação que se segue considerar-se-ão dois casos.

1. Lida-se com duas populações normais independentes: $X_1 \sim \text{normal}(\mu_1, \sigma_1^2)$ e $X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$, onde σ_1^2 e σ_2^2 são desconhecidos mas que se assume serem iguais. Para além disso, as dimensões das amostrais não são suficientemente grandes que justifiquem o recurso a um resultado assintótico.

2. Desta feita as duas populações são independentes, com distribuições arbitrárias (possivelmente normais) e valores esperados e variâncias iguais a μ_i e σ_i^2, $i = 1, 2$, onde σ_1^2 e σ_2^2 são desconhecidos mas não necessariamente iguais. Mais, as dimensões das duas amostras, n_1 e n_2, são suficientemente grandes para invocar o TLC e o Teorema de Slutsky.

CASO 1 — IC para a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos, no entanto, assume-se que serem IGUAIS: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

  $n_1 \leq 30$ ou $n_2 \leq 30$.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Desta feita é suposto utilizarmos:

  $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}, \tag{7.32}$$

  onde $\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ é um estimador razoável para a variância comum das populações X_1 e X_2, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. (Ver formulário.)

Para justificar este resultado distribucional basta reescrever a v.a. fulcral, ter presente a forma como é derivada a distribuição t-Student e notar que se assumiu que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Assim, temos:

$$Z = \frac{\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}}{\sqrt{\frac{\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}}{n_1+n_2-2}}} \stackrel{d}{=} \frac{\text{normal}(0,1)}{\sqrt{\frac{\chi^2_{(n_1+n_2-2)}}{n_1+n_2-2}}} \sim t_{(n_1+n_2-2)}. \tag{7.33}$$

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$\begin{cases} a_\alpha = -F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) \\ b_\alpha = F_{t_{(n_1+n_2-2)}}(1-\alpha/2). \end{cases} \tag{7.34}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  A inversão é análoga...

- **Passo 4 — Concretização**

$$\begin{aligned} & IC_{(1-\alpha)\times 100\%}(\mu_1-\mu_2) \\ & = \Bigg[(\bar{x}_1-\bar{x}_2) - F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) \times \sqrt{\tfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}(\tfrac{1}{n_1}+\tfrac{1}{n_2})}, \\ & \qquad (\bar{x}_1-\bar{x}_2) + F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) \times \sqrt{\tfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}(\tfrac{1}{n_1}+\tfrac{1}{n_2})}\Bigg]. \end{aligned} \tag{7.35}$$

Exemplo 7.20 — IC para a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais

Para efectuar uma análise comparativa entre duas marcas de computadores pessoais, análise essa baseada no tempo (em meses) até à necessidade da primeira reparação, recolheram-se observações que conduziram aos resultados

- Amostra 1: $\sum_{i=1}^{20} x_{i\,1} = 248$, $\sum_{i=1}^{20} x_{i\,1}^2 = 3560$
- Amostra 2: $\sum_{i=1}^{12} x_{i\,2} = 152$, $\sum_{i=1}^{12} x_{i\,2}^2 = 2730$.

Admitindo que os tempos são independentes e têm, para ambas as marcas, distribuições normais com variâncias desconhecidas mas iguais, determine um intervalo de confiança a 90% para a diferença dos respectivos valores esperados.[14]

[14] Adaptado do Exame de 13 de Julho de 2002.

- **V.a.**

  $X_i =$ tempo até 1a. reparação de computador pessoal da marca $i, \ i = 1, 2$

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos mas que se assume serem IGUAIS: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

  $n_1 = 20 \leq 30$ ou $n_2 = 12 \leq 30$.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

  uma vez que é suposto determinar um IC EXACTO para a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais. (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Ao ter em conta que, neste caso, $n_1 = 20$, $n_2 = 12$ e $(1 - \alpha) \times 100\% = 90\%$ i.e., $\alpha = 0.1$, usaremos os quantis de probabilidade:

$$\begin{cases} a_\alpha = -F^{-1}_{t_{(n_1+n_2-2)}}(1 - \alpha/2) = -F^{-1}_{t_{(30)}}(0.95) \stackrel{tabela}{=} -1.697 \\ b_\alpha = F_{t_{(n_1+n_2-2)}}(1 - \alpha/2) = F^{-1}_{t_{(30)}}(0.95) = 1.697. \end{cases}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Este passo é mais uma vez omitido...

- **Passo 4 — Concretização**

  Ao notar que

$$\begin{aligned} & IC_{(1-\alpha)\times 100\%}(\mu_1 - \mu_2) \\ &= \Bigg[(\bar{x}_1 - \bar{x}_2) - F^{-1}_{t_{(n_1+n_2-2)}}(1 - \alpha/2) \times \sqrt{\tfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}(\tfrac{1}{n_1} + \tfrac{1}{n_2})}, \\ &\qquad (\bar{x}_1 - \bar{x}_2) + F^{-1}_{t_{(n_1+n_2-2)}}(1 - \alpha/2) \times \sqrt{\tfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}(\tfrac{1}{n_1} + \tfrac{1}{n_2})}\Bigg], \end{aligned}$$

onde

$$n_1 = 20,\ \bar{x}_1 = 12.4,\ s_1^2 = 25.516$$
$$n_2 = 12,\ \bar{x}_2 = 12.67,\ s_2^2 = 73.152$$
$$F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) = 1.697,$$

segue-se

$$\begin{aligned}
&IC_{90\%}(\mu_1-\mu_2)\\
&= \Big[(12.4-12.67) - 1.697 \times \sqrt{\tfrac{(20-1)\times 25.516+(12-1)\times 73.152}{20+12-2}(\tfrac{1}{20}+\tfrac{1}{12})},\\
&\qquad (12.4-12.67) + 1.697 \times \sqrt{\tfrac{(20-1)\times 25.516+(12-1)\times 73.152}{20+12-2}(\tfrac{1}{20}+\tfrac{1}{12})}\Big]\\
&= [-4.33254, 3.79254].
\end{aligned}$$

•

Motivação 7.21 — IC aproximado para a diferença de valores esperados de populações arbitrárias independentes com variâncias desconhecidas
Nem sempre é razoável admitir que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, muito em particular quando as estimativas de σ_1^2 e σ_2^2 são bem distintas. Contudo, ao lidarmos com $n_1 \leq 30$ e $n_2 \leq 30$, não podemos adiantar uma solução satisfatória no âmbito desta disciplina.

Mas a situação muda totalmente de figura caso $n_1 > 30$ e $n_2 > 30$, já que nesta situação podemos invocar dois resultados assintóticos, nomeadamente, o TLC e o Teorema de Slutsky, e deixamos de necessitar de assumir que as duas populações são normais e que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. •

CASO 2 — IC aproximado para a diferença de valores esperados de duas populações arbitrárias independentes com variâncias desconhecidas

- **Situação**

  $X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (possivelmente normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1, 2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos mas não necessariamente iguais

  $n_1 > 30$ e $n_2 > 30$, i.e., ambas as amostras são suficientemente grandes.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Recorrendo ao TLC e ao Teorema de Slutsky obtém-se a v.a. fulcral para $\mu_1 - \mu_2$ que pode encontrar-se no formulário. Trata-se de uma variante da v.a. fulcral usada no caso em que se pretendia IC para a diferença de valores esperados de populações normais com variâncias conhecidas.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \stackrel{a}{\sim} \text{normal}(0,1). \tag{7.36}$$

  (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \tag{7.37}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  A inversão é análoga...

- **Passo 4 — Concretização**

  O IC é dado por

$$IC(\mu_1 - \mu_2) = \left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \right. \\ \left. (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] \tag{7.38}$$

  e possui grau de confiança aproximadamente igual a $(1-\alpha) \times 100\%$.

Exemplo/Exercício 7.22 — IC aproximado para a diferença de valores esperados de duas populações arbitrárias independentes com variâncias desconhecidas

Para comparar a resistência ao uso de dois tipos de materiais cerâmicos usados em pavimentos, foram colocados 81 mosaicos do primeiro tipo e 121 do segundo tipo num corredor movimentado de uma superfície comercial.

Após um ano o seu desgaste foi medido numa escala conveniente. Para os mosaicos do primeiro tipo obteve-se $\bar{x}_1 = 290$ e $s_1 = 12$, enquanto que para os do segundo tipo os resultados foram $\bar{x}_2 = 321$ e $s_2 = 14$.

Assuma que, em ambos os casos, a distribuição do desgaste é normal e que os desgastes em mosaicos de tipo diferente são independentes. Construa um intervalo de confiança a 95% para a diferença entre os valores esperados do desgaste dos dois tipos de materiais, assumindo que

(a) as variâncias são desconhecidas MAS NÃO SÃO NECESSARIAMENTE IGUAIS;[15]

- **V.a.**

  $X_i =$ desgaste do material do tipo $i, i = 1, 2$

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos, no entanto, assume-se que NÃO são IGUAIS

  $n_1 = 81 >> 30$, $n_2 = 121 >> 30$ (suficientemente grandes)

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

  Uma vez que as dimensões das amostras são suficientemente grandes podemos obter um IC aproximado para a diferença de valores esperados de duas populações arbitrárias[16] independentes, com variâncias desconhecidas mas que se assume serem iguais, devemos utilizar a seguinte v.a. fulcral para $(\mu_1 - \mu_2)$:

  $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \stackrel{a}{\sim} \text{normal}(0, 1).$$

  (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Como o nível aproximado de confiança é de $(1 - \alpha) \times 100\% = 95\%$ (i.e., $\alpha = 0.05$), recorre-se aos quantis

  $$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.975) \stackrel{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) \stackrel{tabela}{=} 1.9600 \end{cases}$$

  e é sabido que enquadram a v.a. fulcral Z com probabilidade aproximada $(1 - \alpha)$.

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omitimos este passo...

[15] Adaptado do Exame de 7 de Fevereiro de 2004.

[16] Por sinal estas populações são normais neste caso.

- **Passo 4 — Concretização**

  Neste caso o IC aproximado é dado por

$$IC(\mu_1 - \mu_2) = \left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \right.$$
$$\left. (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right],$$

  onde

$$n_1 = 81,\ \bar{x}_1 = 290,\ s_1^2 = 12^2$$
$$n_2 = 121,\ \bar{x}_2 = 321,\ s_2^2 = 14^2$$
$$\Phi^{-1}(1-\alpha/2) = 1.9600.$$

  Logo obtém-se

$$\begin{aligned} IC(\mu_1 - \mu_2) &= \left[(290-321) - 1.9600 \times \sqrt{\frac{12^2}{81} + \frac{14^2}{121}}, \right. \\ &\qquad \left. (290-321) + 1.9600 \times \sqrt{\frac{12^2}{81} + \frac{14^2}{121}}\right] \\ &= [-34.6128,\ -27.3872]. \end{aligned}$$

(b) as variâncias são desconhecidas E IGUAIS.

Neste caso podemos obter IC exacto, bastando para o efeito usarmos a v.a. fulcral para $(\mu_1 - \mu_2)$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}.$$

•

7.6 Intervalo de confiança para a variância de uma população normal.

Motivação 7.23 — **IC para a variância de população normal**
Até ao momento só adiantámos IC para valores esperados ou diferenças de valores esperados. É altura de deduzir uma estimativa intervalar para a variância, parâmetro este usualmente desconhecido a par do valor esperado da população. Consideraremos a situação em que estamos a lidar com uma v.a. de interesse normalmente distribuída com valor esperado desconhecido.[17] •

Considere-se então uma população $X \sim normal(\mu, \sigma^2)$, onde quer μ, quer σ^2 são desconhecidos.

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ desconhecido

  σ^2 DESCONHECIDO.

- **Passo 1 — Selecção da v.a. fulcral para σ^2**

  De notar que a v.a. fulcral para σ^2 se encontra no formulário e depende de σ^2 (mas não de μ) e, naturalmente, de um estimador de σ^2, neste caso S^2. Adiante-se ainda que esta v.a. possui distribuição exacta com parâmetros conhecidos, logo independente de σ^2. Trata-se de:

  $$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}. \tag{7.39}$$

  (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Pelo facto da f.d.p. da v.a. do qui-quadrado não ser simétrica em relação à origem e possuir suporte igual a $\mathbb{R}^+$, lidamos efectivamente com dois quantis não só distintos como não simétricos:

[17] Ao considerar-se o valor esperado conhecido, o procedimento descrito a seguir sofre ligeiríssimas alterações, em particular, o número de graus de liberdade da distribuição do qui-quadrado passa a ser igual a n.

$$(a_\alpha, b_\alpha): \quad \begin{cases} P(Z < a_\alpha) = \alpha/2 \\ P(Z > b_\alpha) = \alpha/2 \end{cases}$$

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2). \end{cases} \tag{7.40}$$

(Esquema...)

Escusado será dizer que este par de quantis se encontra tabelado. (Consulta da tabela...)

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$P\left[a_\alpha \leq \tfrac{(n-1)S^2}{\sigma^2} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[\tfrac{1}{b_\alpha} \leq \tfrac{\sigma^2}{(n-1)S^2} \leq \tfrac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\tfrac{(n-1)S^2}{b_\alpha} \leq \sigma^2 \leq \tfrac{(n-1)S^2}{a_\alpha}\right] = 1-\alpha.$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha)\times 100\%}(\sigma^2) = \left[\frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}, \; \frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right]. \tag{7.41}$$

Exemplo 7.24 — IC para a variância de população normal

Retome o Exemplo 7.20 e obtenha um intervalo de confiança a 90% para a variância do tempo até à primeira reparação de computadores da marca 1.[18]

- **V.a.**

  X = tempo até 1a. reparação de computador pessoal da marca 1

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ desconhecido

  σ^2 DESCONHECIDO.

[18] Adaptado do Exame de 13 de Julho de 2002.

- **Passo 1 — Selecção da v.a. fulcral para σ^2**

  Por pretendermos um IC para a variância de população normal faremos uso de:

  $$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

  (Ver formulário.)

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Ao ter-se em consideração que $n = n_1 = 20$ e $(1-\alpha) \times 100\% = 90\%$ far-se-á uso dos quantis não simétricos:

  $$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(19)}}(0.05) \stackrel{tabela}{=} 10.12 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(19)}}(0.95) \stackrel{tabela}{=} 30.14. \end{cases}$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

  Omite-se este passo...

- **Passo 4 — Concretização**

  Ora, como a expressão geral do IC para σ^2 é

  $$IC_{(1-\alpha)\times 100\%}(\sigma^2) = \left[\frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)},\ \frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right],$$

  onde, para este exemplo, se tem

  $$\begin{aligned} & n = n_1 = 20 \\ & s^2 = \tfrac{1}{n-1} \times [(\textstyle\sum_{i=1}^n x_i^2) - n(\bar{x})^2] = \tfrac{1}{20-1}(3560 - 20 \times 12.4^2) = 25.516 \\ & F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = 10.12 \\ & F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = 30.14, \end{aligned}$$

  conclui-se que

  $$\begin{aligned} IC_{90\%}(\sigma^2) &= \left[\frac{(20-1)\times 25.516}{30.14},\ \frac{(20-1)\times 25.516}{10.12}\right] \\ &= [16.085,\ 47.905]. \end{aligned}$$ •

Nota 7.25 — IC para a variância de população arbitrária (não normal)
Ao deixarmos de assumir que a população tem distribuição normal, nada se pode adiantar no que diz respeito à estimação intervalar da variância no âmbito desta disciplina. •

7.7 Intervalos de confiança para uma probabilidade de sucesso e outros parâmetros de populações não normais uniparamétricas

Motivação 7.26 — IC aproximado para uma probabilidade
É também altura de fornecer uma estimativa intervalar para a probabilidade de sucesso de uma prova de *Bernoulli*. •

Consideremos população $X \sim \text{Bernoulli}(p)$, onde a probabilidade de sucesso p é desconhecida. No âmbito desta disciplina consideraremos somente o caso em que dispomos de um grande número de observações (binárias...).

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  p DESCONHECIDO

  n suficientemente grande ($n > 30$).

- **Passo 1 — Selecção da v.a. fulcral para** p

  Neste caso devemos recorrer a:

  $$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} \text{normal}(0,1). \tag{7.42}$$

  Esta v.a. fulcral possui distribuição que decorre do TLC e do Teorema de Slustky.

  Importa notar que Z depende de p e de um estimador deste mesmo parâmetro, $\bar{X}$, a proporção de sucessos na a.a. Note-se ainda que $E(\bar{X}) = E(X) = p$ e $V(\bar{X}) = \frac{V(X)}{n} = \frac{p(1-p)}{n}$. Deste modo, ao aplicar-se o TLC, obteríamos a seguinte v.a. fulcral para p:

  $$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{a}{\sim} \text{normal}(0,1). \tag{7.43}$$

  Contudo, a utilização desta v.a. fulcral torna a inversão da desigualdade do Passo 3 muito difícil uma vez que p figura em numerador e em denominador (neste último caso sob o sinal de uma raíz).

  A v.a. fulcral para p, $Z = \frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}}$, é de longe muito mais conveniente na referida inversão. A respectiva distribuição aproximada decorre da aplicação do TLC e do Teorema de Slutsky.

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi^{-1}(1-\alpha/2). \end{cases} \tag{7.44}$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) \simeq 1-\alpha$$

$$P\left[a_\alpha \leq \frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq b_\alpha\right] \simeq 1-\alpha$$
$$P\left[\bar{X} - b_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} - a_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha$$
$$P\left[\bar{X} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha.$$

- **Passo 4 — Concretização**

O IC seguinte possui grau de confiança aproximadamente igual a $(1-\alpha)\times 100\%$:

$$IC(p) = \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}},\ \ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]. \tag{7.45}$$

Exemplo/Exercício 7.27 — IC aproximado para uma probabilidade

Um inquérito recente feito a 1000 habitantes de uma região rural revelou que 448 apoiam a aplicação de penas de prisão pesadas em crimes de fogo posto.

(a) Construa um intervalo de confiança aproximado a 95% para a probabilidade (p) de uma pessoa escolhida ao acaso na referida região ser favorável à aplicação da referida pena.[19]

- **V.a.**

  $X_i =$ resposta da $i-$ésima ao inquérito, $i = 1, \ldots, n$

  $X_i \overset{i.i.d.}{\sim} X$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  p DESCONHECIDO

  $n = 1000 >> 30$ (suficientemente grande).

[19] Adaptado do Exame de 6 de Setembro de 2006.

- **Passo 1 — Selecção da v.a. fulcral para p**

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} \text{normal}(0,1).$$

uma vez que nos foi solicitada a determinação de um IC aproximado para uma probabilidade e a dimensão da amostra justifica a aplicação do TLC e do Teorema de Slutsky.

- **Passo 2 — Obtenção dos quantis de probabilidade**

Os quantis a utilizar são

$$\begin{cases} a_\alpha = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.975) \stackrel{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

Estes enquadram a v.a. fulcral para p com probabilidade aproximadamente igual a $(1-\alpha)$.

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

Omite-se novamente este passo...

- **Passo 4 — Concretização**

Ao ter-se em consideração que

$$n = 1000$$
$$\bar{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i = \tfrac{448}{1000} = \text{proporção observada de respostas afirmativas}$$
$$\Phi^{-1}(1-\alpha/2) = 1.9600$$

conclui-se que o IC aproximado a 95% para p é dado por

$$\begin{aligned} IC(p) &= \left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}},\right. \\ &\qquad \left.\bar{x} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right] \\ &= \left[0.448 - 1.9600 \times \sqrt{\frac{0.448 \times (1-0.448)}{1000}},\right. \\ &\qquad \left.0.448 + 1.9600 \times \sqrt{\frac{0.448 \times (1-0.448)}{1000}}\right] \\ &= [0.417178,\ 0.478822]. \end{aligned}$$

(b) Obtenha o IC aproximado para p

$$\frac{2n\bar{x} + [\Phi^{-1}(1-\alpha/2)]^2 \pm \Phi^{-1}(1-\alpha/2)\sqrt{4n\bar{x} + [\Phi^{-1}(1-\alpha/2)]^2 - 4n\bar{x}^2}}{2\left\{n + [\Phi^{-1}(1-\alpha/2)]^2\right\}}$$

que se obtém ao recorrer-se à v.a. fulcral $\frac{\bar{X}-p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim}$ normal$(0,1)$, de acordo com Murteira (1990, Vol. II, p. 233). •

É curioso notar que, recorrendo a um procedimento análogo ao acabado de descrever, também se pode obter um IC aproximado para um outro valor esperado, o da v.a. de Poisson, como ilustra o exemplo seguinte.

Exemplo 7.28 — IC aproximado para o valor esperado de população de Poisson

Na famosa experiência realizada pelo físico Rutherford (em 1910), sobre a emissão de partículas α por uma fonte radioactiva, foi observado que em 2612 intervalos de 1/8 de minuto foram emitidas no total 9318 partículas. Admita que a variável aleatória X, que representa o número de partículas emitidas por essa fonte radioactiva num intervalo de 1/8 de minuto, possui distribuição de Poisson com parâmetro μ.

Obtenha um intervalo de confiança aproximado a 90% para o valor esperado de X.[20]

- **V.a.**

  $X =$ número de partículas emitidas em intervalo de 1/8 de minuto

- **Situação**

  $X \sim$ Poisson(μ)

  $\mu = E(X) = V(X)$ DESCONHECIDO

  $n >> 30$ (suficientemente grande).

- **Passo 1 — Selecção da v.a. fulcral para μ**

  Ao recorrermos ao TLC podemos afirmar que

$$\frac{\bar{X}-E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X}-E(X)}{\sqrt{\frac{V(X)}{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{\mu}{n}}} \stackrel{a}{\sim} \text{normal}(0,1).$$

[20] Adaptado do Exame de 8 de Julho de 2006.

No entanto, esta v.a. fulcral não só não consta do formulário, como torna a inversão da desigualdade do Passo 3 muito difícil uma vez que μ figura em numerador e em denominador (neste último caso sob o sinal de uma raíz).

Posto isto invocaremos o TLC e do Teorema de Slutsky que garantem que

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\bar{X}}{n}}} \stackrel{a}{\sim} \text{normal}(0, 1),$$

onde $\bar{X}$ é o estimador de MV de μ. Escusado será dizer que esta v.a. fulcral para μ é mais conveniente na referida inversão.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Usaremos os quantis

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.95) \stackrel{tabela}{=} -1.6449 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.95) = 1.6449 \end{cases}$$

  pois enquadram a v.a. fulcral para μ com probabilidade aproximadamente igual a 0.90.

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omite-se este passo...

- **Passo 4 — Concretização**

  Uma vez que

$$\begin{aligned} & n = 2612 \\ & \bar{x} = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} x_i = \tfrac{9318}{2612} = 3.5674 \\ & \Phi^{-1}(1 - \alpha/2) = 1.6449 \end{aligned}$$

  segue-se o IC aproximado a 90% para μ é

$$\begin{aligned} IC(\mu) &= \left[\bar{x} - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\bar{x}}{n}},\ \ \bar{x} + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\bar{x}}{n}}\right] \\ &= \left[3.5674 - 1.6449 \times \sqrt{\frac{3.5674}{2612}},\ \ 3.5674 + 1.6449 \times \sqrt{\frac{3.5674}{2612}}\right] \\ &= [3.5066,\ 3.6282]. \end{aligned}$$

•

A seguir apresenta-se um quadro que nos parece extremamente útil dado condensar todos os intervalos de confiança que constam deste capítulo. Este quadro possui cinco colunas onde podemos encontrar:

- o parâmetro para o qual se destina o IC;
- em que circunstâncias podemos recorrer a semelhante IC;
- uma estimativa pontual do referido parâmetro;
- a v.a. fulcral a usar nesta situação;
- e, por fim, o IC, que poderá ser exacto ou aproximado.[21]

[21] Esta destrinça é deixada à atenção do leitor/a mas basta relembrar que os intervalos aproximados estão associados à utilização de uma v.a. fulcral com distribuição aproximada e, como tal, após a expressão da v.a. segue-se o sinal $\stackrel{a}{\sim}$.

Quadro-resumo 1: Alguns intervalos de confiança.

Parâmetro	**Situação**	**Estimativa pontual**	**V.a. fulcral**	**Intervalo de confiança**
μ	$X \sim \text{normal}(\mu, \sigma^2)$ σ^2 conhecido	$\bar{x}$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{normal}(0,1)$	$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}},\ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right]$
μ	X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$ σ^2 conhecido n suficientemente grande	$\bar{x}$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} \text{normal}(0,1)$	$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}},\ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{\sigma}{\sqrt{n}}\right]$
$(\mu_1 - \mu_2)$	$X_1 \sim \text{normal}(\mu_1, \sigma_1^2)$ $\perp\!\!\!\perp$ $X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$ σ_1^2, σ_2^2 conhecidos	$(\bar{x}_1 - \bar{x}_2)$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \text{normal}(0,1)$	$\left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$
$(\mu_1 - \mu_2)$	$X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (não normais): $E(X_i) = \mu_i,\ V(X_i) = \sigma_i^2, i = 1,2$ σ_1^2, σ_2^2 conhecidos n_1, n_2 suficientemente grandes	$(\bar{x}_1 - \bar{x}_2)$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \overset{a}{\sim} \text{normal}(0,1)$	$\left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$
μ	$X \sim \text{normal}(\mu, \sigma^2)$ σ^2 desconhecido	$\bar{x}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$	$\left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}},\ \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right]$
μ	X com distribuição arbitrária: $E(X) = \mu, V(X) = \sigma^2$ σ^2 desconhecido n suficientemente grande	$\bar{x}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{a}{\sim} \text{normal}(0,1)$	$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}},\ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right]$
$(\mu_1 - \mu_2)$	$X_1 \sim \text{normal}(\mu_1, \sigma_1^2)$ $\perp\!\!\!\perp$ $X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$ σ_1^2, σ_2^2 desconhecidos mas IGUAIS $n_1 \leq 30$ ou $n_2 \leq 30$	$(\bar{x}_1 - \bar{x}_2)$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$	$\left[(\bar{x}_1 - \bar{x}_2) - F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) \times \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},\ (\bar{x}_1 - \bar{x}_2) + F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) \times \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$
$(\mu_1 - \mu_2)$	$X_1 \perp\!\!\!\perp X_2$ com distribuições arbitrárias (possivelmente normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1,2$ σ_1^2, σ_2^2 desconhecidos mas não necessariamente iguais $n_1 > 30$ e $n_2 > 30$	$(\bar{x}_1 - \bar{x}_2)$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \overset{a}{\sim} \text{normal}(0,1)$	$\left[(\bar{x}_1 - \bar{x}_2) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$
σ^2	$X \sim \text{normal}(\mu, \sigma^2)$ μ desconhecido	s^2	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$	$\left[\frac{(n-1)S^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)},\ \frac{(n-1)S^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right]$
p	$X \sim \text{Bernoulli}(p)$ n suficientemente grande ($n > 30$)	$\bar{x}$	$\frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} \text{normal}(0,1)$	$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}},\ \bar{x} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$

Capítulo 8

Testes de hipóteses

Motivação 8.1 — Testes de hipóteses
Até ao momento procurámos adiantar um valor razoável ou um intervalo de valores razoáveis para um parâmetro desconhecido de interesse, tirando partido da informação amostral de que dispomos.

É altura de tirarmos partido dessa mesma informação para nos pronunciarmos sobre afirmações relativas a esse mesmo parâmetro desconhecido ou a outros aspectos da nossa v.a. de interesse.

Não se trata de um mero exercício académico mas sim de uma resposta a uma necessidade decorrente da prática. Basta pensarmos na averiguação da veracidade/razoabilidade de uma afirmação de:

- um fabricante de automóveis quanto ao valor esperado da quantidade de combustível consumida por cada $100Km$ por certo modelo fabricado;

- uma produtora de um fármaco no que diz respeito à percentagem de pacientes que consideram ter havido melhoras do seu estado de saúde após a administração do fármaco. •

A exposição da matéria deste capítulo muito se assemelha à do capítulo anterior: ao invés de falarmos de intervalos de confiança para parâmetros de interesse, apresentamos testes de hipóteses para esses mesmos parâmetros. Mais, a ordem de apresentação dos testes de hipóteses respeita a do capítulo anterior.

Para além de testes sobre parâmetros, será apresentado o teste de ajustamento do qui-quadrado, bem como o teste de independência do qui-quadrado, pese embora o facto de este último teste nem sempre figurar no programa da disciplina.

8.1 Noções básicas

Antes de prosseguirmos é crucial adiantarmos algumas noções fundamentais no âmbito dos testes de hipóteses, muito em particular o que se entende por hipótese estatística.

Definição informal 8.2 — Hipótese estatística
Qualquer afirmação/conjectura acerca de

- um parâmetro desconhecido ou
- da distribuição da v.a. de interesse, etc.

será denominada de hipótese estatística. •

Definição 8.3 — Hipótese paramétrica
Trata-se de conjectura respeitante a um parâmetro desconhecido, assumindo que se conhece a distribuição da v.a. de interesse (a menos de um ou mais parâmetros desconhecidos). •

Exemplo 8.4 — Hipótese paramétrica
É usual que os fabricantes façam afirmações acerca dos seus produtos. Por exemplo, um fabricante de baterias de 12V, para além de assumir que a duração das mesmas (v.a. de interesse) possui distribuição normal com variância conhecida, pretende convencer um seu grande cliente que o valor esperado da duração das mesmas é de 1000 dias:

- **V.a. de interesse**

  $X =$ duração (em dias) de bateria de 12V

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  $\mu = E(X)$ DESCONHECIDO

  $\sigma^2 = V(X)$ conhecida

- **Hipótese**

  $\mu = 1000$ dias •

É essencialmente com este tipo de hipóteses que iremos lidar ao longo deste capítulo, pelo que omitiremos o adjectivo "paramétrica" sempre que nos parecer desnecessário.

Ao lidar-se com hipóteses paramétricas, assumir-se-á que o conjunto de valores considerados possíveis/razoáveis para o parâmetro desconhecido — i.e., o espaço de parâmetro Θ — possui pelo menos dois valores.

Definição informal 8.5 — Hipótese nula/alternativa
De um modo geral confrontamos duas hipóteses paramétricas:

- a hipótese mais relevante, usualmente designada de hipótese nula e representada por H_0;
- uma hipótese dita alternativa, representada por H_1.

Estas duas hipóteses paramétricas estão associadas a dois sub-espaços disjuntos de Θ: H_0 atribui ao parâmetro valores em $\Theta_0 \subset \Theta$ e H_1 em $\Theta_1 = \Theta \backslash \Theta_0$. •

Exemplo 8.6 — Hipótese nula/alternativa
Há várias possibilidades de hipótese alternativa para a hipótese nula do Exemplo 8.4:

- $H_0 : \mu = \mu_0 = 1000$ dias (hipótese nula)
- $H_1 := \begin{cases} \mu = \mu_1 = 800 \text{ dias} & \text{(ponto de vista do consumidor)} \\ \mu < \mu_0 = 1000 \text{ dias} & \text{(ponto de vista do consumidor)} \\ \mu > \mu_0 = 1000 \text{ dias} & \text{(ponto de vista do fabricante)} \\ \mu \neq \mu_0 = 1000 \text{ dias} & \text{(compromisso entre ambos)} \end{cases}$

Se assumirmos que a variância da duração das baterias de 12V é conhecida estaremos a lidar com os seguintes espaços e sub-espaços de parâmetro:

Espaço de parâmetro	**Hip. Nula**	Θ_0	**Hip. Alternativa**	$\Theta_1 = \Theta\backslash\Theta_0$
$\Theta = \{800, 1000\}$	$H_0 : \mu = 1000$	$\Theta_0 = \{1000\}$	$H_1 : \mu = 800$	$\Theta_1 = \{800\}$
$\Theta = (0, 1000]$	$H_0 : \mu = 1000$	$\Theta_0 = \{1000\}$	$H_1 : \mu < 1000$	$\Theta_1 = (0, 1000)$
$\Theta = [1000, +\infty)$	$H_0 : \mu = 1000$	$\Theta_0 = \{1000\}$	$H_1 : \mu > 1000$	$\Theta_1 = (1000, +\infty)$
$\Theta = (0, +\infty)$	$H_0 : \mu = 1000$	$\Theta_0 = \{1000\}$	$H_1 : \mu \neq 1000$	$\Theta_1 = (0, +\infty)\backslash\{1000\}$

Convém ainda referir que os valores especificados nas hipóteses nula e alternativa nada têm a ver com valores observados na amostra, muito em particular com a estimativa (de MV) do parâmetro desconhecido μ, $\bar{x}$. •

Há várias possibilidades de classificar as hipóteses nula e alternativa, nomeadamente no que diz respeito à cardinalidade do sub-espaço de parâmetro a elas associado...

Definição 8.7 — Hipótese simples/composta
Uma hipótese paramétrica H_i $(i = 0, 1)$ diz-se:

- simples, caso especifique um único valor para o parâmetro desconhecido que corresponde a um único ponto no espaço paramétrico, i.e., caso $\#\Theta_i = 1$ $(i = 0, 1)$;

- composta, caso contrário. •

Exemplo 8.8 — Hipótese simples/composta
Retome-se o Exemplo 8.4. Ao assumir-se que a v.a. de interesse possui distribuição normal com valor esperado desconhecido e variância conhecida tem-se:

- $H_0 : \mu = \mu_0 = 1000$ dias — hipótese simples $(\Theta_0 = \{1000\})$;

- $H_1 : \mu < \mu_0 = 1000$ dias — hipótese composta $(\Theta_1 = (0, 1000))$.

Note-se, no entanto, que, caso se assuma que a variância é igualmente desconhecida, não só passamos a lidar com um espaço de parâmetro ligeiramente diferente, como H_0 passa a ser também uma hipótese composta. Com efeito, $\Theta = \{(\mu, \sigma^2) : \mu, \sigma^2 > 0\} = I\!R^+ \times I\!R^+$ e a hipótese $H_0 : \mu = \mu_0 = 1000$ dias está associada ao sub-espaço $\Theta_0 = \{1000\} \times I\!R^+$, que por sinal não é um conjunto singular. •

Definição informal 8.9 — Hipótese alternativa unilateral/bilateral
É costume ir além na classificação da hipótese alternativa. Esta classificacão depende também da hipótese nula com que estejamos a lidar. H_1 é, em geral, de um dos três seguintes tipos:

- unilateral inferior, caso possua um sinal de menor (e.g., $H_0 : \mu = 1000$ dias e $H_1 : \mu < 1000$ dias);[1]

- unilateral superior, se possuir um sinal de maior (e.g., $H_0 : \mu = 1000$ dias e $H_1 : \mu > 1000$ dias);[2]

- bilateral, caso H_1 tenha um sinal de diferente (e.g., $H_0 : \mu = 1000$ dias e $H_1 : \mu \neq 1000$ dias). •

[1] Ou atribua um valor inferior ao conjecturado em H_0, e.g., $H_1 : \mu = 800$ dias.
[2] Ou atribua um valor superior ao conjecturado em H_0, e.g., $H_1 : \mu = 1200$ dias.

Definição informal 8.10 — Teste de hipóteses
Um teste de hipóteses não passa de um procedimento estatístico que conduz a uma decisão acerca das hipótese nula e alternativa, tendo em conta a informação amostral que recolhemos. •

Nota 8.11 — Decisões em testes de hipóteses
De acordo com a informação recolhida tomamos (de um modo geral) uma de duas decisões:

- rejeitar H_0;
- não rejeitar H_0. •

Definição informal 8.12 — Erros de 1a. e 2a. espécie
As decisões em testes de hipóteses podem ou não ser correctas. Senão vejamos:

	Situação real	
Decisão	H_0 Verdadeira	H_0 Falsa
Rejeitar H_0	**Erro de 1a. Espécie**	Decisão Correcta
Não rejeitar H_0	Decisão Correcta	**Erro de 2a. Espécie**

•

Definição 8.13 — Probabilidade de erro de 1a. e de 2a. espécie
É habitual delinear o teste de hipóteses de modo a minimizar as probabilidades de ocorrência de erros de 1a. e 2a. espécie. Estas probabilidades costumam ser representadas por α e β, respectivamente, e definem-se do seguinte modo:

$$\alpha = P(\text{Erro de 1a. espécie}) = P(\text{Rejeitar } H_0 \mid H_0 \text{ é Verdadeira}) \tag{8.1}$$

$$\beta = P(\text{Erro de 2a. espécie}) = P(\text{Não rejeitar } H_0 \mid H_0 \text{ é Falsa}). \tag{8.2}$$

•

Qualquer destas probabilidades dependerá, como veremos mais tarde, do verdadeiro valor do parâmetro.

Definição 8.14 — Nível de significância
É habitual estabelecer um limite superior para a probabilidade de ocorrência de erro de 1a. espécie. Este limite é usualmente designado de nível de significância (n.s.) do teste e representado doravante por α_0 ($\alpha_0 \in (0,1)$). O teste de hipóteses será delineado de modo que

- $P(\text{Rejeitar } H_0 \mid H_0 \text{ é Verdadeira}) \leq \alpha_0$.

Os valores mais comuns do n.s. do teste são 10%, 5% e 1%. •

Qualquer decisão pela rejeição ou não de H_0 deverá basear-se na informação recolhida, muito em particular no

- valor observado daquilo que chamaremos de uma estatística de teste e designaremos por T.

Definição informal 8.15 — Estatística de teste
Uma estatística de teste — a utilizar no confronto de um par de hipóteses que digam respeito ao parâmetro desconhecido θ — possui as seguintes características:

- reflecte a discrepância entre o estimador de θ e o valor conjecturado para θ em H_0 (θ_0);

- obtém-se, de um modo geral, à custa da v.a. fulcral Z — que usaríamos na construção de um intervalo de confiança para θ —, substituindo θ por θ_0 na expressão de Z;

- tem distribuição (exacta ou aproximada) conhecida, sob a validade de H_0. •

Exemplo 8.16 — Estatística de teste
Retomemos o Exemplo 8.4 e consideremos:

- **V.a. de interesse**

  X = duração de bateria de 12V

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$
  $\mu = E(X)$ DESCONHECIDO
  $\sigma^2 = V(X)$ conhecida

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 1000$ dias vs. $H_1 : \mu \neq \mu_0$

- **Nível de significância**

  $\alpha_0 = 5\%$

Note-se que, caso pretendêssemos construir um IC para μ, deveríamos utilizar a seguinte

- **V.a. fulcral para μ**

  $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{normal}(0,1)$.

E é substituindo μ por μ_0 na expressão de Z que obtemos a...

- **Estatística de teste**

  $T = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0,1)$,

  onde "$\sim_{H_0} \ldots$"deve ler-se "tem distribuição ... sob a validade de H_0". •

Decidir pela rejeição ou não de H_0 pressupõe também que se identifique:

- o conjunto de valores considerados críticos para a estatística de teste T que deverão conduzir à rejeição de H_0.

Este conjunto é, por isso mesmo, denominado de região crítica ou de rejeição de H_0 (para valores da estatística de teste).

Definição informal 8.17 — Região de rejeição de H_0 (para valores da estatística de teste)

Representamo-la por W e escolhemo-la de modo que:

- $P(\text{Rejeitar } H_0 \mid H_0 \text{ é Verdadeira}) = \alpha_0 \ \ (\leq \alpha_0)$.

- seja um intervalo real (ou uma reunião de intervalos reais) que depende de quantil(is) de probabilidade relacionada com α_0 e respeitantes à distribuição (exacta ou aproximada) da estatística de teste sob H_0;

- o seu aspecto dependa da hipótese alternativa H_1 e do que "significa"obter valores inconsistentes com H_0.

Exemplo 8.18 — Região de rejeição de H_0 (para valores da estatística de teste)

No Exemplo 8.16 concluímos rapidamente que quanto mais se distingue a estimativa de μ, $\bar{x}$, do valor conjecturado para μ, $\mu_0 = 1000$ dias, menos consistente é $H_0 : \mu = \mu_0 = 1000$ dias com os dados e mais consistente é $H_1 : \mu \neq \mu_0$ com os mesmos.

Efectivamente, quando, por exemplo, $\bar{x} = 750$ dias ($\bar{x} = 1300$ dias, resp.) devemos, em princípio, rejeitar H_0. Ora, nestes casos temos $\bar{x} << \mu_0$ ($\bar{x} >> \mu_0$, resp.). Equivalentemente,

- $t = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} << 0 \qquad (t = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} >> 0$, resp.).

Assim, a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos — um à esquerda ($t << 0$) e outro à direita ($t >> 0$) — ou seja

- $W = (-\infty, -c) \cup (c, +\infty)$,

onde o valor crítico c é escolhido de modo que

- $P(\text{Rejeitar } H_0 \mid H_0 \text{ é Verdadeira}) = \alpha_0 \Leftrightarrow P(T < -c \text{ ou } T > c \mid H_0) = \alpha_0$.

Assim, à semelhança do que acontecia com os intervalos de confiança lidaremos com o seguinte quantil de probabilidade, para $\alpha_0 = 5\%$:

- $c = \Phi^{-1}(1 - \alpha_0/2) = \Phi^{-1}(1 - 0.05/2) \stackrel{tabela}{=} 1.9600$. •

Uma vez escolhida a estatística de teste e identificada a região de rejeição de H_0 para a mesma, estamos em condições de poder tomar uma decisão.

Proposição 8.19 — Decisão
Para decidir pela rejeição ou não de H_0 é necessário calcular

- $t =$ valor observado da estatística de teste.

Deve tomar-se uma de duas decisões:

- Rejeitar H_0 ao n.s. α_0, se $t \in W$;
- Não rejeitar H_0 ao n.s. α_0, se $t \notin W$. •

Nota 8.20 — Decisão

- Afirmar que
  - H_0 não foi rejeitada ao n.s. α_0

  não significa que
  - H_0 seja verdadeira.
- Analogamente, concluir que
  - H_0 foi rejeitada ao n.s. α_0

  não significa que

– H_0 seja falsa.

Significa sim que

– H_0 não é consistente com os dados ao n.s. α_0.

- Podemos rejeitar H_0 ao n.s. α_0 e não rejeitar esta hipótese a outro n.s. e vice-versa.

  Em suma, a decisão de rejeitar ou não H_0 depende crucialmente do nível de significância considerado. •

Apresentadas que foram as noções fundamentais em testes de hipóteses é altura de apresentar o seu procedimento geral.

Definição informal 8.21 — Procedimento geral de testes de hipóteses
A efectuação de um teste de hipóteses compreende os 7 passos seguintes:

1. **V.a. de interesse**

   Identificar a v.a. de interesse (X).

2. **Situação**

   Identificar a distribuição da v.a. de interesse, o parâmetro DESCONHECIDO que está a ser testado, bem como averiguar se a distribuição de X depende de mais parâmetros e se estes são conhecidos ou desconhecidos.

3. **Hipóteses**

   Especificar as hipóteses nula (H_0) e alternativa (H_1).

4. **Nível de significância**

   Escolher o nível de significância (α_0).

5. **Estatística de teste**

   Escolher a estatística de teste adequada (T) e identificar a sua distribuição (exacta ou aproximada) sob a validade de H_0.

6. **Região de rejeição de** H_0 (para valores da estatística de teste)

   Obter a região de rejeição de H_0 para valores da estatística de teste (W), tendo em conta o nível de significância e a hipótese alternativa.

7. **Decisão**

   Calcular o valor observado da estatística de teste (t) e decidir pela rejeição ou não de H_0 ao n.s. α_0. •

As secções que se seguem consistem essencialmente na apresentação de um ou dois exemplos ilustrativos do teste que dá por título a secção. Note-se, no entanto, que, de um modo geral, em qualquer das secções será introduzido um conceito novo no âmbito dos testes de hipóteses pelo que se recomenda a leitura integral de todas elas.

8.2 Testes de hipóteses para o valor esperado, variância conhecida.

Tal como na obtenção de IC para o valor esperado de uma população com variância conhecida, distinguiremos dois casos:

1. População normal

2. População com distribuição arbitrária e dimensão da amostra suficientemente grande.

Ilustraremos a efectuação destes dois testes com exemplos.

Exemplo 8.22 (CASO 1) — **Teste sobre valor esperado de população normal, variância conhecida**
Um fabricante de baterias de 12V — para além de assumir que a duração das mesmas (v.a. de interesse) possui distribuição normal com desvio-padrão conhecido e igual a 50 dias — pretende convencer um seu grande cliente que o valor esperado da duração das mesmas é de 1000 dias.

Averigue a razoabilidade desta afirmação do fabricante de baterias ao nível de significância de 5% e tendo em conta que a média da duração de 4 baterias testadas foi de 930 dias.

- **V.a. de interesse**
  $X =$ duração de bateria de 12V

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  $\sigma^2 = 50^2 \ dias^2$ conhecido

- **Hipóteses**
  $H_0 : \mu = \mu_0 = 1000$ dias vs. $H_1 : \mu \neq \mu_0 = 1000$ dias

- **Nível de significância**
  $\alpha_0 = 5\%$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0,1) \tag{8.3}$$

pois estamos a efectuar um teste sobre o valor esperado de uma população normal com variância conhecida.

- **Região de rejeição de H_0** (para valores da estatística de teste)

Por estarmos a lidar com teste bilateral, a região de rejeição de H_0, escrita para valores da estatística de teste, é uma reunião de intervalos do tipo

$$W = (-\infty, -c) \cup (c, +\infty), \tag{8.4}$$

onde $c : P(\text{Rejeitar } H_0|H_0) = \alpha_0$, i.e.,

$$c = \Phi^{-1}(1 - \alpha_0/2) = \Phi^{-1}(1 - 0.05/2) \stackrel{tabela}{=} 1.9600. \tag{8.5}$$

- **Decisão**

Uma vez que o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{930 - 1000}{50/\sqrt{4}} \\ &= -2.8 \\ &\in W = (-\infty, -1.9600) \cup (1.9600, +\infty), \end{aligned} \tag{8.6}$$

conclui-se que devemos rejeitar H_0 ao n.s. de 5%, pelo que o cliente tem razões para desconfiar do fabricante. •

Exemplo 8.23 (CASO 2) — **Teste sobre valor esperado de população com distribuição arbitrária e dimensão da amostra suficientemente grande, variância conhecida**

Ao estudar-se a densidade de construção (X) num projecto de urbanização foi recolhida uma amostra de 50 lotes desse projecto, tendo-se obtido $\sum_{i=1}^{50} x_i = 227.2$.[3]

Até que ponto é razoável afirmar que valor esperado da densidade de construção é igual a 5, assumindo que o desvio-padrão de X é igual a 4 e um n.s. de 10% ?

- **V.a. de interesse**

  X = densidade de construção em projecto de urbanização

- **Situação**

  X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$

  μ DESCONHECIDO

  $\sigma^2 = 4^2$ conhecido

  n suficientemente grande ($n = 50 > 30$)

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 5$ vs. $H_1 : \mu \neq \mu_0$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **V.a. fulcral para μ e estatística de teste**

  De acordo com o TLC, pode afirmar-se que:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} \text{normal}(0,1).$$

  Substituindo μ por μ_0 na expressão desta v.a. fulcral, obtemos a seguinte estatística de teste com distribuição aproximada conhecida sob H_0:

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{a}{\sim}_{H_0} \text{normal}(0,1). \tag{8.7}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  O teste é bilateral, logo a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) \simeq \alpha_0$, i.e.,

[3] Adaptado do Exame de 19 de Janeiro de 2002.

$$c = \Phi^{-1}(1-\alpha_0/2) = \Phi^{-1}(1-0.10/2) \stackrel{tabela}{=} 1.6449. \tag{8.8}$$

- **Decisão**

  Como o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \\ &= \frac{4.544-5}{4/\sqrt{50}} \\ &= -0.81 \\ &\notin W = (-\infty, -1.6449) \cup (1.6449, +\infty), \end{aligned} \tag{8.9}$$

  deve afirmar-se que não devemos rejeitar H_0 ao n.s. de 10%. •

8.3 Testes de hipóteses sobre a igualdade de dois valores esperados, variâncias conhecidas.

Motivação 8.24 — Testes de hipóteses sobre a igualdade de dois valores esperados, variâncias conhecidas
Ao confrontar duas populações independentes — representando, por exemplo, os desempenhos de dois tipos de artigos (resistência, duração, etc.) — é usual testar a igualdade dos seus valores esperados, sejam eles μ_1 e μ_2. E importa notar que a hipótese de igualdade de valores esperados,

$$H_0' : \mu_1 = \mu_2 \tag{8.10}$$

é equivalente a

$$H_0 : \mu_1 - \mu_2 = \mu_0 = 0. \tag{8.11}$$

•

Distinguiremos novamente dois casos:

1. Duas populações independentes, com distribuições normais com variâncias conhecidas
2. Duas populações independentes, com distribuições arbitrárias com variâncias conhecidas, e dimensões das duas amostras suficientemente grandes.

Exemplo 8.25 (CASO 1) — **Teste sobre a igualdade de valores esperados de duas populações normais, variâncias conhecidas**
Um fabricante produz dois tipos de plástico e pretende saber se as suas resistências possuem valores esperados iguais. Com objectivo de esclarecer esta dúvida, recolheu duas amostras, tendo registado as seguintes observações:

- Tipo 1 — $\underline{x}_1 = (26, 24, 22, 30)$
- Tipo 2 — $\underline{x}_2 = (25, 31, 33, 29)$.

Assumindo que as resistências de Tipo 1 e 2 são v.a. independentes com distribuição normal e desvios-padrão iguais a $\sigma_1 = 7$ e $\sigma_2 = 3$, respectivamente, teste

- a hipótese de igualdade dos valores esperados das resistências

contra

- a hipótese do valor esperado da resistência do plástico de Tipo 1 ser inferior à do Tipo 2.

Considere para o efeito o n.s. de 5%.

- **V.a. de interesse**

  $X_1 =$ resistência do plástico de Tipo 1

  $X_2 =$ resistência do plástico de Tipo 2

- **Situação**

  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  $\sigma_1^2 = 7^2$ e $\sigma_2^2 = 3^2$ (conhecidos)

- **Hipóteses**

  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0 \ (\mu_1 = \mu_2)$ vs. $H_1 : \mu_1 - \mu_2 < \mu_0 = 0 \ (\mu_1 < \mu_2)$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim_{H_0} \text{normal}(0,1) \tag{8.12}$$

  uma vez que pretendemos efectuar teste sobre a igualdade de valores esperados de duas populações independentes com distribuição normal e variâncias conhecidas.

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Tratando-se de

  – um teste unilateral inferior ($H_1 : \mu_1 - \mu_2 < \mu_0 = 0$),

  concluímos que, quanto menor for a estimativa de MV de $(\mu_1 - \mu_2)$, $(\bar{x}_1 - \bar{x}_2)$, mais razões temos para rejeitar H_0. Assim, a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à esquerda

$$W = (-\infty, c), \tag{8.13}$$

  onde

$$c = \Phi^{-1}(\alpha_0) = -\Phi^{-1}(1 - \alpha_0) = -\Phi^{-1}(1 - 0.05) \stackrel{tabela}{=} -1.6449. \tag{8.14}$$

- **Decisão**

  Uma vez que $n_1 = n_2 = 4$ e as médias das duas amostras são iguais a $\bar{x}_1 = 25.5$ e $\bar{x}_2 = 29.5$, o valor observado da estatística de teste é dado por

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{(25.5 - 29.5) - 0}{\sqrt{\frac{7^2}{4} + \frac{3^2}{4}}} \\ &= -1.05 \\ &\notin W = (-\infty, -1.6449). \end{aligned} \tag{8.15}$$

  Deste modo, concluímos que não devemos rejeitar a hipótese de igualdade dos valores esperados das resistências dos dois tipos de plástico (H_0) ao n.s. de 5%. (Faça um comentário que entender pertinente...) •

Nota 8.26 — Decisão a diversos níveis de significância
É curioso notar que no Exemplo 9.4 não se rejeita H_0 a qualquer nível de significância menor ou igual a 5% pois

$$-1.05 \notin \begin{cases} W_{5\%} = (-\infty, -1.6449) \\ W_{2.5\%} = (-\infty, -1.9600) \\ W_{1\%} = (-\infty, -2.3263) \end{cases} \tag{8.16}$$

sugerindo que enunciemos os seguintes resultados. •

Proposição 8.27 — Decisão a diversos níveis de significância
É crucial ter presente os seguintes resultados:

- Não rejeitar H_0 ao n.s. $\alpha_0 \Rightarrow$ Não rejeitar H_0 a qualquer n.s. $\alpha_0' \leq \alpha_0$;

- Rejeitar H_0 ao n.s. $\alpha_0 \Rightarrow$ Rejeitar H_0 a qualquer n.s. $\alpha_0' \geq \alpha_0$.

(Justifique elaborando um esquema!) •

Exercício 8.28 (CASO 2) — **Teste sobre a igualdade de valores esperados de populações independentes com distribuição arbitrária e dimensão da amostra suficientemente grande, variâncias conhecidas**
Conceba um enunciado de um teste de hipóteses para o CASO 2 e complete os passos seguintes do procedimento geral de hipóteses:

- **V.a. de interesse**

- **Situação**

  $X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (não normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1, 2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 conhecidos

  n_1 e n_2 suficientemente grandes.

- **Hipóteses**

- **Nível de significância**

- **Estatística de teste**

$$\frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \overset{a}{\sim}_{H_0} \text{normal}(0, 1) \tag{8.17}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

- **Decisão**

•

8.4 Função potência de um teste

Motivação 8.29 — **Função potência de um teste**
Ao efectuar-se um teste é usual saber qual a probabilidade de rejeitar H_0 quando esta hipótese é verdadeira. É igualmente importante determinar a probabilidade de tomar a seguinte decisão acertada: rejeitar H_0 quando H_0 é falsa. •

Comecemos por relembrar que, ao definirmos as hipóteses nula (H_0) e alternativa (H_1) sobre o parâmetro desconhecido θ, subdividimos o espaço de parâmetro em dois sub-espaços Θ_0 e Θ_1, respectivamente. Para além disso, recorde-se que a probabilidade de cometer erros de 1a. e 2a. espécie são funções de θ definidas por

$$\begin{aligned} \alpha &= \alpha(\theta) \\ &= P(\text{Rejeitar } H_0 \mid H_0 \text{ é verdadeira}) \\ &= P(T \in W \mid \theta),\ \theta \in \Theta_0 \end{aligned} \tag{8.18}$$

$$\begin{aligned} \beta &= \beta(\theta) \\ &= P(\text{Não rejeitar } H_0 \mid H_0 \text{ é falsa}) \\ &= P(T \notin W \mid \theta),\ \theta \in \Theta_1. \end{aligned} \tag{8.19}$$

Definição 8.30 — **Função potência de um teste**
A função potência de um teste corresponde à probabilidade de rejeição da hipótese nula (quer esta seja verdadeira, quer seja falsa):

$$\begin{aligned} p(\theta) &= P(\text{Rejeitar } H_0 \mid \theta) \\ &= P(T \in W \mid \theta),\ \theta \in \Theta \\ &= \begin{cases} \alpha(\theta), & \theta \in \Theta_0 \\ 1 - \beta(\theta), & \theta \in \Theta_1. \end{cases} \end{aligned} \tag{8.20}$$

•

Caso H_0 seja uma hipótese nula simples, o sub-espaço Θ_0 é singular e temos

$$p(\theta) = \begin{cases} \alpha(\theta_0), & \theta \in \Theta_0 = \{\theta_0\} \\ 1 - \beta(\theta), & \theta \in \Theta_1. \end{cases} \tag{8.21}$$

Exemplo 8.31 — **Função potência de um teste**
Para definir a função potência do teste descrito no Exemplo 8.22, é necessário relembrar o seguinte:

- **V.a. de interesse**

  $X =$ duração de bateria de 12V

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  $\sigma^2 = 50^2$ dias conhecido

  $n = 4$

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 1000$ dias vs. $H_1 : \mu \neq \mu_0$ dias

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0, 1) \tag{8.22}$$

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  $W = (-\infty, -1.9600) \cup (1.9600, +\infty)$.

Assim sendo, segue-se:

- **Espaço e sub-espaços de parâmetro**

$$\Theta = (0, +\infty)$$
$$\Theta_0 = \{\mu_0\}$$
$$\Theta_1 = \Theta \backslash \{\mu_0\}$$

- **Função potência do teste**

$$\begin{aligned}
p(\mu) &= P(\text{Rejeitar } H_0 \mid \mu) \\
&= P(T \in W \mid \mu) \\
&= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < -1.9600 \text{ ou } \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > 1.9600 \mid \mu\right) \\
&= P\left(\frac{\bar{X}-\mu+\mu-\mu_0}{\sigma/\sqrt{n}} < -1.9600 \mid \mu\right) \\
&\quad +P\left(\frac{\bar{X}-\mu+\mu-\mu_0}{\sigma/\sqrt{n}} > 1.9600 \mid \mu\right) \\
&= P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < -1.9600 + \frac{\mu_0-\mu}{\sigma/\sqrt{n}} \mid \mu\right) \\
&\quad +P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > 1.9600 + \frac{\mu_0-\mu}{\sigma/\sqrt{n}} \mid \mu\right) \\
&= \Phi\left(-1.9600 + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right) + 1 - \Phi\left(1.9600 + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right), \ \mu \in \Theta. \quad (8.23)
\end{aligned}$$

- **Concretizações**

  Note-se, por exemplo, que:

  $\mu = \mu_0 = 1000$ dias $\rightarrow$
  $p(\mu_0) = \Phi(-1.9600) + 1 - \Phi(1.9600) = 2 \times [1 - \Phi(1.9600)] = 0.05 = \alpha_0$;

  $\mu = \mu_1 = 1025$ dias $\rightarrow$
  $p(\mu_1) = \Phi(-1.9600 - 1) + 1 - \Phi(1.9600 - 1) = [1 - \Phi(2.9600)] + [1 - \Phi(0.9600)] \stackrel{tabela}{=} (1 - 0.9985) + (1 - 0.8315) = 0.1700$.

  A interpretação do último destes resultados passa por afirmar que, caso o verdadeiro valor esperado da duração das baterias de $12V$ seja igual a $\mu = \mu_1 = 1025$ dias, o teste é capaz de rejeitar H_0 em 17% das vezes. •

Exercício 8.32 — **Função potência de um teste**

(a) Recorra a um *software* que lhe seja familiar para elaborar o gráfico da função potência do teste descrito no Exemplo 8.31. Comente.

(b) Elabore também o gráfico da função potência de teste quando se confrontam as hipóteses $H_0 : \mu = \mu_0$ e $H_1 : \mu \neq \mu_0$ mas se utiliza a região de rejeição $W = (-\infty, -1.96)$. Comente este gráfico e compare-o com o que obteve na alínea (a). •

8.5 Testes de hipóteses para o valor esperado, variância desconhecida.

É de longe mais realista efectuar um teste sobre μ assumindo que σ^2 é igualmente desconhecido. Mais, tal como na Secção 8.2, será necessário fazer a destrinça entre os dois casos seguintes:

1. População normal
2. População com distribuição arbitrária e dimensão da amostra suficientemente grande.

Exemplo 8.33 (CASO 1) — **Teste sobre o valor esperado de população normal, variância desconhecida**
A duração (em horas de uso) de certa marca de pilha para máquinas fotográficas digitais possui distribuição que se admite normal com valor esperado μ e variância σ^2 desconhecidos. Um revendedor dessas baterias adquiriu recentemente um grande lote e seleccionou uma amostra de 10 baterias, tendo obtido as seguintes durações $\underline{x} = (251, 238, 236, 229, 252, 253, 245, 242, 235, 230)$. De notar que a esta amostra estão associados $\sum_{i=1}^{10} x_i = 2411$ e $\sum_{i=1}^{10} x_i^2 = 582009$.

O revendedor pretende averiguar se o valor esperado da duração é de 250 horas. Teste esta hipótese ao nível de significância de 10%, contra a hipótese alternativa defendida pelo produtor que defende que tal valor esperado é superior a 250 horas.

- **V.a. de interesse**

  X = duração (em horas de uso) de certa marca de pilha

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ^2 desconhecido

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 250$ horas vs. $H_1 : \mu > \mu_0 = 250$ horas

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim_{H_0} t_{(n-1)} \tag{8.24}$$

dado que pretendemos efectuar teste sobre o valor esperado de uma população normal com variância desconhecida.

- **Região de rejeição de** H_0 (para valores da estatística de teste)

Estamos a lidar com

  – um teste unilateral superior ($H_1 : \mu > \mu_0$),

pelo que, quanto maior for a estimativa de MV de μ, $\bar{x}$, mais nos devemos inclinar para a rejeição H_0. Assim sendo, a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à direita

$$W = (c, +\infty), \tag{8.25}$$

onde $c : P(\text{Rejeitar}H_0|H_0) = \alpha_0$, i.e.,

$$c = F^{-1}_{t_{(n-1)}}(1 - \alpha_0) = F^{-1}_{t_{(10-1)}}(1 - 0.10) \stackrel{tabela}{=} 1.383. \tag{8.26}$$

- **Decisão**

Tendo em conta que $n = 10$ e a média e a variância corrigida da amostra são iguais a

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2411}{10} = 241.1 \tag{8.27}$$

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_i^2\right) - n\,(\bar{x})^2\right] \\ &= \frac{1}{10-1}[582009 - 10 \times (241.1)^2] \\ &= 79.66 \end{aligned} \tag{8.28}$$

(respectivamente), o valor observado da estatística de teste é dado por

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{241.1 - 250}{\sqrt{79.66/10}} \\ &= -3.15 \\ &\notin W = (1.383, +\infty). \end{aligned} \tag{8.29}$$

Deste modo, concluímos que não devemos rejeitar H_0 ao n.s. de 10% nem a qualquer outro n.s. menor que 10%.[4] •

Exemplo 8.34 (CASO 2) — **Teste sobre o valor esperado de população com distribuição arbitrária, variância desconhecida**
Retome o Exemplo 8.33 e desta feita assuma que o revendedor desconhece a distribuição da duração dessas baterias, adquiriu um grande lote e seleccionou uma amostra não de 10 mas sim de 50 baterias, tendo obtido $\bar{x} = 241.1$ e $s^2 = 398.3$.

Volte a testar as hipóteses nula e alternativa consideradas anteriormente ao n.s. de 10%.

- **V.a. de interesse**

  $X =$ duração (em horas de uso) de certa marca de pilha

- **Situação**

  X distribuição arbitrária

  $\mu = E(X)$ DESCONHECIDO

  $\sigma^2 = V(X)$ desconhecido

  n suficientemente grande ($n = 50 > 30$)

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 250$ horas vs. $H_1 : \mu > \mu_0 = 250$ horas

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{a}{\sim}_{H_0} \text{normal}(0,1) \tag{8.30}$$

  porque pretendemos efectuar teste sobre o valor esperado de uma população com distribuição arbitrária e variância desconhecida e a dimensão da amostra é suficientemente grande e justifica o recurso a um resultado limite.

[4] É curioso notar que caso estivéssemos a testar $H_0 : \mu = \mu_0 = 250$ horas vs. $H_1 : \mu < \mu_0$ rejeitaríamos H_0 ao referido n.s. (Justique...)

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Tratando-se de um teste unilateral superior ($H_1 : \mu > \mu_0$), a região de rejeição de H_0 (para valores da estatística de teste) é à mesma um intervalo à direita

  $$W = (c, +\infty),$$

  onde agora $c : P(\text{Rejeitar}H_0|H_0) \simeq \alpha_0$, pelo que
  $$c = \Phi^{-1}(1 - \alpha_0) = \Phi^{-1}(1 - 0.10) \stackrel{tabela}{=} 1.6449.$$

- **Decisão**

  Uma vez que $n = 50$, $\bar{x} = 241.1$ e $s^2 = 398.3$ tem-se:

  $$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{241.1 - 250}{\sqrt{398.3/50}} \\ &= -3.15 \\ &\not\in W = (1.6449, +\infty). \end{aligned} \tag{8.31}$$

  Pode então afirmar-se que não devemos rejeitar H_0 a qualquer outro n.s. menor ou igual a 10%. •

Terminamos esta secção apresentando um exemplo similar ao apresentado no final da secção 7.4. Diz respeito a um teste sobre o valor esperado de uma v.a. de interesse X cuja distribuição depende de um único parâmetro desconhecido e que, embora o valor esperado $E(X)$ e a variância $V(X)$ dependam desse mesmo parâmetro, o TLC só por si basta para obter uma v.a. fulcral (logo uma estatística de teste para o valor esperado sem que seja necessário estimar $V(X)$.

Exemplo/Exercício 8.35 — Teste sobre valor esperado de população com distribuição arbitrária e dimensão da amostra suficientemente grande, variância desconhecida mas que não carece de estimação...
Admita que o desvio absoluto de uma medição instrumental em relação a uma norma é uma v.a. X com distribuição exponencial com parâmetro λ desconhecido.[5]

(a) Calcule $E(\bar{X})$ e $V(\bar{X})$, onde $\bar{X}$ representa, naturalmente, a média de uma amostra aleatória de dimensão n proveniente da população X. Tirando partido dos resultados anteriores mostre que, para n suficientemente grande, se tem

$$Z = (\lambda\bar{X} - 1)\sqrt{n} \stackrel{a}{\sim} \text{normal}(0,1).$$

(b) Tendo em conta a v.a. fulcral Z, teste a hipótese $H_0 : \lambda = \lambda_0 = 1$ contra a alternativa bilateral $H_1 : \lambda \neq \lambda_0$, ao nível de significância de $\alpha_0 = 10\%$ e à luz do facto da soma dos desvios absolutos da amostra $\underline{x} = (x_1, \ldots, x_{40})$ ser igual a $\sum_{i=1}^{40} x_i = 47.5408$.

- **V.a. de interesse**

  X = desvio absoluto de uma medição instrumental em relação a uma norma

- **Situação**

  $X \sim$ exponencial(λ)

  λ DESCONHECIDO

  n suficientemente grande ($n = 40 > 30$)

- **Hipóteses**

  $H_0 : \lambda = \lambda_0 = 1$ vs. $H_1 : \lambda \neq \lambda_0$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **V.a. fulcral para λ e estatística de teste**

  Comece-se por notar que:

$$\begin{aligned} E(\bar{X}) &= E(X) = \frac{1}{\lambda} \\ V(\bar{X}) &= \frac{V(X)}{n} = \frac{1}{n\,\lambda^2} < +\infty. \end{aligned}$$

[5] Adaptado do Exame de 18 de Janeiro de 2003.

Então, de acordo com o TLC, pode afirmar-se que

$$\begin{aligned}
Z &= \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} \\
&= \frac{\bar{X} - \frac{1}{\lambda}}{\sqrt{\frac{1}{n\,\lambda^2}}} \\
&= (\lambda \bar{X} - 1)\sqrt{n} \\
&\stackrel{a}{\sim} \text{normal}(0,1). \qquad (8.32)
\end{aligned}$$

Substituindo λ por λ_0 na expressão desta v.a. fulcral, obtemos a seguinte estatística de teste com distribuição aproximada conhecida sob H_0:

$$T = (\lambda_0 \bar{X} - 1)\sqrt{n} \stackrel{a}{\sim}_{H_0} \text{normal}(0,1). \qquad (8.33)$$

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Tratando-se de um teste bilateral, a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0|H_0) \simeq \alpha_0$, i.e.,

$$c = \Phi^{-1}(1 - \alpha_0/2) = \Phi^{-1}(1 - 0.10/2) \stackrel{tabela}{=} 1.6449.$$

- **Decisão**

  Dado que o valor observado da estatística de teste é igual a

$$\begin{aligned}
t &= (\lambda_0 \bar{x} - 1)\sqrt{n} \\
&= (1 \times 47.5408/40 - 1)\sqrt{40} \\
&= 1.192 \\
&\notin W = (-\infty, -1.6449) \cup (1.6449, +\infty),
\end{aligned}$$

  conclui-se que não devemos rejeitar H_0 ao n.s. de 10%, nem a qualquer outro n.s. menor que 10%. •

8.6 Um método alternativo de decisão em testes de hipóteses: cálculo do p-value

Motivação 8.36 — Cálculo do p-value
A decisão pela rejeição ou não de H_0 depende crucialmente do nível de significância α_0 que tenhamos considerado.

Ora, em vez de fixarmos o n.s. do teste, de identificarmos a região de rejeição de H_0 e de verificarmos se o valor observado da estatística de teste (t) pertence ou não a tal região, podemos proceder do seguinte modo:

- tomar o valor de t

e averiguar

- para que níveis de significância se decide pela REJEIÇÃO de H_0 e
- para que níveis de significância se decide pela NÃO REJEIÇÃO de H_0. •

Passemos agora a um exemplo por forma a continuar a motivar o cálculo do que designaremos por p-value.

Exemplo 8.37 — Cálculo do p-value
Retomemos o problema do fabricante de baterias de 12V, descrito no Exemplo 8.22. Recordemos que temos:

- **V.a. de interesse**

  X = duração de bateria de 12V

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  $\sigma^2 = 50^2 \text{ dias}^2$ conhecido.

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 1000$ dias vs. $H_1 : \mu \neq \mu_0 = 1000$ dias

- **Nível de significância**

  Desta feita consideremos um n.s. arbitrário α_0.

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0,1) \tag{8.34}$$

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Sabemos que se trata de uma reunião de intervalos do tipo $W_{\alpha_0} = (-\infty, -c_{\alpha_0}) \cup (c_{\alpha_0}, +\infty)$, onde

$$c_{\alpha_0} = \Phi^{-1}(1 - \alpha_0/2). \tag{8.35}$$

- **Decisão**

  Ora, já vimos que o valor observado da estatística de teste é igual a $t = -2.8$. Logo, consultando a tabela de quantis da distribuição normal$(0,1)$ conclui-se que:

  - por um lado,

$$t = -2.8 \in W_{\alpha_0} = \begin{cases} (-\infty, -1.6449) \cup (1.6449, +\infty), & \text{para } \alpha_0 = 10\% \\ (-\infty, -1.9600) \cup (1.9600, +\infty), & \text{para } \alpha_0 = 5\% \\ (-\infty, -2.5758) \cup (2.5758, +\infty), & \text{para } \alpha_0 = 1\% \\ (-\infty, -2.7478) \cup (2.7478, +\infty), & \text{para } \alpha_0 = 0.6\%, \end{cases} \tag{8.36}$$

    devendo rejeitar-se H_0 a qualquer n.s. $\alpha_0 \geq 0.6\%$;

  - por outro lado,

$$t = -2.8 \notin W_{\alpha_0} = \begin{cases} (-\infty, -2.8782) \cup (2.8782, +\infty), & \text{para } \alpha_0 = 0.4\% \\ (-\infty, -3.0902) \cup (3.0902, +\infty), & \text{para } \alpha_0 = 0.2\% \\ \ldots, \end{cases} \tag{8.37}$$

    e neste caso não se deve rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.4\%$.

  Importa ainda notar que, caso o ponto crítico que define a região de rejeição de H_0 fosse

$$c = |t|, \tag{8.38}$$

  teríamos a seguinte região de rejeição de H_0

$$W = (-\infty, -|t|) \cup (|t|, +\infty), \tag{8.39}$$

  com nível de significância associado igual a (Esquema...)

$$
\begin{aligned}
P(T < -|t| \text{ ou } T > |t| \mid H_0) &= 2 \times [1 - \Phi(|t|)] \\
&= 2 \times [1 - \Phi(2.8)] \\
&= 2 \times [1 - 0.9974] \\
&= 0.0052 \\
&= 0.52\%. \qquad (8.40)
\end{aligned}
$$

Para além disso,

$$t = -2.8 \notin W_{0.52\%} = (-\infty, -2.8) \cup (2.8, +\infty) \qquad (8.41)$$

donde

- não rejeitaríamos H_0 ao n.s. de 0.52% nem a qualquer n.s. menor que 0.52%;
- rejeitaríamos H_0 a qualquer n.s. maior que 0.52%.

0.52% é, neste exemplo, o "ponto de viragem" da nossa decisão pela rejeição ou não de H_0. Este ponto é genericamente designado de... •

Definição informal 8.38 — P-value

Dado o valor observado da estatística de teste

- o $p - value$ é o maior nível de significância que leva à não rejeição de H_0.

Para além disso, devemos agir do seguinte modo:

- não rejeitar H_0 a qualquer nível de significância $\alpha_0 \leq p - value$;
- rejeitar H_0 a qualquer nível de significância $\alpha_0 > p - value$.[6]

E quanto menor for o $p - value$, maior é a evidência contra H_0. •

[6]Posto isto, podemos acrescentar que o $p - value$ é o "menor" (aqui com as devidas aspas) nível de significância que leva à rejeição de H_0.

Nota 8.39 — P-value

O cálculo do $p-value$ depende do aspecto da região de rejeição de H_0 (para valores da estatística de teste:

W	**Teste**	**P-value**	**Esquema...**
$(-\infty, c)$	unilateral inferior	$P(T < t \mid H_0) = F_{T\|H_0}(t)$	
$(c, +\infty)$	unilateral superior	$P(T > t \mid H_0) = 1 - F_{T\|H_0}(t)$	
$(-\infty, -c) \cup (c, +\infty)$	bilateral e $T\|H_0$ com distrib. simétrica em relação à origem	$P(T < -\|t\| \text{ ou } T > \|t\| \mid H_0)$ $= 2 \times [1 - F_{T\|H_0}(\|t\|)]$	

•

8.7 Testes de hipóteses sobre a igualdade de valores esperados de duas populações, variâncias desconhecidas.

No âmbito desta disciplina, caso pretendamos confrontar os valores esperados de duas populações normais independentes com variâncias desconhecidas e estejamos a lidar com amostras de dimensões que não são suficientemente grandes que justifiquem o recurso a um resultado assintótico, teremos que assumir que as variâncias são iguais, aliás, tal como aconteceu no capítulo anterior.

Exemplo 8.40 (CASO 1) — **Teste sobre a igualdade de valores esperados de populações normais, variâncias desconhecidas**
Foram efectuados estudos em Los Angeles e New York com o objectivo de determinar a concentração de monóxido de carbono (CO) perto de vias rápidas. Para isso recolheram-se amostras de ar para as quais se determinou a respectiva concentração de CO (usando para tal um espectómetro). Os resultados das medições em *ppm* (partes por milhão) foram no período de uma semana:

- Los Angeles — $\underline{x}_1 = (112.2, 118.4, 114.1)$
- New York — $\underline{x}_2 = (101.1, 102.2, 100.4, 98.6, 88.2)$.

Posteriormente, num programa televisivo em que se debatiam questões ambientais, o presidente da câmara de New York afirmou que: "A média da concentração de CO em Los Angeles é superior ou igual à de New York".

Diga se esta afirmação é consistente com os dados, determinando para o efeito um intervalo para o $p - value$.

- **V.a. de interesse**

  $X_1 =$ concentração de CO em Los Angeles

  $X_2 =$ concentração de CO em New York

- **Situação**

  $X_1 \sim normal(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim normal(\mu_2, \sigma_2^2)$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos, no entanto, assume-se que são IGUAIS: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

  $n_1 \leq 30$ ou $n_2 \leq 30$

- **Hipóteses**

  $H_0 : \mu_1 - \mu_2 \geq \mu_0 = 0$ vs. $H_1 : \mu_1 - \mu_2 < \mu_0$

- **Nível de significância**

  α_0 com valor arbitrário

- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \times (\frac{1}{n_1} + \frac{1}{n_2})}} \sim_{\mu_1-\mu_2=\mu_0} t_{(n_1+n_2-2)} \tag{8.42}$$

  pois estamos a efectuar teste sobre a igualdade dos valores esperados de duas populações independentes com distribuições normais, cujas variâncias, embora desconhecidas, se assume serem iguais.

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Tratando-se de

  – um teste unilateral inferior ($H_1 : \mu_1 - \mu_2 < \mu_0 = 0$),

  concluímos que a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à esquerda do tipo $W = (-\infty, c)$, onde $c : P(\text{Rejeitar } H_0|\mu_1 - \mu_2 = \mu_0) = \alpha_0$, ou seja,

$$c = F^{-1}_{t_{(n_1+n_2-2)}}(\alpha_0) = -F^{-1}_{t_{(n_1+n_2-2)}}(1 - \alpha_0). \tag{8.43}$$

- **Decisão**

  Ora,

  $n_1 = 3$, $\bar{x}_1 = 114.9$ e $s_1^2 = 10.09$
  $n_2 = 5$, $\bar{x}_2 = 98.1$ e $s_2^2 = 32.34$,

  logo o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \times (\frac{1}{n_1} + \frac{1}{n_2})}} \\ &= \frac{(114.9 - 98.1) - 0}{\sqrt{\frac{(3-1)10.09+(5-1)32.34}{3+5-2} \times (\frac{1}{3} + \frac{1}{5})}} \\ &= 3.237. \end{aligned} \tag{8.44}$$

Uma vez que este teste está associado a uma região de rejeição de H_0 que é um intervalo à esquerda temos:

$$\begin{aligned} p - value &= P(T < t \mid \mu_1 - \mu_2 = \mu_0) \\ &= F_{T|\mu_1-\mu_2=\mu_0}(t) \\ &= F_{t_{(3+5-2)}}(3.237) \\ &\stackrel{maq.}{=} 0.991123. \end{aligned} \tag{8.45}$$

Atendendo a este valor:

- não devemos rejeitar H_0 a n.s. $\alpha_0 \leq 99.1123\%$;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 99.1123\%$.

Recorrendo às tabelas de quantis da distribuição de t-student com 6 graus de liberdade podemos adiantar um intervalo para o $p-value$ deste teste. Com efeito, ao enquadrarmos convenientemente $t = 3.237$ por dois quantis, obtemos sucessivamente

$$\begin{aligned} F^{-1}_{t_{(3+5-2)}}(0.99) = 3.143 &< t = 3.237 < 3.707 = F^{-1}_{t_{(3+5-2)}}(0.995) \\ 0.99 &< p - value = F_{t_{(3+5-2)}}(3.237) < 0.995. \end{aligned} \tag{8.46}$$

Deste modo, o intervalo para o $p - value$ é (0.99,0.995), pelo que

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 99\%$;
- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 99.5\%$.

Mais, pela ordem de grandeza do $p - value$, a hipótese defendida pelo presidente da câmara de New York (H_0) afigura-se altamente consistente com os dados obtidos, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$. •

Ainda a propósito do exemplo anterior, note-se que, pelo facto das amostras possuirem dimensão 3 e 5, não seria lícito recorrer a qualquer resultado assintótico, pelo que a solução satisfatória no âmbito desta disciplina passa por assumir que as variâncias, embora desconhecidas, são iguais. Acrescente-se, no entanto, que um teste prévio de igualdade das mesmas permite-nos concluir que esta hipótese não deve ser rejeitada a qualquer dos níveis usuais de significância.

Mas nem sempre é razoável admitir que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, muito em particular quando as estimativas de σ_1^2 e σ_2^2 são bem distintas. Recorde-se que, caso $n_1 > 30$ e $n_2 > 30$, podemos invocar dois resultados assintóticos, nomeadamente, o TLC e o Teorema de Slutsky, e deixa de ser necessário assumir que as duas populações são normais e que as variâncias, embora desconhecidas, são iguais.

Exemplo 8.41 (CASO 2) — **Teste sobre a igualdade de valores esperados de populações independentes com distribuição arbitrária e amostras suficientemente grandes, variâncias desconhecidas**

Para comparar a resistência ao uso de dois tipos de materiais cerâmicos usados em pavimentos, foram instalados 81 mosaicos do primeiro tipo e 121 do segundo tipo num corredor movimentado de uma superfície comercial. Após um ano o seu desgaste foi medido numa escala conveniente. Para os mosaicos do primeiro tipo obteve-se $\bar{x}_1 = 290$ e $s_1 = 12$, enquanto que para os do segundo tipo os resultados foram $\bar{x}_2 = 321$ e $s_2 = 14$.

O fabricante do primeiro tipo de material afirma que o valor esperado do desgaste dos seus mosaicos é inferior ao dos mosaicos do segundo tipo de material. Efectue um teste de hipóteses que entender razoável, assumindo que os desgastes em mosaicos diferentes são v.a. independentes e um nível de significância de 1%.[7]

- **V.a. de interesse**

  X_1 = desgaste dos mosaicos do primeiro tipo de material

  X_2 = desgaste dos mosaicos do segundo tipo de material

- **Situação**

  $X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (possivelmente normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2,\ i = 1, 2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos mas não necessariamente iguais

  $n_1 > 30$ e $n_2 > 30$, i.e., ambas as amostras são suficientemente grandes

- **Hipóteses**

  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$ vs. $H_1 : \mu_1 - \mu_2 < \mu_0 = 0$ (pretensão do fabricante do primeiro tipo de material)

- **Nível de significância**

  $\alpha_0 = 1\%$

- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \stackrel{a}{\sim}_{H_0} \text{normal}(0, 1) \tag{8.47}$$

[7] Adaptado do Exame de 7 de Fevereiro de 2004.

por tratar-se de teste sobre a igualdade dos valores esperados de duas populações independentes com distribuição arbitrária com variâncias desconhecidas e estarmos a lidar com amostras suficientemente grandes.

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Lidamos novamente com um teste unilateral inferior ($H_1 : \mu_1 - \mu_2 < \mu_0 = 0$), logo a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à esquerda $W = (-\infty, c)$, onde $c : P(\text{Rejeitar } H_0|H_0) \simeq \alpha_0$, i.e.,

$$c = \Phi^{-1}(\alpha_0) = -\Phi^{-1}(1-\alpha_0) = -\Phi^{-1}(0.99) = -2.3263. \tag{8.48}$$

- **Decisão**

  Uma vez que

  $n_1 = 81$, $\bar{x}_1 = 290$ e $s_1 = 12$

  $n_2 = 121$, $\bar{x}_2 = 321$ e $s_2 = 14$,

  o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{(290 - 321) - 0}{\sqrt{\frac{12^2}{81} + \frac{14^2}{121}}} \\ &= -16.82 \\ &\in W = (-\infty, -2.3263). \end{aligned} \tag{8.49}$$

  Assim sendo, devemos rejeitar H_0 a qualquer n.s. $\alpha_0' \geq 1\%$. Posto isto podemos afirmar que a pretensão do fabricante do primeiro tipo de material (desgaste esperado menor) é consistente com os dados aos níveis usuais de significância. •

8.8 Testes de hipóteses para a variância de uma população normal.

É também costume especular sobre a variância de uma v.a. de interesse. Caso esta v.a. possua distribuição normal, podemos adiantar uma estatística de teste com distribuição exacta do qui-quadrado sob a validade de H_0.

Exemplo/Exercício 8.42 — Teste de hipóteses sobre a variância de uma população normal

(a) Admita que a resistência à tensão de uma fibra têxtil (em *psi*) possui distribuição normal, com valor esperado μ desconhecido, e averigue se a variância é igual a $\sigma_0^2 = 50\,psi^2$ ou se pelo contrário é superior a este valor, à luz da amostra $\underline{x} = (1018, 982, 1007, 1015, 978)$ e ao n.s. de 5%.

- **V.a. de interesse**

  $X =$ resistência à tensão de uma fibra têxtil (em *psi*)

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ desconhecido

  σ^2 DESCONHECIDO.

- **Hipóteses**

  $H_0 : \sigma^2 = \sigma_0^2 = 50\,psi^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

  $$T = \frac{(n-1)\,S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{(n-1)} \tag{8.50}$$

  pois pretendemos efectuar um teste sobre a variância de uma população normal com valor esperado desconhecido.

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Tratando-se de um teste unilateral superior ($H_1 : \sigma^2 > \sigma_0^2$), a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à direita do tipo $W = (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, ou seja,

  $$c = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(5-1)}}(1-0.05) \stackrel{tabela}{=} 9.488. \tag{8.51}$$

- **Decisão**

  O valor observado da estatística é igual a

$$\begin{aligned} t &= \frac{(n-1)\,s^2}{\sigma_0^2} \\ &= \frac{(5-1)\times 351.5}{50} \\ &= 28.12 \\ &\in W = (9.488, +\infty). \end{aligned} \tag{8.52}$$

  Deste modo devemos rejeitar H_0 ao n.s. de 5% ou a qualquer outro n.s. maior que 5%.

(b) Faça uso de uma máquina (resp. das tabelas) para determinar um valor (resp. um intervalo) para o $p-value$ do teste da alínea anterior. Que decisões deverá tomar ao recorrer ao valor do (resp. intervalo para o) $p-value$? •

8.9 Outro método alternativo de decisão em testes de hipóteses: relação entre intervalos de confiança e testes bilaterais.

Existe ainda uma outra alternativa de efectuação de testes de hipóteses que passa por invocar uma analogia entre intervalos de confiança e testes, nomeadamente bilaterais, como se pode ver na proposição que se segue.

Proposição 8.43 — **Relação entre intervalos de confiança e testes de hipóteses bilaterais**
Seja

- $IC_{(1-\alpha_0)\times 100\%}(\theta) = [l, u]$

um intervalo de confiança para θ. Então este IC leva à

- rejeição de $H_0 : \theta = \theta_0$ ao nível de significância α_0 [8]

a favor da hipótese alternativa bilateral $H_1 : \theta \neq \theta_0$, caso

- $\theta_0 \notin IC_{(1-\alpha_0)\times 100\%}(\theta)$. •

Este resultado faz todo o sentido uma vez que, caso o valor conjecturado para θ, θ_0, não pertença ao conjunto de valores razoáveis para θ associados ao grau de confiança $(1 - \alpha_0) \times 100\%$, faz todo o sentido rejeitar H_0 ao n.s. de α_0.

Nota 8.44 — **Relação entre intervalos de confiança e testes bilaterais**
Para invocar esta analogia, é necessário que a estatística de teste tenha sido obtida à custa da v.a. fulcral para θ que usada na construção de $IC_{(1-\alpha_0)\times 100\%}(\theta)$. •

[8] Ou a qualquer outro n.s. superior a α_0.

Exemplo 8.45 — Relação entre intervalos de confiança e testes bilaterais
Retomemos o Exemplo 8.42, construamos um IC a 95% para σ^2 e averiguemos a razoabilidade de $H_0 : \sigma^2 = \sigma_0^2 = 50psi^2$ contra a hipótese alternativa bilateral $H_1 : \sigma^2 \neq \sigma_0^2$ ao nível de significância de $\alpha_0 = 5\%$.

- **Passo 1 — Selecção da v.a. fulcral para** σ^2

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \tag{8.53}$$

já que estamos a obter um IC para a variância de uma população normal com valor esperado desconhecido.

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$\begin{cases} a_{\alpha_0} = F^{-1}_{\chi^2_{(n-1)}}(\alpha_0/2) = F^{-1}_{\chi^2_{(4)}}(0.025) \stackrel{tabela}{=} 0.484 \\ b_{\alpha_0} = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha_0/2) = F^{-1}_{\chi^2_{(4)}}(0.975) \stackrel{tabela}{=} 11.14. \end{cases} \tag{8.54}$$

- **Passo 3 — Inversão da desigualdade** $a_{\alpha_0} \leq Z \leq b_{\alpha_0}$

[TPC]

- **Passo 4 — Concretização**

$$\begin{aligned} IC_{(1-\alpha_0)\times 100\%}(\sigma^2) &= \left[\frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha_0/2)}, \frac{(n-1)s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha_0/2)}\right] \\ &= \left[\frac{(5-1)351.5}{11.14}, \frac{(5-1)351.5}{0.484}\right] \\ &= [126.212, 2904.959]. \end{aligned} \tag{8.55}$$

- **Hipóteses**

$H_0 : \sigma^2 = \sigma_0^2 = 50psi^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$

- **Decisão**

Invocando a relação entre intervalos de confiança e testes de hipóteses, pode concluir-se que, pelo facto de

$$\sigma_0^2 = 50 \notin IC_{95\%}(\sigma^2) = [126.212, 2904.959], \tag{8.56}$$

devemos rejeitar H_0 ao n.s. de 5% (a favor de $H_1 : \sigma^2 \neq \sigma_0^2$) ou a qualquer outro n.s. maior que 5%. •

8.10 Testes de hipóteses para parâmetros de populações não normais uniparamétricas.

No âmbito desta disciplina aprenderemos, por exemplo, a efectuar testes assintóticos sobre uma probabilidade de sucesso desconhecida. Vejamos um exemplo.

Exemplo 8.46 — **Teste sobre uma probabilidade de sucesso**
Um fabricante de cerveja afirma que a sua bebida tem a preferência de 40% dos apreciadores de cerveja. Recolhida uma amostra, constatou-se que 54 de 150 apreciadores preferem tal marca. Averigue a razoabilidade da afirmação do fabricante ao nível de significância de 4%.[9]

- **V.a. de interesse**

  Neste caso estamos a lidar com a seguinte v.a. de interesse:

$$X = \begin{cases} 1, & \text{se o indivíduo prefere a marca de cerveja} \\ 0, & \text{c.c.} \end{cases} \tag{8.57}$$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  p DESCONHECIDO

  n suficientemente grande ($n = 150 >> 30$)

- **Hipóteses**

  $H_0 : p = p_0 = 0.40$ vs. $H_1 : p \neq p_0$

- **Nível de significância**

  $\alpha_0 = 4\%$

- **Estatística de teste**

  Sabe-se que o estimador de MV de p é

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{8.58}$$

  onde $X_i \sim_{i.i.d.} X$. Para além disso,

[9] Adaptado do Exame de 4 de Fevereiro de 2003.

$$E(\bar{X}) = E(X) = p \tag{8.59}$$

$$V(\bar{X}) = \frac{1}{n}V(X) = \frac{p(1-p)}{n} < +\infty. \tag{8.60}$$

Então pelo TLC pode afirmar-se que

$$Z' = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{a}{\sim} \text{normal}(0,1), \tag{8.61}$$

pelo que a estatística de teste é

$$T = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{a}{\sim}_{H_0} \text{normal}(0,1). \tag{8.62}$$

De notar que esta estatística de teste, ao contrário das que usámos até ao momento, não foi obtida à custa da v.a. fulcral

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} \text{normal}(0,1), \tag{8.63}$$

que teríamos utilizado para obter um IC assintótico para o parâmetro p. Esta decisão prende-se essencialmente com o facto de não ser necessária qualquer inversão de duplas desigualdades do tipo $-c \leq Z \leq c$ no âmbito dos testes de hipóteses.[10]

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Tratando-se de um teste bilateral, a região de rejeição de H_0, escrita para valores da estatística de teste, é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde c : $P(\text{Rejeitar } H_0|H_0) \simeq \alpha_0$, ou seja,

$$c = \Phi^{-1}(1 - \alpha_0/2) = \Phi^{-1}(1 - 0.04/2) \stackrel{tabela}{=} 2.0537. \tag{8.64}$$

- **Decisão**

  Dado que o valor observado da estatística de teste é igual a

$$t = \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

[10]Relembre-se que a utilização da v.a. fulcral para p, Z', foi evitada pois tornava a inversão da desigualdade do Passo 3 da obtenção de ICs extraordinariamente difícil uma vez que p figura em numerador e em denominador (neste último caso sob o sinal de uma raíz).

$$\begin{aligned}
&= \frac{\frac{54}{150} - 0.4}{\sqrt{\frac{0.4(1-0.4)}{150}}} \\
&= \frac{.36 - 0.4}{0.04} \\
&= -1 \\
&\notin W = (-\infty, -2.0537) \cup (2.0537, +\infty),
\end{aligned}$$

conclui-se que não devemos rejeitar H_0 ao n.s. de 4%. Deste modo, pode dizer-se que a afirmação do fabricante é consistente com os dados ao n.s. de 4%, bem como a qualquer outro n.s. menor que 4%. •

Exercício 8.47 — P-value de teste sobre uma probabilidade de sucesso
Prove que o valor aproximado para o $p - value$ do teste assintótico efectuado no Exemplo 8.46 é igual a $2 \times [1 - \Phi(|-1|)] = 0.3174$. •

Refira-se ainda que a seguir podem encontrar-se dois quadros-resumo que sistematizam os testes de hipóteses paramétricas aqui leccionados considerando em qualquer dos casos um nível de significância de α_0. Nas suas cinco colunas constam:

- as circunstâncias em que estamos a efectuar o teste de hipóteses;
- a hipótese nula (H_0);
- a estatística de teste a utilizar nesta situação (T);
- a hipótese alternativa (H_1);
- a região de rejeição de H_0 escrita para valores da estatística de teste (W).

Quadro-resumo 2a: Alguns testes de hipóteses paramétricas.

Situação	Hipótese nula	Estatística de teste	Hipótese alternativa	Região de Rejeição
$X \sim \text{normal}(\mu, \sigma^2)$ σ^2 conhecido	$H_0 : \mu = \mu_0$	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0,1)$	$H_1 : \mu \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$
X com distribuição arbitrária (não normal): $E(X) = \mu$, $V(X) = \sigma^2$ σ^2 conhecido n suficientemente grande	$H_0 : \mu = \mu_0$	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$	$H_1 : \mu \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$
$X_1 \sim \text{normal}(\mu_1, \sigma_1^2)$ $\perp\!\!\!\perp$ $X_2 \sim \text{normal}(\mu_2, \sigma_2^2)$ σ_1^2, σ_2^2 conhecidos	$H_0 : \mu_1 - \mu_2 = \mu_0$	$\frac{(\bar{X}_1-\bar{X}_2)-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim_{H_0} \text{normal}(0,1)$	$H_1 : \mu_1 - \mu_2 \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu_1 - \mu_2 > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu_1 - \mu_2 < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$
$X_1 \perp\!\!\!\perp X_2$, com distribuições arbitrárias (não normais): $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2, i = 1, 2$ σ_1^2, σ_2^2 conhecidos n_1, n_2 suficientemente grandes	$H_0 : \mu_1 - \mu_2 = \mu_0$	$\frac{(\bar{X}_1-\bar{X}_2)-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$	$H_1 : \mu_1 - \mu_2 \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu_1 - \mu_2 > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu_1 - \mu_2 < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$
$X \sim \text{normal}(\mu, \sigma^2)$ σ^2 desconhecido	$H_0 : \mu = \mu_0$	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim_{H_0} t_{(n-1)}$	$H_1 : \mu \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = F^{-1}_{t_{(n-1)}}(1-\alpha_0/2)$
			$H_1 : \mu > \mu_0$	$(c, +\infty),\ c = F^{-1}_{t_{(n-1)}}(1-\alpha_0)$
			$H_1 : \mu < \mu_0$	$(-\infty, c),\ c = -F^{-1}_{t_{(n-1)}}(1-\alpha_0)$
X com distribuição arbitrária: $E(X) = \mu, V(X) = \sigma^2$ σ^2 desconhecido n suficientemente grande	$H_0 : \mu = \mu_0$	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$	$H_1 : \mu \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$

Quadro-resumo 2b: Alguns testes de hipóteses paramétricas (cont.).

Situação	Hipótese nula	Estatística de teste	Hipótese alternativa	Região de Rejeição
$X_1 \sim$ normal(μ_1, σ_1^2) $\perp\!\!\!\perp$ $X_2 \sim$ normal(μ_2, σ_2^2) σ_1^2, σ_2^2 desconhecidos mas IGUAIS $n_1 \leq 30$ ou $n_2 \leq 30$	$H_0 : \mu_1 - \mu_2 = \mu_0$	$\frac{(\bar{X}_1-\bar{X}_2)-\mu_0}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim_{H_0} t_{(n_1+n_2-2)}$	$H_1 : \mu_1 - \mu_2 \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha_0/2)$
			$H_1 : \mu_1 - \mu_2 > \mu_0$	$(c, +\infty),\ c = F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha_0)$
			$H_1 : \mu_1 - \mu_2 < \mu_0$	$(-\infty, c),\ c = -F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha_0)$
$X_1 \perp\!\!\!\perp X_2$ com distribuições arbitrárias (possivelmente normais): $E(X_i) = \mu_i, V(X_i) = \sigma_i^2, i = 1, 2$ σ_1^2, σ_2^2 desconhecidos mas não necessariamente iguais $n_1 > 30$ e $n_2 > 30$	$H_0 : \mu_1 - \mu_2 = \mu_0$	$\frac{(\bar{X}_1-\bar{X}_2)-\mu_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \overset{a}{\sim}_{H_0}$ normal$(0,1)$	$H_1 : \mu_1 - \mu_2 \neq \mu_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : \mu_1 - \mu_2 > \mu_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : \mu_1 - \mu_2 < \mu_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$
$X \sim$ normal(μ, σ^2) μ desconhecido	$H_0 : \sigma^2 = \sigma_0^2$	$\frac{(n-1)S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{(n-1)}$	$H_1 : \sigma^2 \neq \sigma_0^2$	$(0, a) \cup (b, +\infty)$, $a = F^{-1}_{\chi_{(n-1)}}(\alpha_0/2),\ b = F^{-1}_{\chi_{(n-1)}}(1-\alpha_0/2)$
			$H_1 : \sigma^2 > \sigma_0^2$	$(c, +\infty),\ c = F^{-1}_{\chi_{(n-1)}}(1-\alpha_0)$
			$H_1 : \sigma^2 < \sigma_0^2$	$(0, c),\ c = F^{-1}_{\chi_{(n-1)}}(\alpha_0)$
$X \sim$ Bernoulli(p) n suficientemente grande $(n > 30)$	$H_0 : p = p_0$	$\frac{\bar{X}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{H_0}$ normal$(0,1)$	$H_1 : p \neq p_0$	$(-\infty, -c) \cup (c, +\infty),\ c = \Phi^{-1}(1-\alpha_0/2)$
			$H_1 : p > p_0$	$(c, +\infty),\ c = \Phi^{-1}(1-\alpha_0)$
			$H_1 : p < p_0$	$(-\infty, c),\ c = -\Phi^{-1}(1-\alpha_0)$

8.11 Teste de ajustamento do qui-quadrado de Pearson.

O teste de ajustamento do qui-quadrado permite averiguar a adequação de

- uma distribuição com todos os parâmetros conhecidos (hipótese simples — subsecções 8.11.1 e 8.11.4) ou de
- uma distribuição com pelo menos um parâmetro desconhecido, i.e., uma família de distribuições (hipótese composta — subsecções 8.11.2 e 8.11.4)

a um conjunto de dados referente a

- uma v.a. de interesse do tipo discreto ou contínuo.

Este teste exige, no entanto,

- um grande número de observações (uma vez que se baseia num resultado assintótico)

e pressupõe que

- os dados estejam organizados em classes e disponhamos de uma tabela de frequências.

De notar que estas classes devem não só ser disjuntas, como cobrir todo o contradomínio da v.a. de interesse.

A apresentação deste mesmo teste far-se-á essencialmente à custa de exemplos. Contudo importa referir que recorreremos ao procedimento geral de testes de hipóteses e que lidaremos com uma estatística de teste cujo valor observado reflecte as

- discrepâncias (relativas)

entre

- as frequências absolutas observadas das classes da tabela de frequências e
- as frequências esperadas (ou suas estimativas) dessas mesmas classes sob a validade da hipótese nula.

Esta estatística de teste possui distribuição assintótica do qui-quadrado, sob a validade de H_0. Acrescente-se também que o número de graus de liberdade desta distribuição do qui-quadrado depende do número de:

- classes em que estão organizados os dados da tabela de frequências;
- parâmetros não especificados em H_0 e que, consequentemente, é necessário estimar.

Convém ainda referir que no decurso do teste de ajustamento do qui-quadrado é por vezes necessária o agrupamento de classes adjacentes e refazer parcialmente os cálculos. Recorreremos mais uma vez a um exemplo para ilustrar esta situação, na sub-secção 8.11.3.

Por fim, ilustrar-se-á o recurso a classes equiprováveis no teste de ajustamento de qui-quadrado, na sub-secção 8.11.5.[11]

[11] Ainda à laia de conclusão refira-se que há outros testes de ajustamento que não requerem tantas observações como é o caso do teste de ajustamento de Kolmogorov-Smirnov que não faz parte do programa da disciplina.

8.11.1 Ajustamento de uma distribuição discreta

É curioso notar que a frequência observada da classe i é a concretização da v.a.

$$O_i \sim \text{binomial}(n, p_i), \tag{8.65}$$

onde n representa o total de observações e p_i a probabilidade de uma observação escolhida ao acaso pertencer à classe i (justificação!).

Caso a hipótese nula seja simples, somos capazes de calcular o valor da probabilidade p_i sob a validade de H_0, valor esse que será doravante representado por p_i^0. Somos igualmente capazes de calcular o valor esperado de O_i sob H_0, pelo que igualaremos a frequência esperada sob H_0 da classe i, E_i, a este valor esperado. Assim,

$$E_i = E(O_i|H_0) = E[\text{binomial}(n, p_i^0)] = n \times p_i^0. \tag{8.66}$$

Exemplo 8.48 — Teste de ajustamento do qui-quadrado: dados discretos (hipótese simples — uniforme discreta)
Um dado é lançado 1000 vezes tendo conduzido à tabela de frequências observadas a seguir.

Resultado	Freq. Obs.
1	174
2	174
3	154
4	179
5	154
6	165

Coloca-se, naturalmente, a questão: *Será o dado equilibrado/perfeito*? Resta-nos responder a esta questão, considerando para o efeito um nível de significância de, por exemplo, 5%.

- **V.a. de interesse**

  $X =$ resultado do lançamento do dado

- **Hipóteses**

  $H_0 : X \sim$ uniforme discreta$(\{1, \ldots, 6\})$ vs.
  $H_1 : X \not\sim$ uniforme discreta$(\{1, \ldots, 6\})$ [12]

[12] A hipótese nula é simples dado que em H_0 se conjectura uma única distribuição completamente definida.

Importa notar que, ao considerarmos $p_i = P(X = i), i = 1, \ldots, 6$, as hipóteses nula e alternativa podem reescrever-se do seguinte modo:

$H_0 : p_i = p_i^0 = 1/6, i = 1, \ldots, 6$ vs.
$H_1 : \exists i : p_i \neq p_i^0$.

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \stackrel{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)} \tag{8.67}$$

onde:

$k =$ No. de classes em que estão organizados os dados da tabela de frequências $= 6$

$O_i =$ Frequência absoluta observável da classe i

$E_i =$ Frequência absoluta esperada, sob H_0, da classe i
$= E(O_i|H_0) = np_i^0$ [13]

$\beta =$ No. de parâmetros a estimar $= 0$. [14]

Escusado será dizer que as frequências absolutas esperadas sob H_0 serão iguais ao valor esperado de O_i sob H_0. Com efeito,

$$E_i = n \times p_i^0 = 1000 \times 1/6, \; i = 1, \ldots, 6. \tag{8.68}$$

De referir que estas frequências são dependentes de i, salvo em raríssimas excepções como é o caso deste exemplo. O exemplo seguinte ilustrará tal facto.

- **Região de rejeição de** H_0 (para valores de T)

  Quanto maior a discrepância entre a frequência absoluta observada da classe i (o_i) e o valor conhecido da frequência absoluta esperada sob H_0 da classe i (E_i), maior é o valor observado da estatística de teste e menos consistente é a hipótese H_0 com os dados recolhidos. Logo a região de rejeição de H_0 (para valores de T) é um intervalo à direita

[13] Estas frequências absolutas esperadas serão igualadas a $n \times p_i^0$, onde $p_i^0 = 1/6$ uma vez que estamos a averiguar o ajustamento da distribuição uniforme discreta com espaço de resultados $\{1, \ldots, 6\}$.

[14] Uma vez que H_0 é uma hipótese simples não é necessário estimar qualquer parâmetro desconhecido.

$$W = (c, +\infty), \tag{8.69}$$

onde $c : P(\text{Rejeitar } H_0|H_0) \simeq \alpha_0$, i.e.,

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(6-0-1)}}(1-0.05) = 11.07. \tag{8.70}$$

- **Decisão**

Para calcular o valor observado da estatística de teste é necessário construir uma tabela de frequências onde deverão constar, entre muitos outros valores, as frequências absolutas observadas e os valores conhecidos das frequências absolutas esperadas sob H_0.

i	Classe i	Freq. abs. observada o_i	Freq. abs. esperada sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$\{1\}$	174	$1000 \times \frac{1}{6} = 166.(6)$	$\frac{[174-166.(6)]^2}{166.(6)} = 0.322$
2	$\{2\}$	174	"	0.322
3	$\{3\}$	154	"	0.963
4	$\{4\}$	179	"	0.912
5	$\{5\}$	154	"	0.963
6	$\{6\}$	165	"	$\frac{[165-166.(6)]^2}{166.(6)} = 0.017$
		$\sum_{i=1}^k o_i = n$ $= 1000$	$\sum_{i=1}^k E_i = n$ $= 1000.00$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i}$ $= 3.499$

Então o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \\ &= 0.322 + 0.322 + 0.963 + 0.912 + 0.963 + 0.017 \\ &= 3.499 \\ &\notin W = (11.07, +\infty). \end{aligned} \tag{8.71}$$

Consequentemente não devemos rejeitar a hipótese de estarmos a lidar com um dado equilibrado, ao n.s. 5%, nem a qualquer outro n.s. menor que 5%. •

Exemplo 8.49 — Teste de ajustamento do qui-quadrado: dados discretos (hipótese simples — distribuição de Poisson)

O departamento de defesa pretende saber qual a distribuição de probabilidade do número de avarias, durante uma dada missão, ocorridas numa determinada zona de um submarino. Com esse objectivo foram recolhidos dados relativos a 500 destas missões condensados na tabela abaixo.

No. de falhas por missão	0	1	2	3	4
No. de missões (com tal no. de falhas)	185	180	95	30	10

Teste ao nível de significância de 5% a hipótese de os dados seguirem uma distribuição de Poisson com valor esperado igual a 1.

- **V.a. de interesse**

  $X =$ número de falhas em uma missão do submarino

- **Hipóteses**

  $H_0 : X \sim \text{Poisson}(1)$ vs. $H_1 : X \not\sim \text{Poisson}(1)$ [15]

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \overset{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)}, \tag{8.72}$$

  onde:

  $k =$ No. de classes $= 5$

  $O_i =$ Frequência absoluta observável da classe i

  $E_i =$ Frequência absoluta esperada, sob H_0, da classe i [16]

  $\beta =$ No. de parâmetros a estimar $= 0$.[17]

[15] A hipótese nula é simples já que em H_0 se conjectura uma única distribuição e não uma família de distribuições como seria o caso de $H_0 : X \sim \text{Poisson}(\lambda)$ onde λ é desconhecido.

[16] Estas frequências serão igualadas a valores de fácil cálculo já que estamos a averiguar o ajustamento da distribuição de Poisson com valor esperado igual a 1.

[17] Dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.

- **Região de rejeição de H_0** (para valores de T)

  Pelo mesmo motivo apontado no exemplo anterior, a região de rejeição de H_0 escrita para valores de T é um intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(5-0-1)}}(1-0.05) = 9.488. \tag{8.73}$$

- **Decisão**

  O cálculo do valor observado da estatística de teste pressupõe, como anteriormente, a construção de uma tabela auxiliar de frequências de vário tipo.

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$\{0\}$	185	$500 \times 0.3679 = 183.95$	$\frac{(185-183.95)^2}{183.95} = 0.0060$
2	$\{1\}$	180	$500 \times 0.3679 = 183.95$	0.0848
3	$\{2\}$	95	$500 \times 0.1839 = 91.95$	0.1012
4	$\{3\}$	30	$500 \times 0.0613 = 30.65$	0.0138
5	$\{4,5,\ldots\}$	10	$500 \times 0.0190 = 9.50$	0.0263
		$\sum_{i=1}^k o_i = n$ $= 500$	$\sum_{i=1}^k E_i = n$ $= 500.00$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i}$ $= 0.2321$

Note-se que

$$\begin{aligned}
E_i &= n \times p_i^0 \\
&= n \times P(X \in \text{ Classe } i \mid H_0) \\
&= n \times P[X \in \text{ Classe } i \mid X \sim \text{Poisson}(1)] \\
&= n \times \begin{cases} P[X = i-1 \mid X \sim \text{Poisson}(1)], & i = 1,2,3,4 \\ P[X \geq 4 \mid X \sim \text{Poisson}(1)], & i = 5 \end{cases} \\
&= n \times \begin{cases} \frac{e^{-\lambda}\times\lambda^{i-1}}{(i-1)!} = \frac{e^{-1}\times 1^{i-1}}{(i-1)!}, & i = 1,2,3,4 \\ 1 - (p_1^0 + \ldots p_4^0), & i = 5. \end{cases}
\end{aligned} \tag{8.74}$$

Assim, temos

$$\begin{aligned}
t &= \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \\
&= 0.0060 + 0.0848 + 0.1012 + 0.0138 + 0.0263 \\
&= 0.2321 \\
&\notin W = (9.488, +\infty),
\end{aligned} \tag{8.75}$$

pelo que não devemos rejeitar a hipótese de os dados serem provenientes de uma população com distribuição de Poisson com parâmetro conhecido e igual a 1, a qualquer nível de significância menor ou igual que 5%. •

8.11.2 Ajustamento de uma família de distribuições discretas

É altura de considerar o caso em que se conjectura não uma distribuição específica para a nossa v.a. de interesse mas sim um modelo ou família de distribuições. Basta pensar, por exemplo, no modelo de Poisson de parâmetro desconhecido λ .

Tratando-se H_0 de uma hipótese composta, só podemos adiantar uma expressão para $n \times p_i^0$, expressão essa que depende de parâmetro(s) desconhecido(s). Assim sendo, $n \times p_i^0$ não pode figurar na expressão da estatística de teste,[18] pelo que só nos resta considerar que E_i passe a ser um estimador de $n \times p_i^0$.[19]

Retomaremos o Exemplo 8.49 por forma a deixar claras as diferenças entre os testes de ajustamento do qui-quadrado com hipóteses nulas simples e composta.

Exemplo 8.50 — Teste de ajustamento do qui-quadrado: dados discretos (hipótese composta — modelo de Poisson)

Voltemos a considerar o problema do número de avarias durante missões do submarino do Exemplo 8.49 e testemos agora a adequação do modelo de Poisson — família de todas as distribuições de Poisson — a este mesmo conjunto de dados e ao mesmo nível de significância, 5%.

- **Hipóteses**

  $H_0 : X \sim \text{Poisson}(\lambda)$ vs. $H_1 : X \not\sim \text{Poisson}(\lambda)$

  H_0 é uma hipótese composta dado que λ é uma constante positiva desconhecida.

- **Nível de significância**

  $\alpha_0 = 5\%$

[18] As estatísticas, por definição, não podem depender de quaisquer parâmetros desconhecidos.

[19] Convém referir que esta substituição conduz a uma alteração da distribuição assintótica da estatística de teste. Caso as estimativas de MV do(s) β parâmetro(s) desconhecido(s) seja(m) calculada(s), como com base nos dados agrupados em classes, a estatística de teste possui distribuição do qui-quadrado com $k - \beta - 1$ graus de liberdade. Caso as estimativas de MV do(s) parâmetro(s) desconhecido(s) seja(m) calculada(s), como, aliás, é costume, com base nos dados originais e não nos dados agrupados em classes, a cauda direita da f.d.p. assintótica da estatística de teste está entre as caudas direitas das f.d.p. das distribuições do qui-quadrado com $k - 1$ e $k - \beta - 1$ graus de liberdade. No entanto, iremos considerar (incorrectamente) que, em qualquer dos dois casos, a distribuição assintótica da estatística de teste é do qui-quadrado com $k - \beta - 1$ graus de liberdade; ao fazê-lo, estamos a lidar com um teste que conduzirá a uma percentagem superior de rejeições indevidas de H_0 no primeir dos dois casos. Para mais detalhes recomenda-se a leitura de Paulino (1992, pp. 54–56).

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \overset{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)}. \tag{8.76}$$

Assinale-se somente o que distingue esta estatística de teste daquela que figura em (8.72):

E_i = Estimador da frequência absoluta esperada, sob H_0, da classe i

β = No. de parâmetros a estimar = 1.[20]

- **Estimação de λ**

A estimativa de MV de λ é dada por

$$\hat{\lambda} = \frac{185 \times 0 + 180 \times 1 + 95 \times 2 + 30 \times 3 + 10 \times 4}{500} = 1, \tag{8.77}$$

por sinal igual ao valor considerado no Exemplo 8.49.

- **Região de rejeição de H_0** (para valores de T)

Desta feita lidamos com o intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1 - \alpha_0) = F^{-1}_{\chi^2_{(5-1-1)}}(1 - 0.05) = 7.815. \tag{8.78}$$

- **Decisão**

Como λ é desconhecido o mesmo acontece com p_i^0 e com a frequência absoluta esperada, sob H_0, da classe i,

$$n \times p_i^0 \quad = \quad n \times \begin{cases} \frac{e^{-\lambda} \times \lambda^{i-1}}{(i-1)!}, & i = 1,2,3,4 \\ 1 - (p_1^0 + \ldots p_4^0), & i = 5. \end{cases} \tag{8.79}$$

Mas uma vez que $\hat{\lambda} = 1$ as estimativas das frequências absolutas esperadas sob H_0 são iguais a

$$e_i = n \times \hat{p}_i^0 \quad = \quad n \times \begin{cases} \frac{e^{-\hat{\lambda}} \times \hat{\lambda}^{i-1}}{(i-1)!} = \frac{e^{-1} \times 1^{i-1}}{(i-1)!}, & i = 1,2,3,4 \\ 1 - (\hat{p}_1^0 + \ldots \hat{p}_4^0), & i = 5. \end{cases} \tag{8.80}$$

De realçar que estes dados foram claramente manipulados por forma a evitar mais cálculos na obtenção do habitual quadro usado no cálculo do valor observado da estatística de teste. Mas atente-se a algumas diferenças — subtis é certo (identifique-as!) — entre o quadro que se segue e o do Exemplo 8.49.

[20] É necessário estimar λ para efectuar este teste.

i	Classe i	Freq. abs. obs. o_i	Estim. freq. abs. esp. sob H_0 $e_i = n \times \hat{p}_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$\{0\}$	185	$500 \times 0.3679 = 183.95$	$\frac{(185-183.95)^2}{183.95} = 0.0060$
2	$\{1\}$	180	$500 \times 0.3679 = 183.95$	0.0848
3	$\{2\}$	95	$500 \times 0.1839 = 91.95$	0.1012
4	$\{3\}$	30	$500 \times 0.0613 = 30.65$	0.0138
5	$\{4, 5, \ldots\}$	10	$500 \times 0.0190 = 9.50$	0.0263
		$\sum_{i=1}^k o_i = n$ $= 500$	$\sum_{i=1}^k e_i = n$ $= 500.00$	$t = \sum_{i=1}^k \frac{(o_i-e_i)^2}{e_i}$ $= 0.2321$

Posto isto, obtemos o mesmo valor observado da estatística de teste:

$$t = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = 0.2321 \notin W = (7.815, +\infty). \tag{8.81}$$

Conclui-se então que a hipótese dos dados serem provenientes de uma população com distribuição pertencente ao modelo de Poisson é razoável a qualquer nível de significância menor ou igual a 5%. •

8.11.3 Agrupamento de classes

Uma vez que o teste de ajustamento do qui-quadrado se baseia na convergência em distribuição de binomiais (O_i) para normais afigura-se razoável exigir que as frequências esperadas sob H_0, E_i, não sejam demasiado pequenas, em particular exigir que sejam superiores ou iguais a 5.[21]

Deste modo, se registarmos, para algum i,

- $E_i < 5$

devemos agrupar esta classe à classe adjacente com menor frequência absoluta esperada sob H_0.

Não há, contudo, uma base teórica para tal regra, havendo mesmo estudos que indicam que pode registar-se uma ou duas frequências esperadas sob H_0 inferiores a 5, sem que seja violada a distribuição assintótica, como refere Paulino (1992, p. 52). Com efeito, há autores que sugerem que não há a necessidade de qualquer agrupamento de classes se:

[21] Recorde-se que: é suposto aproximar a v.a. $X \sim \text{binomial}(n, p)$ pela v.a. $\tilde{X} \sim \text{normal}(np, np(1-p))$ quando $np > 5$ e $n(1-p) > 5$; o quadrado de uma normal-padrão tem distribuição do qui-quadrado...

- em pelo menos 80% das classes se verificar $E_i \geq 5$ e
- nas restantes classes $E_i \geq 1$.

E será este o critério que usaremos doravante para não agrupar classes.

Aproveitaremos um exemplo de Montgomery e Runger (2003, pp. 316–8) para ilustrar não só o agrupamento de classes no teste de ajustamento do qui-quadrado mas também o cálculo de intervalo aproximado para o p-value no referido teste. A apresentação deste exemplo será acompanhada por poucos comentários já que este exemplo é similar ao Exemplo 8.50.

Exemplo 8.51 — Teste de ajustamento do qui-quadrado: dados discretos (hipótese composta, agrupamento de classes, p-value)

Alguns estudos prévios levam a crer que o número de defeitos em circuitos de determinado tipo possui uma distribuição de Poisson. Os resultados na tabela seguinte dizem respeito a uma amostra de 60 circuitos.

No. defeitos	Freq. Obs.
0	32
1	15
2	9
3	4

Testemos a adequação do modelo de Poisson a este conjunto de dados calculando para o efeito um intervalo (aproximado) para o p-value.

- **V.a. de interesse**

  X = número de defeitos em circuitos de determinado tipo

- **Hipóteses**

  $H_0 : X \sim \text{Poisson}(\lambda)$ vs. $H_1 : X \not\sim \text{Poisson}(\lambda)$[22]

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \stackrel{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)}. \tag{8.82}$$

- **Estimação de λ**

  A estimativa de MV de λ é igual a

$$\hat{\lambda} = \frac{32 \times 0 + 15 \times 1 + 9 \times 2 + 4 \times 3}{60} = 0.75. \tag{8.83}$$

[22] H_0 é uma hipótese composta dado que λ é uma constante positiva desconhecida.

- **Estimação das probabilidades de pertença e das frequências absolutas esperadas sob H_0**

  As estimativas das probabilidade de pertença a cada uma das classes sob H_0 são dadas por:

$$\begin{aligned}
\hat{p}_1^0 &= \hat{P}(X=0|H_0) = \frac{e^{-\hat{\lambda}} \times \hat{\lambda}^0}{0!} = 0.472 \\
\hat{p}_2^0 &= \hat{P}(X=1|H_0) = \frac{e^{-\hat{\lambda}} \times \hat{\lambda}^1}{1!} = 0.354 \\
\hat{p}_3^0 &= \hat{P}(X=2|H_0) = \frac{e^{-\hat{\lambda}} \times \hat{\lambda}^2}{2!} = 0.133 \\
\hat{p}_4^0 &= \hat{P}(X \geq 3|H_0) = 1 - (\hat{p}_1^0 + \hat{p}_2^0 + \hat{p}_3^0) = 0.041.
\end{aligned}$$

  Consequentemente, obtemos as seguintes estimativas das frequências absolutas esperadas sob H_0:

$$\begin{aligned}
e_1 &= n \times \hat{p}_1^0 = 28.32 \\
e_2 &= n \times \hat{p}_2^0 = 21.24 \\
e_3 &= n \times \hat{p}_3^0 = 7.98 \\
e_4 &= n \times \hat{p}_4^0 = 2.46.
\end{aligned}$$

- **Agrupamento de classes e decisão**

  Tendo em conta que a 4a. das 4 classes possui $e_i < 5$ — e como tal menos de 80% das classes possuem $e_i \geq 5$ — deve proceder-se ao agrupamento da 4a. classe àquela que é a sua única classe adjacente, a 3a. classe. Passamos a

  - dispor de $k = 3$ classes $(\{0\}, \{1\}, \{2, 3, \ldots\})$

  ao invés das 4 iniciais e a lidar com o seguinte quadro de frequências:

i	Classe i	Freq. abs. obs. o_i	Estim. freq. abs. esp. sob H_0 $e_i = n \times \hat{p}_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$\{0\}$	32	28.32	$\frac{(32-28.32)^2}{28.32} = 0.478$
2	$\{1\}$	15	21.24	$\frac{(15-21.24)^2}{21.24} = 1.833$
3	$\{2, 3, \ldots\}$	$\mathbf{9+4} = 13$	$\mathbf{7.98+2.46} = 10.44$	$\frac{(13-10.44)^2}{10.44} = 0.628$
		$\sum_{i=1}^k o_i = n = 60$	$\sum_{i=1}^k e_i = n = 60$	$t = \sum_{i=1}^k \frac{(o_i-e_i)^2}{e_i} = 2.939$

- **Intervalo aproximado para o p-value**

  Tratando-se de teste associado a uma região de rejeição de H_0 que é um intervalo à direita temos

$$\begin{aligned} p-value &= P(T > t \mid H_0) \\ &= 1 - F_{T|H_0}(t) \\ &\simeq 1 - F_{\chi^2_{(3-1-1)}}(2.939). \end{aligned} \tag{8.84}$$

  Recorrendo às tabelas de quantis da distribuição do qui-quadrado com um grau de liberdade podemos obter um intervalo (aproximado) para o p-value deste teste. Para tal basta enquadrar convenientemente $t = 2.939$ por dois quantis, obtendo-se sucessivamente

$$\begin{aligned} F^{-1}_{\chi^2_{(1)}}(0.900) = 2.706 &< t = 2.939 < 3.170 = F^{-1}_{\chi^2_{(1)}}(0.925) \\ 0.900 &< F_{\chi^2_{(1)}}(2.939) < 0.925 \\ 1 - 0.925 &< 1 - F_{\chi^2_{(1)}}(2.939) < 1 - 0.900 \end{aligned}$$

  e por fim o intervalo aproximado para o p-value: (0.075,0.10). Deste modo, conclui-se que:

  - não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 7.5\%$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 10\%$. •

8.11.4 Dados contínuos — hipótese simples/composta

Para v.a. contínuas o procedimento de teste é análogo: as observações devem ser previamente organizadas em classes, i.e., em intervalos disjuntos que cubram todo o contradomínio da v.a. de interesse.

Por exemplo, caso estivéssemos a testar a adequação da distribuição normal(0,1) as classes seriam do tipo:

- Classe 1 — $(-\infty, a_1]$
- Classe 2 — $(a_1, a_2]$

  ...

- Classe k — $(a_{k-1}, +\infty)$.

Exemplo 8.52 — Teste de ajustamento do qui-quadrado: dados contínuos (hipótese simples/composta)

(a) Um engenheiro admite que a duração de certa marca de lâmpadas eléctricas é uma v.a. com distribuição exponencial de parâmetro $\lambda = 0.05$. Numa experiência laboratorial envolvendo 100 dessas lâmpadas obtiveram-se as seguintes durações resumidas abaixo:

Classe	[0, 10)	[10, 20)	[20, 30)	$[30, +\infty)$
Frequência	30	27	23	20

Com base nos dados fornecidos, teste a hipótese considerada pelo engenheiro ao nível de significância de 5%. Obtenha e discuta o intervalo aproximado para o valor–p associado à amostra considerada.[23]

- **Hipóteses**

  $H_0 : X \sim$ exponencial(0.05) vs. $H_1 : X \not\sim$ exponencial(0.05)

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \stackrel{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)},$$

  onde:

  $k =$ No. de classes $= 4$

  $O_i =$ Frequência absoluta observável da classe i

  $E_i =$ Frequência absoluta esperada, sob H_0, da classe i

  $\beta =$ No. de parâmetros a estimar $= 0$.

- **Região de rejeição de H_0** (para valores de T)

  É um intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1 - \alpha_0) = F^{-1}_{\chi^2_{(4-0-1)}}(1 - 0.05) = 7.815.$$

[23] Adaptado do Exame de 24 de Junho de 2006.

- **Decisão**

  Tirando partido do facto de

$$\begin{aligned} F_{X|H_0}(x) &= P[X \leq x \mid \text{exponencial}(0.05)] \\ &= 1 - e^{-0.05x},\ x \geq 0, \end{aligned}$$

  conclui-se que as frequências absolutas esperadas sob H_0 são iguais a

$$\begin{aligned} p_1^0 &= P(0 \leq X < 10|H_0) \\ &= F_{X|H_0}(10) - F_{X|H_0}(0) = 0.39347 - 0 = 0.39347 \\ p_2^0 &= P(10 \leq X < 20|H_0) \\ &= F_{X|H_0}(20) - F_{X|H_0}(10) = 0.63212 - 0.39347 = 0.23865 \\ p_3^0 &= P(20 \leq X < 30|H_0) \\ &= F_{X|H_0}(30) - F_{X|H_0}(20) = 0.77687 - 0.63212 = 0.14475 \\ p_4^0 &= P(X \geq 30|H_0) \\ &= 1 - F_{X|H_0}(30) = 1 - 0.77687 = 0.22313 \end{aligned}$$

  Consequentemente, obtemos o seguinte quadro

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$[0, 10)$	30	$100 \times 0.39347 = 39.347$	$\frac{(30-39.347)^2}{39.347} = 2.2204$
2	$[10, 20)$	27	$100 \times 0.23865 = 23.865$	0.4118
3	$[20, 30)$	23	$100 \times 0.14475 = 14.475$	5.0208
4	$[30, +\infty)$	20	$100 \times 0.22313 = 22.313$	0.2398
		$\sum_{i=1}^k o_i = n$ $= 100$	$\sum_{i=1}^k E_i = n$ $= 100.00$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i}$ $= 7.8928$

  e o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= 7.8928 \\ &\in W = (7.815, +\infty), \end{aligned}$$

  pelo que devemos rejeitar a hipótese de ajustamento da distribuição de exponencial com parâmetro 0.05 ao conjunto de dados, a qualquer nível de significância maior ou igual que 5%.

- **Intervalo aproximado para o p-value**

  Uma vez que a região de rejeição de H_0 é um intervalo à direita segue-se:

$$\begin{aligned} p - value &= P(T > t \mid H_0) \\ &= 1 - F_{T|H_0}(t) \\ &\simeq 1 - F_{\chi^2_{(4-0-1)}}(7.8928). \end{aligned}$$

Ora, ao recorrer às tabelas de quantis da distribuição do qui-quadrado com três graus de liberdade podemos obter um intervalo (aproximado) para o p-value deste teste enquadrando convenientemente $t = 7.8928$ por dois quantis:

$$\begin{aligned}
F^{-1}_{\chi^2_{(3)}}(0.95) = 7.815 &< t = 7.8928 < 9.348 = F^{-1}_{\chi^2_{(3)}}(0.975) \\
0.95 &< F_{\chi^2_{(3)}}(7.8928) < 0.975 \\
1 - 0.975 &< 1 - F_{\chi^2_{(3)}}(7.8928) < 1 - 0.95.
\end{aligned}$$

Deste modo conclui-se que o intervalo aproximado para o p-value é (0.025,0.05) e que:

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 2.5\%$;
- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 5\%$.

(b) Um engenheiro admite que a duração de certa marca de lâmpadas eléctricas é uma v.a. com distribuição pertencente ao modelo $\{\text{Exponencial}(\lambda),\ \lambda \in I\!R^+\}$. Numa experiência laboratorial envolvendo 100 dessas lâmpadas obtiveram-se as seguintes durações resumidas abaixo:

Classe	$[0,10)$	$[10,20)$	$[20,30)$	$[30,+\infty)$
Frequência	30	27	23	20

Com base nos dados fornecidos, teste a hipótese considerada pelo engenheiro ao nível de significância de 5%, sabendo que a estimativa de MV calculada numericamente com base nos dados agrupados é igual a 0.047094.

- **Hipóteses**

  $H_0 : X \sim \text{Exponencial}(\lambda),\ \lambda > 0$ DESCONHECIDO vs. $H_1 : X \not\sim \text{Exponencial}(\lambda)$[24]

- **Estatística de Teste**

  $$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)}$$

  $k =$ No. de classes $= 4$

  $O_i =$ Frequência absoluta observável da classe i

  $E_i =$ ESTIMADOR da frequência absoluta esperada, sob H_0, da classe i

  $\beta =$ No. de parâmetros a estimar $= 1$.

[24] A hipótese nula é COMPOSTA já que em H_0 se conjectura uma família de distribuições.

- **Cálculo das estimativas das frequências absolutas observadas sob H_0**

  Como λ é desconhecido o mesmo acontece com

  $$F_{X|H_0}(x) = P[X \leq x \mid \text{Exponencial}(\lambda)] = 1 - e^{-\lambda\, x},\ x \geq 0,$$

  com p_i^0 e com a frequência absoluta esperada, sob H_0, da classe i,

  $$\begin{aligned}
  n \times p_i^0 &= n \times \begin{cases} P(0 < X \leq 10|H_0) = F_{X|H_0}(10) - F_{X|H_0}(0), & i=1 \\ P(10 < X \leq 20|H_0) = F_{X|H_0}(20) - F_{X|H_0}(10), & i=2 \\ P(20 < X \leq 30|H_0) = F_{X|H_0}(30) - F_{X|H_0}(20), & i=3 \\ P(X > 30|H_0) = 1 - F_{X|H_0}(30), & i=4 \end{cases} \\
  &= \begin{cases} 1 - e^{-10\,\lambda}, & i=1 \\ e^{-10\,\lambda} - e^{-20\,\lambda}, & i=2 \\ e^{-20\,\lambda} - e^{-30\,\lambda}, & i=3 \\ e^{-30\,\lambda}, & i=4. \end{cases}
  \end{aligned}$$

  Mas uma vez que a estimativa de MV de λ foi avaliada numericamente com base na amostra agrupada e é igual a $\hat{\lambda} = 0.047094$, podemos obter as estimativas das frequências absolutas esperadas sob H_0, tirando partido das expressões acima e da propriedade de invariância dos estimadores de MV:

  $$\begin{aligned}
  e_i &= n \times \hat{p}_i^0 \\
  &= n \times \begin{cases} 1 - e^{-10\,\hat{\lambda}} = 1 - e^{-10 \times 0.047094}, & i=1 \\ e^{-10\,\hat{\lambda}} - e^{-20\,\lambda} = e^{-10 \times 0.047094} - e^{-20 \times 0.047094}, & i=2 \\ e^{-20\,\hat{\lambda}} - e^{-30\,\lambda} = e^{-20 \times 0.047094} - e^{-30 \times 0.047094}, & i=3 \\ e^{-30\,\hat{\lambda}} = e^{-30 \times 0.047094}, & i=4. \end{cases} \\
  &= \begin{cases} 100 \times 0.375587 = 37.5587, & i=1 \\ 100 \times 0.234521 = 23.4521, & i=2 \\ 100 \times 0.146438 = 14.6438, & i=3 \\ 100 \times 0.243454 = 24.3454, & i=4. \end{cases}
  \end{aligned}$$

  Constata-se que não é necessário agrupar qualquer classe já que em pelo menos 80% delas $e_i \geq 5$ e nas restantes $e_i \geq 1$.

- **Região de rejeição de H_0** (para valores de T)

  Lidamos com um intervalo à direita $W = (c, +\infty)$, onde

  $$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(4-1-1)}}(1-0.05) = F^{-1}_{\chi^2_{(2)}}(0.95) \overset{tabela}{=} 5.991.$$

- **Decisão**

  Tendo em conta que

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $e_i = n \times \hat{p}_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$(0, 10]$	30	$100 \times 0.375587 = 37.5587$	$\frac{(30-37.5587)^2}{37.5587} = 1.52119$
2	$(10, 20]$	27	$100 \times 0.234521 = 23.4521$	0.536724
3	$(20, 30]$	23	$100 \times 0.146438 = 14.6438$	4.76827
4	$(30, +\infty)$	20	$100 \times 0.243454 = 24.3454$	0.775594
		$\sum_{i=1}^k o_i = n$ $= 100$	$\sum_{i=1}^k e_i = n$ $= 100.00$	$t = \sum_{i=1}^k \frac{(o_i-e_i)^2}{e_i}$ $= 7.60177$

o valor observado da estatística de teste é dado por

$$\begin{aligned} t &= 7.60177 \\ &\in W = (5.991, +\infty), \end{aligned}$$

pelo que devemos rejeitar a hipótese de os dados serem provenientes do modelo exponencial, a qualquer nível de significância maior ou igual que 5%. •

Terminamos esta sub-secção recomendando a leitura do Exemplo 9.13 de Montgomery e Runger (2003, pp. 318–319) e remetendo o/a leitor/a para a sub-secção seguinte onde se encontra um exercício com dados contínuos.

8.11.5 Classes equiprováveis e dados contínuos

Ao lidarmos com dados contínuos, é costume tirar partido do agrupamento efectuado na análise descritiva, usualmente com classes com a mesma amplitude. Este procedimento pode conduzir ao agrupamento de classes adjacentes.

Por forma a evitar este inconveniente e sistematizar a escolha de classes, deve adoptar-se o que usualmente se designa por classes equiprováveis sob H_0, i.e., os extremos das k classes são escolhidos de forma que

$$p_i^0 = P(X \in \text{Classe } i \mid H_0) = 1/k, i = 1, \ldots, k, \tag{8.85}$$

onde $k : n \times p_i^0 = \frac{n}{k} \geq 5$, i.e., o número de classes deverá satisfazer $k \leq \frac{n}{5}$. Para tal basta que se considere os seguintes extremos:

- a_0 = limite inferior do contradomínio da distribuição conjecturada em H_0 para a v.a. X;

- $a_i = F^{-1}_{X|H_0}(i/k)$ = quantil de probabilidade i/k da distribuição conjecturada em H_0 para a v.a. de interesse X, $i = 1, \ldots, k-1$;

- a_k = limite superior do contradomínio da distribuição conjecturada em H_0 para a v.a. X.

Este procedimento tem ainda a particularidade de simplificar o cálculo do valor observado da estatística de teste dado que, neste caso, $E_i = n/k$ e, consequentemente, a estatística passa a escrever-se

$$\begin{aligned} T &= \sum_{i=1}^{k} \frac{(O_i - n/k)^2}{n/k} \\ &= \frac{k}{n} \sum_{i=1}^{k} O_i^2 - n. \end{aligned} \tag{8.86}$$

Com efeito:

$$\begin{aligned} T &= \sum_{i=1}^{k} \frac{(O_i - n/k)^2}{n/k} \\ &= \frac{k}{n} \sum_{i=1}^{k} \left[O_i^2 - 2\frac{n}{k} O_i + \left(\frac{n}{k}\right)^2 \right] \\ &= \frac{k}{n} \left[\sum_{i=1}^{k} O_i^2 - 2\frac{n}{k} \sum_{i=1}^{k} O_i + \sum_{i=1}^{k} \left(\frac{n}{k}\right)^2 \right] \\ &= \frac{k}{n} \left[\sum_{i=1}^{k} O_i^2 - 2\frac{n}{k} \times n + k \times \left(\frac{n}{k}\right)^2 \right] \\ &= \frac{k}{n} \sum_{i=1}^{k} O_i^2 - 2n + n \\ &= \frac{k}{n} \sum_{i=1}^{k} O_i^2 - n. \end{aligned}$$

Exemplo 8.53 — Teste de ajustamento do qui-quadrado: dados contínuos (hipótese simples, classes equiprováveis)

Com o objectivo de estudar o tempo até falha de certo equipamento electrónico (em milhares de horas), X, foram recolhidas e ordenadas 50 observações na tabela seguinte. Dada a natureza dos dados e alguns estudos prévios, suspeita-se que as observações sejam provenientes de uma população X com distribuição de Pareto com parâmetros 2.001 e 2.822, i.e., com f.d.

$$1 - \frac{2.001^{2.822}}{x^{2.822}},\ x \geq 2.001.$$

2.001	2.007	2.017	2.026	2.036	2.075	2.077	2.082	2.101	2.137
2.156	2.161	2.181	2.196	2.214	2.227	2.320	2.367	2.424	2.443
2.444	2.449	2.478	2.520	2.579	2.581	2.598	2.637	2.691	2.715
2.720	2.825	2.863	2.867	3.016	3.176	3.360	3.413	3.567	3.721
3.727	3.769	3.803	4.329	4.420	4.795	6.009	6.281	6.784	8.305

(a) Prove que as classes $[2.001, 2.1656]$, $(2.1656, 2.3981]$, $(2.3981, 2.7686]$, $(2.7686, 3.5394]$ e $(3.5394, +\infty)$ são equiprováveis sob a hipótese $H_0 : X \sim \text{Pareto}(2.001, 2.822)$. De facto, para $i = 1, \ldots, k$ e $k = 5$:

$$\begin{aligned}
p_1^0 &= P(X \in [2.001, 2.1656] \mid H_0) \\
&= F_{X|H_0}(2.1656) - F_{X|H_0}(2.001) \\
&= \left(1 - \frac{2.001^{2.822}}{2.1656^{2.822}}\right) - \left(1 - \frac{2.001^{2.822}}{2.001^{2.822}}\right) \\
&= 0.2000 \qquad (8.87)\\
&= \frac{1}{k} \qquad (8.88)\\
&\ \vdots
\end{aligned}$$

(b) Averigue a adequação da distribuição de *Pareto* com os referidos parâmetros ao conjunto de dados, ao nível de significância de 10%.

- **Hipóteses**

  $H_0 : X \sim \text{Pareto}(2.001, 2.822)$ vs. $H_1 : X \not\sim \text{Pareto}(2.001, 2.822)$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \overset{a}{\sim}_{H_0} \quad \chi^2_{(k-\beta-1)}, \tag{8.89}$$

onde:

$k =$ No. de classes $= 5$

$O_i =$ Frequência absoluta observável da classe i

$E_i =$ Frequência absoluta esperada, sob H_0, da classe $i = n/k = 10$

$\beta =$ No. de parâmetros a estimar $= 0$.

- **Região de rejeição de H_0** (para valores de T)

É um intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1 - \alpha_0) = F^{-1}_{\chi^2_{(5-0-1)}}(1 - 0.1) = 7.779. \tag{8.90}$$

- **Decisão**

Ao lidarmos com classes equiprováveis sob H_0 basta recorrer a (8.86) para obter o valor observado da estatística de teste. Assim sendo, há somente a necessidade de determinar as frequências absolutas observadas das 5 classes.

Ora, da consulta da tabela com as $n = 50$ observações conclui-se que são iguais a: 12, 6, 13, 7, 12. Logo

$$\begin{aligned}
t &= \sum_{i=1}^{k} \frac{(o_i - n/k)^2}{n/k} \\
&= \frac{k}{n} \sum_{i=1}^{k} o_i^2 - n \\
&= \frac{5}{50}(12^2 + 6^2 + 13^2 + 7^2 + 12^2) - 50 \\
&= 4.2 \\
&\notin W = (7.779, +\infty),
\end{aligned} \tag{8.91}$$

pelo que não devemos rejeitar a hipótese de os dados serem provenientes de uma população com distribuição de Pareto com parâmetros 2.001 e 2.822, a qualquer nível de significância menor ou igual que 10%. •

8.12 Teste de independência do qui-quadrado de Pearson em tabelas de contingência.

Dada a pertinência do teste de independência do qui-quadrado em tabelas de contigência decidimos acrescentá-lo neste capítulo. Limitamo-nos, no entanto, a apresentar um exemplo. Para mais detalhes e considerações teóricas acerca deste teste de hipóteses,[25] remetemos o/a leitor/a mais interessado para Montgomery e Runger (2003, pp. 315-320).

Exemplo 8.54 — Teste de independência do qui-quadrado em tabelas de contingência

Num estudo clínico seleccionaram-se aleatoriamente $n = 1000$ indivíduos que foram classificados segundo o género e a presença ou ausência de daltonismo, obtendo-se os seguintes resultados:

	Masculino	**Feminino**
Daltónic@s	39	6
Não Daltónic@s	461	494

- **Par aleatório de interesse**

  Neste exemplo estamos a lidar com uma tabela de contigência e duas v.a. de interesse. A saber:

$$X = \begin{cases} 1 & \text{se o indivíduo for daltónico} \\ 2 & \text{c.c.} \end{cases} \quad (8.92)$$

$$Y = \begin{cases} 1 & \text{se o indivíduo for do género masculino} \\ 2 & \text{c.c.} \end{cases} \quad (8.93)$$

- **Situação**

  Considere-se, para $i = 1, \ldots, r$ e $j = 1, \ldots, s$ $(r, s = 2)$:

  $p_{ij} = P(X = i, Y = j)$ DESCONHECIDO;

  $p_{i.} = P(X = i) = \sum_{j=1}^{s} p_{ij}$ DESCONHECIDO;

  $p_{.j} = P(Y = j) = \sum_{i=1}^{r} p_{ij}$ DESCONHECIDO.

[25] Por sinal também ele assintótico e como tal requerendo um número suficientemente grande de observações.

- **Hipóteses**

  $H_0 : p_{ij} = p_{i.} \times p_{.j},\ i, j = 1, 2$ [26] vs. $H_1 : \exists (i,j) : p_{ij} \neq p_{i.} \times p_{.j}$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

$$\begin{aligned} T &= \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(O_{ij} - \frac{O_{i.} \times O_{.j}}{n}\right)^2}{\frac{O_{i.} \times O_{.j}}{n}} \ \stackrel{a}{\sim}_{H_0} \ \chi^2_{(r-1)(s-1)}, \end{aligned} \tag{8.94}$$

  onde, para $i = 1, \ldots, r$ e $j = 1, \ldots, s$:

  O_{ij} = Frequência absoluta observável da célula (i, j) da tabela de contingência
  $O_{ij} \sim \text{binomial}(n, p_{ij})$;

  $O_{i.} = \sum_{j=1}^{s} O_{ij}$ = Frequência absoluta observável da linha i da tabela de contingência
  $O_{i.} \sim \text{binomial}(n, p_{i.})$;

  $O_{.j} = \sum_{i=1}^{r} O_{ij}$ = Frequência absoluta observável da coluna j da tabela de contingência
  $O_{.j} \sim \text{binomial}(n, p_{.j})$

  $E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$ = Estimador de $E(O_{ij} \mid H_0)$.[27]

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Quanto maior for a discrepância entre as frequências das células da tabela de frequências (o_{ij}) e a estimativa da frequência absoluta esperada dessa mesma célula sob a validade da hipótese de independência ($o_{i.} \times o_{.j}/n$), mais inconsistente será H_0 com os dados. Logo a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à direita da distribuição assintótica de T sob H_0, i.e., $W = (c, +\infty)$, onde

[26] Recorde-se que duas v.a. discretas X e Y são independentes caso a f.p. conjunta de (X, Y) se escrevear à custa do produtos das f.p. marginais de X e de Y.

[27] De notar que $E(O_{ij} \mid H_0) = E[\text{binomial}(n, p_{ij}) \mid H_0] = n\ p_{i.} \times p_{.j}$, pelo que o estimador natural desta quantidade é, efectivamente, $n \times \frac{O_{i.}}{n} \times \frac{O_{.j}}{n} = \frac{O_{i.} \times O_{.j}}{n}$.

$$c = F^{-1}_{\chi^2_{(r-1)(s-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(2-1)(2-1)}}(1-0.1) = 2.706. \tag{8.95}$$

- **Decisão**

  Tendo em consideração que

o_{ij}	**Masculino**	**Feminino**	$o_{i.}$
Daltónic@s	39	6	45
Não Daltónic@s	461	494	955
$o_{.j}$	500	500	$n = 1000$

o valor observado da estatística de teste é igual a

$$\begin{aligned}
t &= \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(o_{ij}-o_{i.}\times o_{.j}/n)^2}{o_{i.}\times o_{.j}/n} \\
&= \frac{\left(39-\frac{45\times500}{1000}\right)^2}{\frac{45\times500}{1000}} + \frac{\left(6-\frac{45\times500}{1000}\right)^2}{\frac{45\times500}{1000}} \\
&\quad + \frac{\left(461-\frac{955\times500}{1000}\right)^2}{\frac{955\times500}{1000}} + \frac{\left(494-\frac{955\times500}{1000}\right)^2}{\frac{955\times500}{1000}} \\
&= \frac{(39-22.5)^2}{22.5} + \frac{(6-22.5)^2}{22.5} \\
&\quad + \frac{(461-477.5)^2}{477.5} + \frac{(494-477.5)^2}{477.5} \\
&= 25.34 \\
&\in W = (2.706, +\infty).
\end{aligned} \tag{8.96}$$

A presença de daltonismo num indivíduo parece depender do respectivo género a qualquer n.s. superior ou igual a 10%. •

Capítulo 9

Introdução à regressão linear simples

9.1 Modelos de regressão.

Motivação 9.1 — **Modelos de regressão**
Em engenharia é frequente estarmos interessados em estabelecer uma relação entre

- uma variável dependente, Y, e
- uma (ou mais) variável(is) independente(s), x (ou $(x_1, \ldots, x_k)$) .

Esta relação é, de um modo geral, traduzida por um

- modelo de regressão. •

Exemplo 9.2 — **Modelos de regressão**
Considere as variáveis

- Y = pressão atmosférica (variável dependente ou resposta)
- x = altitude (variável independente ou explicativa ou regressora).

Há várias possibilidades de modelos de regressão, entre elas:

1. Modelo determinístico

$$Y = \beta_0 + \beta_1 x \tag{9.1}$$

2. Modelo de regressão linear simples (RLS)

$$Y = \beta_0 + \beta_1 x + \epsilon, \tag{9.2}$$

onde se assume que ϵ é um erro aleatório tal que $E(\epsilon) = 0$.

3. Modelo de regressão linear múltipla

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \tag{9.3}$$

onde x_1 e x_2 representam a altitude e a temperatura (resp.). •

A obtenção de informação no âmbito da RLS passa pela recolha de uma amostra de n pontos

$$(x_i, y_i),\ i = 1, \ldots, n, \tag{9.4}$$

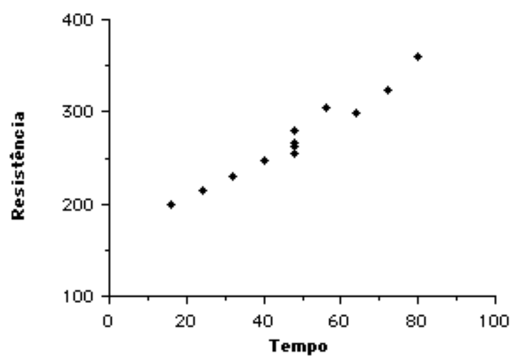
onde podemos entender que

- x_i representa o estímulo a que é submetido o indivíduo i e
- y_i representa a resposta do indivíduo i a esse mesmo estímulo.

Nota 9.3 — **Representação gráfica**
É crucial representar graficamente os pontos $(x_i, y_i), i = 1, \ldots, n$, para averiguar se a relação entre a variável independente/explicativa x e a variável resposta Y é de facto do tipo linear ou se há a necessidade de uma transformação dos dados para que tal ocorra. •

Exemplo 9.4 — **Regressão linear simples**
Pretende estudar-se a relação entre a resistência de um determinado tipo de plástico (Y) e o tempo (em horas) que decorre a partir da conclusão do processo de moldagem até ao momento de medição da resistência (x). Para tal foram testadas 12 peças construídas com esse plástico e obtidas as seguintes observações:

i	1	2	3	4	5	6	7	8	9	10	11	12
Tempo (x_i)	16	24	32	40	48	48	48	48	56	64	72	80
Resistência (y_i)	199	214	230	248	255	262	279	267	305	298	323	359



Apesar de nenhuma curva simples passar exactamente por todos os pontos, há forte indicação no sentido de os pontos do gráfico se dispersarem aleatoriamente em torno de uma recta. •

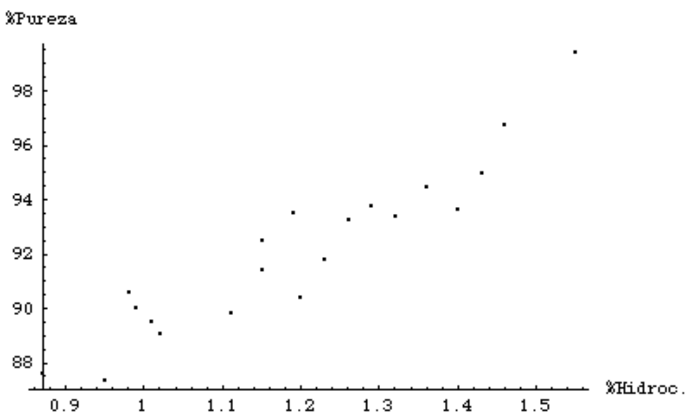
9.2 Métodos dos mínimos quadrados e da máxima verosimilhança em regressão linear simples.

Comecemos por uma ilustração e pela definição informal do modelo de regressão linear simples (RLS).

Exemplo 9.5 — **Regressão linear simples**
Pretende averiguar-se se a relação entre a percentagem de hidrocarbonetos presentes no condensador principal de uma unidade de destilação (x) e a pureza do oxigénio produzido (Y) é do tipo linear, tendo sido recolhidas para o efeito as 20 observações que constam de Montgomery e Runger (2003, pp. 373–374):

i	1	2	3	4	5	6	7	8	9	10
%Hidrocarbonetos	0.99	1.02	1.15	1.29	1.46	1.36	0.87	1.23	1.55	1.4
%Pureza	90.01	89.05	91.43	93.74	96.73	94.45	87.59	91.77	99.42	93.65

i	11	12	13	14	15	16	17	18	19	20
%Hidrocarbonetos	1.19	1.15	0.98	1.01	1.11	1.2	1.26	1.32	1.43	0.95
%Pureza	93.54	92.52	90.56	89.54	89.85	90.39	93.25	93.41	94.98	87.33



Tal como no Exemplo 9.4, o gráfico parece apontar no sentido de os pontos se dispersarem aleatoriamente em torno de uma recta. •

Definição informal 9.6 — Modelo de RLS
O modelo de RLS é definido por

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \; i = 1, \ldots, n, \tag{9.5}$$

onde:

- Y_i = resposta aleatória do indivíduo i (variável dependente aleatória)
- $x_i = i-$ observação da variável independente[1]
- β_0 = ordenada na origem (constante DESCONHECIDA)
- β_1 = declive (constante DESCONHECIDA)[2]
- ϵ_i = erro aleatório associado à observação da resposta do indivíduo i. •

A esta definição informal é necessário acrescentar dois sub-conjuntos de hipóteses de trabalho fundamentais quer para a estimação pontual, quer para outro tipo de inferências sobre os parâmetros do modelo tais como a ordenada na origem e o declive.

Nota 9.7 — Hipóteses de trabalho do modelo de RLS
É costume assumir que os erros aleatórios

- ϵ_i são v.a. não correlacionadas[3]

tais que

- $E(\epsilon_i) = 0$
- $V(\epsilon_i) = \sigma^2$ (constante DESCONHECIDA). •

Nota 9.8 — Consequências
À luz destas hipóteses de trabalho e da equação do modelo de RLS (9.5) segue-se, para $i = 1, \ldots, n$:

- $E(Y_i) = \beta_0 + \beta_1 x_i + E(\epsilon_i) = \beta_0 + \beta_1 x_i$
- $V(Y_i) = V(\epsilon_i) = \sigma^2$. •

---

[1] É habitual assumir que se trata de variável não aleatória — ou por tomar valor fixo, ou por tratar-se de medição sem erro ou com erro desprezvel

[2] Na verdade β_0 e β_1 são a ordenada na origem e o declive do valor esperado da resposta, respectivamente (ver nota 9.8).

[3] I.e., $corr(\epsilon_i, \epsilon_j) = 0, \; i \neq j$.

9.2.1 Estimação de β_0 e β_1 — método dos mínimos quadrados

O ponto de partida para a estimação pontual dos parâmetros β_0 e β_1 é o conjunto de dados:

- $(x_i, y_i),\ i = 1, \ldots, n.$

Motivação 9.9 — Método dos mínimos quadrados
A obtenção das estimativas dos mínimos quadrados de β_0 e β_1 passa pela minimização das discrepâncias entre o que é "esperado" pelo modelo de RLS — $E(Y_i) = \beta_0 + \beta_1 x_i$ — e o que é efectivamente observado — y_i. (Esquema...)

Com efeito, pretendemos encontrar estimativas de β_0 e β_1 que minimizem a soma dos quadrados dos desvios verticais entre y_i e $\beta_0 + \beta_1 x_i$, soma essa igual a

$$Q = Q(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2. \tag{9.6}$$

•

Proposição 9.10 — Estimativas de mínimos quadrados
As estimativas de mínimos quadrados de β_0 e β_1 — representadas doravante por $\hat{\beta}_0$ e $\hat{\beta}_1$ — são a solução do seguinte sistema de duas equações lineares:

$$(\hat{\beta}_0, \hat{\beta}_1) : \left\{ \begin{array}{lcl} \left.\frac{\partial Q}{\partial \beta_0}\right|_{\beta_0=\hat{\beta}_0, \beta_1=\hat{\beta}_1} & = & 0 \\ \left.\frac{\partial Q}{\partial \beta_1}\right|_{\beta_0=\hat{\beta}_0, \beta_1=\hat{\beta}_1} & = & 0 \end{array} \right. \tag{9.7}$$

i.e.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{9.8}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}. \tag{9.9}$$

•

Proposição 9.11 — Estimadores de mínimos quadrados
Os estimadores de mínimos quadrados de β_0 e β_1 serão igualmente representados por $\hat{\beta}_0$ e $\hat{\beta}_1$,[4] encontram-se no formulário da disciplina e são dados por:

$$\hat{\beta}_0 \stackrel{form}{=} \bar{Y} - \hat{\beta}_1 \bar{x} \tag{9.10}$$

$$\hat{\beta}_1 \stackrel{form}{=} \frac{\sum_{i=1}^{n} x_i Y_i - n\,\bar{x}\,\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}. \tag{9.11}$$

•

[4] Neste capítulo não faremos a destrinça entre estimador e estimativa, em termos notacionais, por mera conveniência.

Nota 9.12 — Estimativas de mínimos quadrados

- O sistema (9.7) é equivalente a

$$(\hat\beta_0,\hat\beta_1) \quad : \quad \begin{cases} -2\sum_{i=1}^n [y_i - (\hat\beta_0 + \hat\beta_1 x_i)] & = \quad 0 \\ -2\sum_{i=1}^n x_i[y_i - (\hat\beta_0 + \hat\beta_1 x_i)] & = \quad 0 \end{cases} \tag{9.12}$$

$$\begin{cases} n\,\hat\beta_0 + \hat\beta_1 \sum_{i=1}^n x_i & = \quad \sum_{i=1}^n y_i \\ \hat\beta_0 \sum_{i=1}^n x_i + \hat\beta_1 \sum_{i=1}^n x_i^2 & = \quad \sum_{i=1}^n x_i y_i \end{cases} \tag{9.13}$$

$$\begin{cases} n\,\hat\beta_0 + \hat\beta_1 \sum_{i=1}^n x_i & = \quad \sum_{i=1}^n y_i \\ \hat\beta_0 \sum_{i=1}^n x_i + \hat\beta_1 \sum_{i=1}^n x_i^2 & = \quad \sum_{i=1}^n x_i y_i \end{cases} \tag{9.14}$$

$$\begin{cases} \hat\beta_0 & = \quad \bar y - \hat\beta_1 \bar x \\ (\bar y - \hat\beta_1 \bar x) \sum_{i=1}^n x_i + \hat\beta_1 \sum_{i=1}^n x_i^2 & = \quad \sum_{i=1}^n x_i y_i \end{cases} \tag{9.15}$$

$$\begin{cases} \hat\beta_0 & = \quad \bar y - \hat\beta_1 \bar x \\ \hat\beta_1 \left(\sum_{i=1}^n x_i^2 - n\,\bar x^2\right) & = \quad \sum_{i=1}^n x_i y_i - n\,\bar x\,\bar y. \end{cases} \tag{9.16}$$

- Convém realçar que o sistema (9.7) possui solução sse

$$\sum_{i=1}^n x_i^2 - n\,\bar x^2 \neq 0, \tag{9.17}$$

i.e., sse na amostra existirem pelo menos dois valores distintos da variável explicativa x.

- Acrescente-se ainda que pode mostrar-se que a matriz hessiana de Q é semi-definida positiva, pelo que, caso exista solução $(\hat\beta_0, \hat\beta_1)$, ela corresponderá a um mínimo.

- As diferenças $e_i = y_i - (\hat\beta_0 + \hat\beta_1 x_i)$ são denominadas de resíduos. E será à custa destes resíduos que obteremos uma estimativa da variância σ^2. A sua análise (empírica) permitirá avaliar a adequação do modelo de RLS e será abordada na última secção deste capítulo. •

Nota 9.13 — Fórmula alternativa de $\hat\beta_1$

É frequente encontrar a seguinte fórmula alternativa de $\hat\beta_1$ na literatura. Esta fórmula é, por sinal, menos prática que a da Proposição 9.10 e é dada por:

$$\frac{\sum_{i=1}^n (x_i - \bar x)(y_i - \bar y)}{\sum_{i=1}^n (x_i - \bar x)^2}. \tag{9.18}$$

•

Exemplo 9.14 — **Estimativas de mínimos quadrados**

O custo de manutenção (em euros) de tractores parece aumentar com a idade (em anos) do tractor. Para verificar esta suposição, obtiveram-se os seguintes dados:

i	1	2	3	4	5	6	7	8	9
Idade (em anos)	0.5	0.5	1.0	1.0	1.0	4.0	4.0	4.0	4.5
Custo (em euros)	163	182	978	466	549	495	723	681	619

i	10	11	12	13	14	15	16	17
Idade (em anos)	4.5	4.5	5.0	5.0	5.0	5.5	6.0	6.0
Custo (em euros)	1049	1033	890	1522	1194	987	764	1373

Considerando um modelo de RLS, obtenha e interprete as estimativas de mínimos quadrados de β_0 e β_1.

- **Estimativas de mínimos quadrados**

  Tirando partido das expressões das estimativas de mínimos quadrados de β_0 e β_1 e do facto de $n = 17$ e de

  $\sum_{i=1}^n x_i = 62$
  $\bar{x} = 3.647058824$
  $\sum_{i=1}^n x_i^2 = 289.5$
  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 63.38235294$

  $\sum_{i=1}^n y_i = 13668$
  $\bar{y} = 804$
  $\sum_{i=1}^n y_i^2 = 13294114$
  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 2305042$ [5]

  $\sum_{i=1}^n x_i y_i = 58196.5$
  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 8348.5,$

  obtém-se:

[5] Embora estas quantidades sejam desnecessárias no curso deste exemplo, o seu cálculo é mais que recomendável já que o valor de $\sum_{i=1}^n y_i^2$ e muito em particular o de $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2$ é usado em outros tipos de inferências que faremos sobre os parâmetros do modelo.

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} \\
&= \frac{58196.5 - 17 \times 3.647058824 \times 804}{289.5 - 17 \times 3.647058824^2} \\
&= \frac{8348.5}{63.38235294} \\
&= 131.7164733 \text{ (Euro/ano)} \qquad (9.19)
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 804 - 131.7164733 \times 3.647058824 \\
&= 323.6222738 \text{ (Euro).} \qquad (9.20)
\end{aligned}$$

- **Interpretação das estimativas de mínimos quadrados**

  $\hat{\beta}_0 \simeq 323.6$

  Estima-se que o valor esperado do custo de manutenção de um tractor novo (com idade igual a 0 anos) seja de aproximadamente 323.6 Euros.

  $\hat{\beta}_1 \simeq 131.72$

  Por cada ano que passa, o valor esperado do custo de manutenção aumenta aproximadamente 131.72 Euros. •

Nota 9.15 — **Estimativas de mínimos quadrados**
Recomenda-se o uso do máximo de algarismos possíveis nos passos intermédios da obtenção das estimativas de mínimos quadrados, por forma a evitar a acumulação de erros de arredondamento.

Recomenda-se também que os arredondamentos só sejam efectuados aquando da apresentação dos valores das referidas estimativas. •

9.2.2 Estimação de β_0 e β_1 — método da MV

Para obter as estimativas de MV de β_0 e β_1 são necessárias hipóteses de trabalho adicionais que dizem mais uma vez respeito aos erros aleatórios. Estas hipóteses de trabalho adicionais possuem implicações distribuicionais que nos permitirão efectuar inferências de vário tipo, nomeadamente, obter intervalos de confiança e efectuar testes de hipóteses sobre β_0 e β_1 e outros parâmetros de interesse.

Nota 9.16 — Hipóteses de trabalho adicionais do modelo de RLS
Adicionalmente, assuma-se que

$$\epsilon_i \sim_{i.i.d.} \text{normal}(0, \sigma^2),\ i = 1, \ldots, n. \tag{9.21}$$

•

Nota 9.17 — Consequências
Ao lidar com erros aleatórios ϵ_i i.i.d. e normalmente distribuídos conclui-se que as seguintes funções afim destes erros

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \sim_{indep} \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n, \tag{9.22}$$

i.e.,

$$f_{Y_i|\beta_0,\beta_1}(y_i) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\left[y_i - (\beta_0 + \beta_1 x_i)\right]^2\right\},\ i = 1, \ldots, n. \tag{9.23}$$

•

Proposição 9.18 — Estimativas de MV
Pode adiantar-se que as estimativas de MV de β_0 e β_1 coincidem com as estimativas de mínimos quadrados deste par de parâmetros. Assim, não só são representadas por $\hat{\beta}_0$ e $\hat{\beta}_1$, como são iguais a

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{9.24}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}. \tag{9.25}$$

•

Nota 9.19 — Dedução das estimativas de MV

- **Função de verosimilhança**

  A função de verosimilhança é dada por

$$
\begin{aligned}
L(\beta_0, \beta_1|\underline{y}) &= \prod_{i=1}^{n} f_{Y_i|\beta_0,\beta_1}(y_i) \\
&= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\left[y_i - (\beta_0 + \beta_1 x_i)\right]^2\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i)\right]^2\right\}. \qquad (9.26)
\end{aligned}
$$

- **Log-verosimilhança**

  Por seu lado o logaritmo da função de verosimilhança é

$$
\begin{aligned}
\ln L(\beta_0, \beta_1|\underline{y}) &= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i)\right]^2 \\
&= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}Q, \qquad (9.27)
\end{aligned}
$$

  onde, recorde-se, Q é definido por (9.6).

- **Maximização**

  Tratando-se $\ln L(\beta_0, \beta_1|\underline{y})$ de uma função proporcional a $-Q$, a sua maximização no que diz respeito a β_0 e β_1 é equivalente à minimização de Q que conduziu, como se sabe, às estimativas seguintes:

$$
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (9.28)
$$

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}. \qquad (9.29)
$$

•

Nota 9.20 — **Notação alternativa**

Em alguns livros de texto pode encontrar-se a seguinte notação:

$$
\begin{aligned}
S_{XX} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 \qquad (9.30) \\
S_{XY} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n} x_i(y_i - \bar{y}) \\
&= \sum_{i=1}^{n} y_i(x_i - \bar{x}) \\
&= \sum_{i=1}^{n} x_i y_i - n\,\bar{x}\bar{y}. \qquad (9.31)
\end{aligned}
$$

Posto isto,

$$\hat{\beta}_0 = \bar{y} - \frac{S_{XY}}{S_{XX}}\,\bar{x} \tag{9.32}$$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}. \tag{9.33}$$

•

9.2.3 Recta de regressão

É usual estimar o valor esperado da resposta associada a um valor arbitrário x da variável explicativa. A estimativa pontual de $E(Y|x) = \beta_0 + \beta_1 x$ é igual a

$$\begin{aligned}\hat{y} &= \hat{E}(Y|x) \\ &= \hat{\beta}_0 + \hat{\beta}_1 x \qquad (9.34)\end{aligned}$$

e é habitual dar-se-lhe o nome de "fitted value".

A recta definida por

$$\hat{\beta}_0 + \hat{\beta}_1 x, \text{ para } x \in [x_{(1)}, x_{(n)}]$$

(onde $x_{(1)} = \min_{i=1,\ldots,n} x_i$ e $x_{(n)} = \max_{i=1,\ldots,n} x_i$), é usualmente designada de recta de regressão, recta estimada ou recta ajustada.

Exemplo 9.21 — Recta de regressão

Retome o Exemplo 9.4 (resistência do plástico) e obtenha uma estimativa para o valor esperado da resistência de uma peça de plástico que seja testada 20 horas após a sua moldagem.

- **Estimativas de β_0 e β_1**

  Note-se que $n = 12$ e

  $\sum_{i=1}^n x_i = 576$
  $\bar{x} = 48$
  $\sum_{i=1}^n x_i^2 = 31488$
  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 3840$

  $\sum_{i=1}^n y_i = 3239$
  $\bar{y} = 269.9166667$
  $\sum_{i=1}^n y_i^2 = 897639$
  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 23378.91667$

$\sum_{i=1}^{n} x_i y_i = 164752$

$\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 9280.$

Logo

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\ &= \frac{164752 - 12 \times 48 \times 269.9166667}{31488 - 12 \times 48^2} \\ &= \frac{9280}{3840} \\ &= 2.416666667 \qquad (9.35)\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 269.9166667 - 2.416666667 \times 48 \\ &= 153.9166667 \qquad (9.36)\end{aligned}$$

- **Recta de regressão**

  A estimativa do valor esperado da resistência de uma peça de plástico que seja testada x horas após a sua moldagem é igual a

$$\begin{aligned}\hat{y} &= \hat{E}(Y|x) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times x \\ &= 153.9166667 + 2.416666667 \times x. \qquad (9.37)\end{aligned}$$

  É esta a equação da recta de regressão. Ora, para $x = 20$, obtém-se

$$\begin{aligned}\hat{E}(Y|x = 20) &= 153.9166667 + 2.416666667 \times 20 \\ &= 202.25. \qquad (9.38)\end{aligned}$$

•

Exercício 9.22 — **Recta de regressão**
Trace a recta de regressão sobre o gráfico do Exemplo 9.4 (resistência do plástico) e obtenha alguns resíduos ($e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$) à sua escolha. •

9.3 Propriedades dos estimadores dos mínimos quadrados e estimação da variância.

É possível adiantar o valor esperado e a variância dos estimadores de mínimos quadrados de β_0 e β_1, bem como a covariância entre estes dois estimadores.

Proposição 9.23 — **Propriedades dos estimadores dos mínimos quadrados**

Parâmetro	**Estimador**	**Valor esperado**	**Variância**
β_0	$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$	β_0	$\sigma^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)$
β_1	$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - n\,\bar{x}\,\bar{Y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}$	β_1	$\frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}$
			$cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}$

Convém salientar que:

- $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores centrados de β_0 e β_1;
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são v.a. dependentes correlacionadas negativamente.[6] •

Motivação 9.24 — **Estimação de σ^2**
Para além de β_0 e β_1 é necessário estimar um outro parâmetro DESCONHECIDO:

- $V(\epsilon_i) = V(Y_i) = \sigma^2$. •

Tal como se teve ocasião de referir a estimativa de σ^2 obtém-se à custa dos resíduos, tal como podemos ver na proposição que se segue.

Proposição 9.25 — **Estimativa de σ^2**
A estimativa de σ^2 representar-se-á por $\hat{\sigma}^2$ e é igual a

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
&= \frac{1}{n-2}\sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 \times x_i)]^2 \\
&= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right], \qquad (9.39)
\end{aligned}$$

[6] Este facto é óbvio já que $\hat{\beta}_0$ depende de $\hat{\beta}_1$ através de uma sua função decrescente, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

sendo que a última expressão é a mais conveniente para o cálculo desta estimativa já que tira partido de quantidades previamente obtidas aquando da determinação de $\hat{\beta}_0$ e $\hat{\beta}_1$. •

Proposição 9.26 — Estimador de σ^2 e uma sua propriedade
O estimador de σ^2 consta do formulário e é definido por

$$\hat{\sigma}^2 \stackrel{form}{=} \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} Y_i^2 - n\,\bar{Y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right], \tag{9.40}$$

Refira-se também que se trata de um estimador centrado de σ^2.[7] •

Exemplo 9.27 — Estimativa de σ^2
Retome mais uma vez o Exemplo 9.4 (resistência do plástico) e obtenha a estimativa da variância da resistência.

A estimativa de σ^2 é, neste caso, igual a

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{12-2}\Big[\Big(897639 - 12 \times 269.9166667^2\Big) \\
&\qquad -2.416666667^2 \times \Big(31488 - 12 \times 48^2\Big)\Big] \\
&= \frac{1}{12-2}\Big(23378.91667 - 2.416666667^2 \times 3840\Big) \\
&= 95.225.
\end{aligned} \tag{9.41}$$

•

Exercício 9.28 — Estimativas das variâncias dos estimadores de β_0 e β_1
Retome o Exemplo 9.27 e obtenha estimativas das variâncias dos estimadores de β_0 e β_1, i.e., determine estimativas de $\sigma^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right)$ e de $\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}$, resp. •

[7] Embora não se trate do estimador de MV de σ^2 sob a hipótese de trabalho de erros aleatórios i.i.d. normalmente distribuídos.

Mais uma vez representamos do mesmo modo quer a estimativa, quer o estimador de um parâmetro desconhecido, neste caso σ^2.

9.4 Alguns abusos do modelo de regressão.

Escusado será dizer que a tentação de estabelecer uma relação entre uma variável resposta e uma ou mais variáveis explicativas e de estimar o valor esperado da resposta para valores arbitrários da variável explicativa pode conduzir a abusos na utilização de modelos de regressão. Entre eles destacaríamos dois. A saber:

- **Escolha incorrecta das variáveis explicativas**

  Uma escolha incorrecta das variáveis explicativas pode levar a conclusões absurdas já que uma associação estatística não implica necessariamente uma relação de causa e efeito.

- **Erro de extrapolação**

  A relação traduzida pelo modelo de RLS só é válida para a gama de valores observados da variável explicativa x,

  $$[x_{(1)}, x_{(n)}],$$

  onde, recorde-se, $x_{(1)} = \min_{i=1,\dots,n} x_i$ e $x_{(n)} = \max_{i=1,\dots,n} x_i$.

  Caso se obtenha um "fitted value"

  $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, \text{ para } x \notin [x_{(1)}, x_{(n)}],$$

  estamos a cometer aquilo que se designa por

  erro de extrapolação. (Esquema...)

Exemplo 9.29 — **Erro de extrapolação**
Tendo em vista o teste de uma nova solução de insulina foram administradas 3 doses diferentes dessa mesma solução a coelhos e registadas as quedas na quantidade de açúcar no sangue depois de um período fixo de tempo.

As observações referentes ao logaritmo da dose (x) e à queda da quantidade de açúcar no sangue (Y) encontram-se na tabela abaixo.

i	1	2	3	4	5	6	7	8	9	10	11	12
LnDose (x_i)	0.36	0.36	0.36	0.36	0.56	0.56	0.56	0.56	0.76	0.76	0.76	0.76
Queda açúcar (y_i)	17	21	49	54	64	48	34	63	62	72	61	91

(a) Obtenha a recta de regressão e uma estimativa do valor esperado da queda da quantidade de açúcar no sangue quando o logaritmo da dose é unitário, tendo em conta que $\sum_{i=1}^{n} x_i = 6.72$, $\sum_{i=1}^{n} x_i^2 = 4.0832$, $\sum_{i=1}^{n} y_i = 636$, $\sum_{i=1}^{n} y_i^2 = 38602$ e $\sum_{i=1}^{n} x_i y_i = 385.16$.

Se por um lado o logaritmo da dose administrada não é uma v.a. (por ser perfeitamente controlável), por outro lado a resposta, queda da quantidade de açúcar no sangue, é claramente uma v.a. Então, faz sentido recorrer à partida ao modelo de RLS.

- **Recta de regressão**

  Dado que $\bar{x} = \frac{6.72}{12} = 0.56$ e $\bar{y} = \frac{636}{12} = 53$ as estimativas de β_0 e β_1 são iguais a:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\ &= \frac{385.16 - 12 \times 0.56 \times 53}{4.0832 - 12 \times 0.56^2} \\ &= \frac{29}{0.32} \\ &= 90.625 \end{aligned} \tag{9.42}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 53 - 90.625 \times 0.56 \\ &= 2.25. \end{aligned} \tag{9.43}$$

  Logo a recta de regressão é dada por:

$$\begin{aligned}\hat{y} &= \hat{E}(Y|x) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times x \\ &= 2.25 + 90.625 \times x. \end{aligned} \tag{9.44}$$

- **Estimativa de $E(Y|x=1)$**

  A estimativa do valor esperado da queda da quantidade de açúcar no sangue quando o logaritmo da dose é unitário é $\hat{y} = 2.25 + 90.625 \times 1 = 92.875$. Note-se, no entanto, que esta estimativa foi obtida por EXTRAPOLAÇÃO uma vez que a gama de valores da variável explicativa é $[0.36, 0.76]$.

  Assim, esta estimativa deve ser usada com muito CAUTELA ou, melhor ainda, não deve ser utilizada.

(b) Calcule o logaritmo da dose a administrar de forma a obter uma estimativa do valor esperado da queda da quantidade de açúcar no sangue igual a 70.

- **Obtenção da dose a administrar**

  Pretende-se a abcissa x tal que $\hat{E}(Y|x) = 70$. Ora,

$$\hat{\beta}_0 + \hat{\beta}_1 \times x = 70 \Leftrightarrow x = \frac{70 - \hat{\beta}_0}{\hat{\beta}_1} \Leftrightarrow x = \frac{70 - 2.25}{90.625} \Leftrightarrow x = 0.7476 \quad (9.45)$$

  é o logaritmo da dose a administrar de modo a obter queda esperada da quantidade de açúcar no sangue igual a 70. •

9.5 Intervalos de confiança para β_0, β_1 e para o valor esperado da resposta.

Motivação 9.30 — Intervalos de confiança para β_0, β_1 e $E(Y|x_0) = \beta_0 + \beta_1 x_0$
Para além das estimativas pontuais, é de extrema utilidade adiantar intervalos de valores razoáveis para os parâmetros desconhecidos β_0 e β_1, bem como para o valor esperado da resposta quando a variável explicativa toma o valor x_0,

- $E(Y|x_0) = \beta_0 + \beta_1 x_0$. •

Para construir intervalos de confiança (e já agora efectuar testes de hipóteses) para β_0 e β_1 ou outros parâmetros de interesse no âmbito do modelo de RLS é crucial assumir que

$$\epsilon_i \sim_{i.i.d.} \text{normal}(0, \sigma^2),\ i = 1, \ldots, n, \tag{9.46}$$

já que nessa situação obtemos os seguintes resultados distribuicionais.

Proposição 9.31 — Distribuição dos estimadores de MV de β_0, β_1 e $\beta_0 + \beta_1 x_0$
Sob a validade de (9.46) temos $Y_i \sim_{indep} \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \ldots, n$, donde:

Parâmetro	**Estimador**	**Distribuição**
β_0	$\hat{\beta}_0$	$\text{normal}\left(\beta_0,\ \ \sigma^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)\right)$
β_1	$\hat{\beta}_1$	$\text{normal}\left(\beta_1,\ \ \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)$
$\beta_0 + \beta_1 x_0$	$\hat{\beta}_0 + \hat{\beta}_1 x_0$	$\text{normal}\left(\beta_0 + \beta_1 x_0,\ \ \sigma^2 \times \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)\right)$

•

Proposição 9.32 — V.a. fulcral para σ^2
Também sob a validade de (9.46) temos:

Parâmetro	**V.a. fulcral** para σ^2	**Distribuição**
σ^2	$\frac{(n-2)\,\hat{\sigma}^2}{\sigma^2}$	$\chi^2_{(n-2)}$

•

É exactamente por $\hat{\beta}_0$ e $\hat{\beta}_1$ terem distribuições normais e serem independentes da v.a. fulcral para σ^2 que seremos capazes de adiantar v.a. fulcrais para β_0, β_1 e $E(Y|x_0) = \beta_0 + \beta_1 x_0$. Elas possuem, em qualquer dos casos, distribuição $t-$Student já que $\frac{(n-2)\,\hat{\sigma}^2}{\sigma^2}$ tem distribuição do qui-quadrado.

Com efeito, ao reduzir os três estimadores da Proposição 9.31 e ao substituir σ^2 pelo seu estimador, obtemos as seguintes v.a. fulcrais que, por sinal, também constam do formulário. As concretizações dos intervalos aleatórios de confiança foram acrescentadas à tabela seguinte e obtêm-se sem grande dificuldade.

Proposição 9.33 — V.a. fulcrais para β_0, β_1 e $\beta_0 + \beta_1 x_0$

Os IC para estes três parâmetros de interesse obtêm-se, sob a validade de $\epsilon_i \sim_{i.i.d.}$ normal$(0, \sigma^2)$, $i = 1, \ldots, n$, à custa das seguintes v.a. fulcrais:

Parâmetro	**V.a. fulcral**	$IC_{(1-\alpha)\times 100\%}$
β_0	$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \sim t_{(n-2)}$	$\left[\hat{\beta}_0 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}\right]$
β_1	$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim t_{(n-2)}$	$\left[\hat{\beta}_1 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right]$
$\beta_0 + \beta_1 x_0$	$\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}$	$\left[(\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}\right]$

•

Nota 9.34 — Intervalos de confiança para β_0, β_1 e $E(Y|x_0) = \beta_0 + \beta_1 x_0$

- Estes três intervalos de confiança são do tipo

  estimativa centrada do parâmetro

  $\pm$quantil $\times \sqrt{\text{estimativa da variância do estimador do parâmetro}}$.

- A amplitude de $IC_{(1-\alpha)\times 100\%}(\beta_0 + \beta_1 x_0)$ aumenta à medida que nos afastamos de $\bar{x}$ (i.e., torna-se menos precisa), reflectindo, assim, os "perigos" da extrapolação. (Esquema...) •

Exemplo 9.35 — **Intervalo de confiança para** $E(Y|x) = \beta_0 + \beta_1 x_0$
Um astrónomo resolveu estudar a relação entre a distância e a velocidade de recessão entre nebulosas. Com esse objectivo registou para 24 nebulosas as distâncias a partir da terra (x, em *megaparsecs*) e as respectivas velocidades de recessão (y, em Km/s), tendo obtido

$$\textstyle\sum_{i=1}^{24} x_i = 21.873,\ \sum_{i=1}^{24} y_i = 8955,\ \sum_{i=1}^{24} x_i^2 = 29.5178,\ \sum_{i=1}^{24} y_i^2 = 6511425,$$
$$\textstyle\sum_{i=1}^{24} x_i y_i = 12513.7$$

e $\min(x_1, \dots, x_{24}) = 0.032$ e $\max(x_1, \dots, x_{24}) = 2.$[8]
Obtenha um intervalo de confiança a 90% para o valor esperado da velocidade de recessão de uma nebulosa a uma distância da terra de 0.55 *megaparsecs*, assumindo que o modelo de RLS é apropriado.[9]

- **Estimativas de MV**

  Assumindo que $\epsilon_i \sim_{i.i.d.}$ normal$(0, \sigma^2),\ i = 1, \dots, n$, e tirando partido do facto de

$$\textstyle\sum_{i=1}^{n} x_i = 21.873$$
$$\bar{x} = 0.911375$$
$$\textstyle\sum_{i=1}^{n} x_i^2 = 29.5178$$
$$\textstyle\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 9.583294625$$

$$\textstyle\sum_{i=1}^{n} y_i = 8955$$
$$\bar{y} = 373.125$$
$$\textstyle\sum_{i=1}^{n} y_i^2 = 6511425$$
$$\textstyle\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 = 3170090.625$$

$$\textstyle\sum_{i=1}^{n} x_i y_i = 12513.7,$$
$$\textstyle\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 4352.336875,$$

  obtém-se

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\
&= \frac{4352.336875}{9.583294625} \\
&= 454.1587
\end{aligned} \tag{9.47}$$

[8] 1 *parsec* = 3.26 *anos luz*.
[9] Adaptado do Exame de 24 de Junho de 2004.

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 373.125 - 454.1587 \times 0.911375 \\ &= -40.7839. \end{aligned} \tag{9.48}$$

- **Estimativa de MV de** $E(Y|x_0 = 0.55)$

  É igual a

$$\begin{aligned}\hat{E}(Y|x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\ &= -40.7839 + 454.1587 \times 0.55 \\ &= 209.0034. \end{aligned} \tag{9.49}$$

- **Estimativa de** σ^2

  É, neste caso, dada por

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\ &= \frac{1}{24-2}\Big(3170090.625 - 454.1587^2 \times 9.583294625\Big) \\ &= 54247.1805. \end{aligned} \tag{9.50}$$

- **IC a** 90% **para** $E(Y|x_0)$

  - **Passo 1 — V.a. fulcral para** $E(Y|x_0) = \beta_0 + \beta_1 x_0$

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)} \tag{9.51}$$

  - **Passo 2 — Quantis de probabilidade**

    Já que $(1-\alpha) \times 100\% = 90\%$ temos $\alpha = 0.10$ e lidaremos com os dois quantis simétricos:

$$\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) = \pm F^{-1}_{t_{(24-2)}}(1-0.10/2) \stackrel{tabela}{=} 1.717. \tag{9.52}$$

  - **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

    Omitímo-lo...

– **Passo 4 — Concretização**

É dado por

$$
\begin{aligned}
& IC_{(1-\alpha)\times 100\%}(\beta_0 + \beta_1 x_0) \\
&= \left[\; (\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\tfrac{1}{n} + \tfrac{(x_0-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]} \; \right] \\
&= \left[209.0039 \pm 1.717 \times \sqrt{54247.1805 \left[\tfrac{1}{24} + \tfrac{(0.55-0.911375)^2}{9.583294625}\right]}\right] \\
&= [209.0039 \pm 1.717 \times 54.7678] \\
&= [209.0039 \pm 94.0336] \\
&= [114.967, 303.0398].
\end{aligned}
$$

•

Exemplo 9.36 — Intervalos de confiança para β_0 e β_1
Determine IC para os coeficientes do modelo de RLS do Exemplo 9.35.

- **IC a** 90% **para** β_0

  ◦ **Passo 1 — V.a. fulcral para** β_0

  Da consulta do formulário segue-se

$$
Z = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}
$$

  ◦ **Passo 2 — Quantis de probabilidade**

  Já que $(1-\alpha)\times 100\% = 90\%$ temos $\alpha = 0.10$ e lidaremos com os dois quantis simétricos:

$$
\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) = \pm F^{-1}_{t_{(24-2)}}(1-0.10/2) \stackrel{tabela}{=} 1.717.
$$

  ◦ **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

  Omitímo-lo...

  ◦ **Passo 4 — Concretização**

  É dado por

$$
IC_{(1-\alpha)\times 100\%}(\beta_0) = \left[\; \hat{\beta}_0 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}
$$

Ora, tirando partido dos resultados do Exemplo 9.35, segue-se:

$$
\begin{aligned}
IC_{90\%}(\beta_0) &= \left[-40.7839 \pm 1.717 \times \sqrt{54247.1805 \left[\frac{1}{24} + \frac{0.911375^2}{9.583294625}\right]}\right] \\
&= [-184.045, 102.477].
\end{aligned}
$$

Mais, uma vez que $\beta_0 = 0 \in IC_{90\%}(\beta_0)$ não devemos rejeitar $H_0 : \beta_0 = \beta_{0,0} = 0$ (i.e., a hipótese de a recta de regressão passar pela origem) a qualquer n.s. $\alpha_0 \leq 100\% - 90\% = 10\%$.

- **IC a** 90% **para** β_1

  - **Passo 1 — V.a. fulcral para** β_1

    Consultando mais uma vez o formulário segue-se

    $$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim t_{(n-2)}$$

  - **Passo 2 — Quantis de probabilidade**

    Lidaremos novamente com os dois quantis simétricos:

    $$\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) = \pm F^{-1}_{t_{(24-2)}}(1-0.10/2) \stackrel{tabela}{=} 1.717.$$

  - **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

    Omitímo-lo...

  - **Passo 4 — Concretização**

    É dado por

    $$IC_{(1-\alpha)\times 100\%}(\beta_1) = \left[\hat{\beta}_1 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right].$$

    Recorrendo de novo aos resultados do Exemplo 9.35, temos:

    $$
    \begin{aligned}
    IC_{90\%}(\beta_1) &= \left[454.1587 \pm 1.717 \times \sqrt{\frac{54247.1805}{9.583294625}}\right] \\
    &= [324.982, 583.336].
    \end{aligned}
    $$

•

9.6 Testes de hipóteses sobre β_0, β_1 e o valor esperado da resposta.

Escusado será dizer que é a partir das v.a. fulcrais introduzidas na secção anterior que obteremos estatísticas de teste para o confronto de hipóteses sobre β_0, β_1 e $E(Y|x_0) = \beta_0 + \beta_1 x_0$.

Proposição 9.37 — **Estatísticas de teste para β_0 e β_1**
Mais uma vez sob a validade de $\epsilon_i \sim_{i.i.d.} \text{normal}(0, \sigma^2),\ i = 1, \ldots, n$, temos

Hipótese nula	**Estatística de teste**
$H_0 : \beta_0 = \beta_{0,0}$	$\frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \sim_{H_0} t_{(n-2)}$
$H_0 : \beta_1 = \beta_{1,0}$	$\frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$
$H_0 : E(Y|x_0) = \beta_0 + \beta_1 x_0 = E_0(Y|x_0)$	$\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - E_0(Y|x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim_{H_0} t_{(n-2)}$

•

Nota 9.38 — **Testes de hipóteses sobre β_1**
De entre os testes de hipóteses no âmbito do modelo de RLS destaque-se um teste de especial importância por vezes denominado de

- teste de significância da regressão.

A hipótese nula deste teste é

- $H_0 : \beta_1 = \beta_{1,0} = 0$.

Esta hipótese é sinónimo de inexistência de associação linear entre a variável resposta Y e a variável explicativa já que sob a validade de H_0 temos $Y_i = \beta_0 + \epsilon_i,\ i = 1, \ldots, n$. •

Exemplo 9.39 — Testes de hipóteses sobre β_1

Considere o conjunto de dados do Exemplo 9.35 e teste a hipótese de a velocidade de recessão das nebulosas não ser influenciada pela respectiva distância a partir da terra ao nível de significância de 10%.

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ vs. $H_1 : \beta_1 \neq 0$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)} \tag{9.53}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Estamos a lidar com

  – um teste bilateral ($H_1 : \beta_1 \neq 0$),

  pelo que, quanto maior for o valor absoluto da estimativa de MV de β_1, $\hat{\beta}_1$, mais nos devemos inclinar para a rejeição H_0. Logo a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo

$$W = (-\infty, -c) \cup (c, +\infty), \tag{9.54}$$

  onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

$$c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(24-2)}}(1 - 0.10/2) \overset{tabela}{=} 1.717. \tag{9.55}$$

- **Decisão**

  Tendo em conta que, de acordo com o Exemplo 9.35,

  $\hat{\beta}_1 = 454.1587$

  $\hat{\sigma}^2 = 54247.1805$

  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 9.583294625$,

  o valor observado da estatística de teste é dado por

$$
\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{454.1587 - 0}{\sqrt{\frac{54247.1805}{9.583294625}}} \\
&= 6.036378628 \\
&\in W = (-\infty, -1.717) \cup (1.717, +\infty). \qquad (9.56)
\end{aligned}
$$

Deste modo, concluimos que devemos rejeitar a hipótese de a velocidade de recessão das nebulosas não ser influenciada pela respectiva distância a partir da terra, ao n.s. de 10% bem como a qualquer outro n.s. maior que 10%. •

Exemplo 9.40 — **Testes de hipóteses sobre β_1**

Retome o conjunto de dados do Exemplo 9.4 (resistência do plástico) e teste a hipótese de o tempo que decorre a partir da conclusão do processo de moldagem até ao momento de medição da resistência não influenciar a resistência do plástico ao nível de significância de 5%.

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ vs. $H_1 : \beta_1 \neq 0$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$
T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)} \qquad (9.57)
$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Estamos a lidar com

  – um teste bilateral ($H_1 : \beta_1 \neq 0$),

  pelo que, quanto maior for o valor absoluto da estimativa de MV de β_1, $\hat{\beta}_1$, mais nos devemos inclinar para a rejeição H_0. Logo a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo

$$
W = (-\infty, -c) \cup (c, +\infty), \qquad (9.58)
$$

  onde

$$
c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(12-2)}}(1 - 0.05/2) =_{tabela} 2.228. \qquad (9.59)
$$

- **Decisão**

  Tendo em conta os seguintes resultados já obtidos anteriormente

  $\sum_{i=1}^n x_i = 576$

  $\bar{x} = 48$

  $\sum_{i=1}^n x_i^2 = 31488$

  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 3840$

  $\sum_{i=1}^n y_i = 3239$

  $\bar{y} = 269.9166667$

  $\sum_{i=1}^n y_i^2 = 897639$

  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 23378.91667$

  $\sum_{i=1}^n x_i y_i = 164752$

  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 9280.$

  $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{9280}{3840} = 2.416666667$

  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 269.9166667 - 2.416666667 \times 48 = 153.9166667$

  $\hat{\sigma}^2 = \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\,\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right]$
  $= \frac{1}{12-2}\left(23378.91667 - 2.416666667^2 \times 3840\right) = 95.225,$

  conclui-se que o valor observado da estatística de teste é dado por

$$
\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{2.416666667 - 0}{\sqrt{\frac{95.225}{3840}}} \\
&= 15.3464 \\
&\in W = (-\infty, -2.228) \cup (2.228, +\infty).
\end{aligned} \tag{9.60}
$$

  Deste modo, concluimos que devemos rejeitar a hipótese de a resistência do plástico não ser influenciada pelo tempo que decorre a partir da conclusão do processo de moldagem até ao momento de medição da resistência, ao n.s. de 5% bem como a qualquer outro n.s. maior que 5%. •

Exemplo 9.41 — Testes de hipóteses sobre β_1

Teste a significância da regressão para o conjunto de dados do Exemplo 9.5 (hidrocarbonetos), calculando para o efeito um intervalo para o $p - value$.

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ vs. $H_1 : \beta_1 \neq 0$

- **Nível de significância**

  α_0

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)} \tag{9.61}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  A região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2)$.

- **Decisão**

  Tendo em conta os seguintes resultados já obtidos anteriormente

  $\sum_{i=1}^n x_i = 23.92$

  $\bar{x} = 1.196$

  $\sum_{i=1}^n x_i^2 = 29.2892$

  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 0.68088$

  $\sum_{i=1}^n y_i = 1843.21$

  $\bar{y} = 92.1605$

  $\sum_{i=1}^n y_i^2 = 170044.5321$

  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 173.376895$

  $\sum_{i=1}^n x_i y_i = 2214.6566$

  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 10.17744.$

  $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{10.17744}{0.68088} = 14.94747973$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 92.1605 - 14.94747973 \times 1.196 = 74.28331424$

$\hat{\sigma}^2 = \frac{1}{n-2} \left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 \right) \right]$
$= \frac{1}{20-2} \left(173.376895 - 14.94747973^2 \times 0.68088 \right) = 1.18055,$

conclui-se que o valor observado da estatística de teste é dado por

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} = \frac{14.94747973 - 0}{\sqrt{\frac{1.18055}{0.68088}}} = 11.3517. \tag{9.62}$$

- **Intervalo para o p-value**

  Tratando-se de teste associado a uma região de rejeição de H_0 que é uma reunião de intervalos simétricos temos:

$$\begin{aligned} p - value &= 2 \times P(T > t \mid H_0) \\ &= 2 \times [1 - F_{T|H_0}(t)] \\ &= 2 \times [1 - F_{t_{(20-2)}}(11.3517)]. \end{aligned} \tag{9.63}$$

  Recorrendo às tabelas de quantis da distribuição do qui-quadrado com 18 graus de liberdade podemos obter um intervalo para o p-value deste teste. Para tal basta enquadrar convenientemente $t = 11.3517$ por dois quantis. No entanto, as tabelas disponíveis só permitem afirmar que

$$\begin{aligned} F^{-1}_{t_{(18)}}(0.9995) = 3.992 &< t = 11.3517 \\ 0.9995 &< F_{t_{(18)}}(11.3517) \\ 2 \times [1 - F_{t_{(18)}}(11.3517)] &< 2 \times (1 - 0.9995) \end{aligned}$$

  e por fim o intervalo para o p-value é (0,0.001).

  Deste modo, conclui-se que devemos rejeitar H_0 (hipótese de a pureza do oxigénio produzido não ser influenciada pela percentagem de hidrocarbonetos) a qualquer n.s. $\alpha_0 \geq 0.1\%$. •

Exemplo 9.42 — **Testes de hipóteses sobre β_0 e $E(Y|x_0) = \beta_0 + \beta_1 x_0$**
Com o objectivo de estimar a constante de elasticidade $(\beta_1)^{-1}$ (em Newton/metro) de certo tipo de mola, foram seleccionadas 10 dessas molas, todas com o mesmo comprimento em repouso β_0 (em metros). Foi obtido o seguinte conjunto de resultados, após se ter suspendido à mola i $(i = 1, \ldots, 10)$ um corpo com peso x_i (em Newton) e medido o comprimento da mola após tal suspensão y_i:

i	1	2	3	4	5	6	7	8	9	10
x_i	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
y_i	0.206	0.212	0.218	0.225	0.239	0.237	0.245	0.251	0.257	0.263

$\sum_{i=1}^{10} x_i = 13.75$ $\quad \sum_{i=1}^{10} x_i^2 - n\bar{x}^2 = 5.15625$ $\quad \sum_{i=1}^{10} y_i = 2.353$ $\quad \sum_{i=1}^{10} y_i^2 - n\bar{y}^2 = 0.0034021$

$\sum_{i=1}^{10} x_i y_i - n\bar{x}\bar{y} = 0.131375.$

Admita a validade do modelo de regressão linear simples $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ $(i = 1, \ldots, 10)$ para este conjunto de dados e responda às questões que se seguem:[10]

(a) Obtenha estimativas de máxima verosimilhança do declive da recta, da constante de elasticidade $(\beta_1)^{-1}$ e do valor esperado do comprimento da mola ao sujeitá-la a uma força de 3.0 Newton. Comente esta última estimativa.

- **Estimativa de MV do declive da recta** (β_1)

  Ao tomar como hipótese de trabalho $\epsilon_i \sim_{i.i.d.} \text{normal}(0, \sigma^2)$, $i = 1, \ldots, n$, e ao ter consideração que $n = 10$ e

  $\sum_{i=1}^{n} x_i = 13.75$
  $\bar{x} = 1.375$
  $\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 5.15625$

  $\sum_{i=1}^{n} y_i = 2.353$
  $\bar{y} = 0.2353$
  $\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 = 0.0034021$

  $\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 0.131375,$

  depressa se conclui que

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\ &= \frac{0.131375}{5.15625} \\ &= 0.0254788.\end{aligned}$$

[10] Adaptado do Exame de 24 de Junho de 2006.

- **Estimativa de MV da constante de elasticidade** $(\beta_1)^{-1}$

  Ao invocar-se a propriedade de invariância dos estimadores de MV pode adiantar-se que a estimativa de MV da constante de elasticidade,

  $$h(\beta_1) = (\beta_1)^{-1},$$

  é dada por

  $$\begin{aligned} \widehat{h(\beta_1)} &= h(\hat{\beta}_1) \\ &= (\hat{\beta}_1)^{-1} \\ &= \frac{1}{0.0254788} \\ &= 39.2483. \end{aligned}$$

- **Estimativa de MV de** $E(Y|x_0 = 3)$

  Invocando novamente a propriedade de invariância dos estimadores de MV e tirando partido quer de $\hat{\beta}_1 = 0.0254788$ quer de

  $$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 0.2353 - 0.0254788 \times 1.375 \\ &= 0.200267, \end{aligned}$$

  pode adiantar-se que

  $$\begin{aligned} \hat{E}(Y|x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\ &= 0.200267 + 0.0254788 \times 3 \\ &= 0.276703. \end{aligned}$$

- **Comentário**

  A estimativa de MV do valor esperado do comprimento de uma mola após a suspensão de peso $x_0 = 3.0$ deve ser usada com toda alguma cautela uma vez que foi obtida por EXTRAPOLAÇÃO. Com efeito, $x_0 = 3.0 \notin [x_{(1)}, x_{(n)}] = [0.5, 2.5]$, ou seja o valor de x_0 não pertence à gama de pesos considerados para a obtenção da recta de regressão.

(b) Teste a hipótese de o comprimento em repouso comum às 10 molas β_0 (em metros) ser igual a $0.2m$ ao nível de significância de 5%.

- **Hipóteses**

  $H_0 : \beta_0 = \beta_{0,0} = 0.2$ vs. $H_1 : \beta_0 \neq 0.2$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$\frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Dado que

  - o teste é bilateral ($H_1 : \beta_0 \neq 0.2$),

  a região de rejeição de H_0 (para valores da estatística de teste) é a reunião de intervalos

$$W = (-\infty, -c) \cup (c, +\infty),$$

  onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, ou seja,

$$c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(10-2)}}(1 - 0.05/2) \stackrel{tabela}{=} 2.306.$$

- **Decisão**

  Tendo em conta os cálculos anteriores bem como o facto de a estimativa centrada de σ^2 ser igual a

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{10-2}\left(0.0034021 - 0.0254788^2 \times 5.15625\right) \\
&= 0.00000685263,
\end{aligned}$$

  tem-se o seguinte valor observado da estatística de teste:

$$\begin{aligned}
t &= \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \\
&= \frac{0.200267 - 0.2}{\sqrt{0.00000685263 \times \left(\frac{1}{10} + \frac{1.375^2}{5.15625}\right)}} \\
&= 0.149307 \\
&\notin W = (-\infty, -2.306) \cup (2.306, +\infty).
\end{aligned}$$

  Logo não se deve rejeitar a hipótese de o comprimento em repouso das molas (β_0) ser igual a $20cm$ ao n.s. de 5% nem a qualquer outro n.s. menor que 5%.

(c) Teste agora a hipótese de o comprimento esperado após suspensão de um peso de $x_0 = 2$ ser igual a $0.25m$ ao nível de significância de 10%.

- **Hipóteses**

  $H_0 : E(Y|x_0 = 2) = E_0(Y|x_0 = 2) = 0.25$ vs.

  $H_1 : E(Y|x_0 = 2) \neq E_0(Y|x_0 = 2)$

- **Nível de significância**

  $\alpha_0 = 10\%$

- **Estatística de teste**

$$\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - E_0(Y|x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  A região de rejeição de H_0 é a reunião de intervalos $W = (-\infty, -c) \cup (c, +\infty)$, onde

$$c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(10-2)}}(1 - 0.10/2) \stackrel{tabela}{=} 1.860,$$

  já que o teste é bilateral ($H_1 : E(Y|x_0 = 2) \neq E_0(Y|x_0 = 2)$).

- **Decisão**

  Capitalizando nos cálculos das alíneas anteriores obtém-se o seguinte valor observado da estatística de teste

$$\begin{aligned}
t &= \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - E_0(Y|x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \\
&= \frac{(0.200267 + 0.0254788 \times 2) - 0.25}{\sqrt{0.00000685263 \times \left[\frac{1}{10} + \frac{(2.0-1.375)^2}{5.15625}\right]}} \\
&= 1.11586 \\
&\notin W = (-\infty, -1.860) \cup (1.860, +\infty).
\end{aligned}$$

  Assim sendo, não se deve rejeitar a hipótese de o comprimento esperado de mola após suspensão de peso de 2 Newton ser igual a $0.25m$ a qualquer n.s. menor ou igual a 10%. •

9.7 Coeficiente de determinação e análise de resíduos na avaliação do modelo

Motivação 9.43 — Coeficiente de determinação
Para além de averiguarmos se a variável resposta (Y) depende linearmente da variável explicativa (x), é fundamental verificar se a recta de regressão se ajusta bem ao conjunto de dados. Para o efeito limitar-nos-emos, no âmbito desta disciplina, a calcular o coeficiente de determinação. •

Definição informal 9.44 — Coeficiente de determinação
Trata-se de coeficiente que dá ideia aproximada do ajustamento da recta de regressão aos dados. É definido por

$$\begin{aligned} r^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)}. \end{aligned} \tag{9.64}$$

Assim, $r^2 \times 100\%$ corresponde à

- percentagem da variação total ($\sum_{i=1}^n (y_i - \bar{y})^2$) explicada pela variável regressora x ($\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$). •

Nota 9.45 — Coeficiente de determinação
Escusado será dizer que a segunda fórmula do coeficiente de determinação que figura na definição informal 9.44 é de longe a mais conveniente exactamente por tirar partido de quantidades previamente calculadas.[11]

Importa notar ainda que:

- $r^2 \in [0, 1]$
- $r^2 = 1 \Leftrightarrow \hat{y}_i = y_i, i = 1, \ldots, n \rightarrow$ modelo RLS MUITO BOM.

  Com efeito, caso $\hat{y}_i = y_i, i = 1, \ldots, n$, todos os pontos estão sobre a recta de regressão, pelo que o modelo de regressão explica toda a variabilidade observada na variável resposta Y. O modelo de RLS é muito bom.

[11] Esta fórmula leva-nos a afirmar que o coeficiente de determinação corresponde ao quadrado do coeficiente de correlação empírica entre as variáveis x e y.

- $r^2 = 0 \Leftrightarrow \hat{y}_i = \bar{y}, i = 1, \ldots, n \rightarrow$ modelo RLS MUITO MAU.

  De facto, caso $\hat{y}_i = \bar{y}, i = 1, \ldots, n$, o modelo de regressão é incapaz de explicar a variabilidade observada na variável resposta Y. O modelo de RLS é muito mau. •

Exemplo 9.46 — **Coeficiente de determinação**
Retome o Exemplo 9.35 (velocidade de recessão de nebulosas). Calcule e comente o valor do coeficiente de determinação.

- **Coeficiente de determinação**

  É dado por

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{(12513.7 - 24 \times 0.911375 \times 373.125)^2}{(29.5178 - 24 \times 0.911375^2) \times (6511425 - 24 \times 373.125^2)} \\
&= \frac{4352.336875^2}{9.58329 \times 3170090.625} \\
&= 0.6235. \qquad (9.65)
\end{aligned}
$$

  A recta de regressão explica 62.35% da variação total da variável resposta Y. Assim, havendo cerca de 40% de variação não explicada, pode afirmar-se que a recta estimada se ajusta razoavelmente ao nosso conjunto de dados. •

Exercício 9.47 — **Coeficiente de determinação**
Certifique-se que o coeficiente de determinação do Exemplo 9.4 (resistência do plástico) é ligeiramente inferior a 96%, podendo concluir-se que a recta estimada se ajusta muito bem aos dados, o que já era bem visível pelo gráfico que acompanhava este mesmo exemplo. •

Nota 9.48 — **Análise empírica de resíduos**
Por forma a averiguar a adequação do modelo de RLS e a razoabilidade das hipóteses de trabalho deve elaborar-se um gráfico em cujas

- abcissas devem constar os "fitted values" ($\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1\, x_i$)

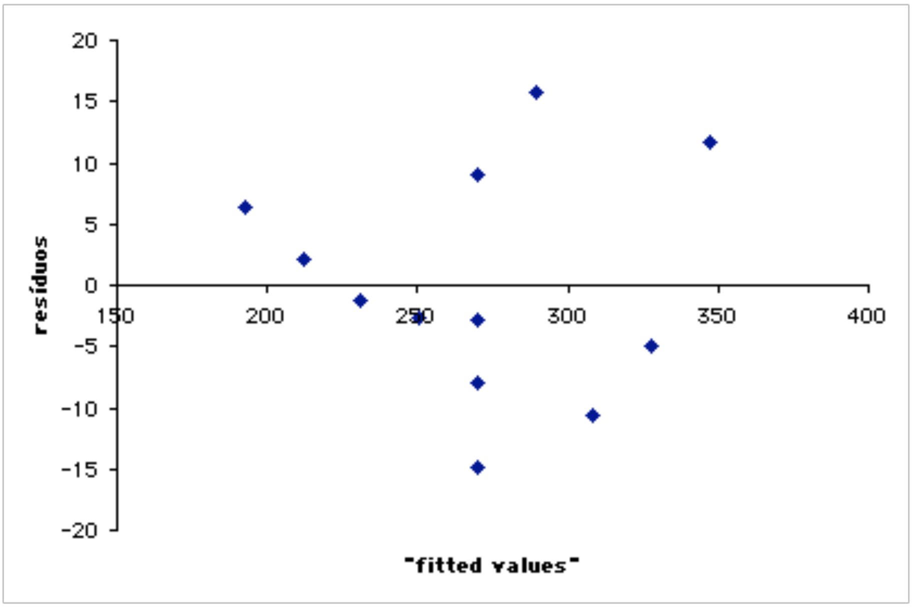
e cujas

- ordenadas correspondem aos resíduos ($e_i = y_i - \hat{y}_i$).

Caso as hipóteses de trabalho sobre os erros aleatórios ($\epsilon_i \sim_{i.i.d.}$ normal$(0,\sigma^2)$) não estejam a ser violadas:

- o gráfico acabado de descrever deve ter o aspecto de uma mancha de pontos dispersos aleatoriamente numa faixa horizontal (Esquema...) e
- o histograma dos resíduos deve ter um aspecto simétrico e semelhante à f.d.p. da v.a normal. •

Exemplo 9.49 — Análise empírica de resíduos
De seguida obtém-se o gráfico de resíduos ($e_i = y_i - \hat{y}_i$) vs. "fitted values" ($\hat{y}_i$), para os dados da resistência do plástico do Exemplo 9.4.



Este gráfico assemelha-se a uma mancha de pontos dispersos ao acaso numa faixa horizontal, pelo que se pode concluir que a recta estimada se ajusta muito bem aos dados, o que, aliás, se coaduna quer com o gráfico dos dados (ver Exemplo 9.4), quer com o valor elevado do coeficiente de determinação (ver Exercício 9.47). •

Nota 9.50 — Análise empírica de resíduos
Dependendo do aspecto do gráfico de resíduos ($e_i = y_i - \hat{y}_i$) vs. "fitted values" ($\hat{y}_i$), podemos ser levados a concluir que estamos a lidar com

- um modelo de regressão linear com erros aleatórios independente com variância constante

ou então com

- um modelo de regressão não linear
- erros aleatórios com variância não constante
- observações discordantes
- falta de independência. (Esquemas...) •

Exemplo 9.51 — **Exame de 9 de Janeiro de 2007**
Numa dada região pretende determinar-se se existe uma relação entre as despesas básicas anuais das famílias (Y) e o seu rendimento anual (x).

A recolha de uma amostra de 10 famílias conduziu aos seguintes resultados, em milhares de euros:

i	1	2	3	4	5	6	7	8	9	10
Rendimento (x_i)	22	26	45	37	28	50	56	34	60	40
Despesas básicas anuais (y_i)	16	17	26	24	22	21	32	18	30	20

$\sum_{i=1}^{10} x_i = 398 \quad \sum_{i=1}^{10} x_i^2 = 17330 \quad \sum_{i=1}^{10} y_i = 226 \quad \sum_{i=1}^{10} y_i^2 = 5370 \quad \sum_{i=1}^{10} x_i y_i = 9522.$

Admita a validade do modelo de regressão linear simples $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ ($i = 1, \ldots, 10$) para este conjunto de dados e responda às seguintes questões:

(a) Estime a recta de regressão.

- **Estimativas de β_0 e β_1**

  Tendo em conta que $n = 10$ e

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n} x_i = 398\\
&\bar{x} = 39.8\\
&\textstyle\sum_{i=1}^{n} x_i^2 = 17330\\
&\textstyle\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 1489.6
\end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n} y_i = 226\\
&\bar{y} = 22.6\\
&\textstyle\sum_{i=1}^{n} y_i^2 = 5370\\
&\textstyle\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 = 262.4
\end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n} x_i y_i = 9522,\\
&\textstyle\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 527.2,
\end{aligned}$$

  obtém-se

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\\
&= \frac{527.2}{1489.6}\\
&= 0.35392\\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}\\
&= 22.6 - 0.35392 \times 39.8\\
&= 8.5196.
\end{aligned}$$

- **Recta de regressão**

$$\hat{y} = \hat{E}(Y|x) = \hat{\beta}_0 + \hat{\beta}_1 x = 8.5196 + 0.35392\,x, \text{ para } x \in [22, 60].$$

(b) Calcule o coeficiente de determinação e comente o valor que obteve.

- **Coeficiente de determinação**

  Tirando partido dos cálculos auxiliares da alínea anterior tem-se:

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{527.2^2}{1489.6 \times 262.4} \\ &= 0.711. \end{aligned}$$

  71.1% da variação total nas despesas básicas anuais das famílias (Y) é explicada pelo rendimento anual (x), pelo que a recta estimada se ajusta razoavelmente ao conjunto de dados.

(c) Será que a amostra de que dispõe permite concluir — ao nível de significância de 5% — que há relação linear positiva entre as despesas básicas anuais e o rendimento anual?

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ vs. $H_1 : \beta_1 > 0$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Ao lidar com

  – um teste unilateral à direita ($H_1 : \beta_1 > 0$),

  quanto maior for o valor da estimativa de MV de β_1, $\hat{\beta}_1$, mais razões temos para rejeitar H_0, pelo que a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo à direita

$$W = (c, +\infty),$$

onde

$$c = F^{-1}_{t_{(n-2)}}(1-\alpha_0) = F^{-1}_{t_{(10-2)}}(1-0.05) \stackrel{tabela}{=} 1.86.$$

- **Decisão**

  Ora,

$$\begin{aligned}
\hat{\beta}_1 &= 0.35392 \\
\hat{\sigma}^2 &= \tfrac{1}{n-2}\left[\left(\textstyle\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\textstyle\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \tfrac{1}{10-2}(262.4 - 0.35392^2 \times 1489.6) \\
&= 9.47671 \\
\textstyle\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 &= 1489.6
\end{aligned}$$

  logo o valor observado da estatística de teste é igual a

$$\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{0.35392 - 0}{\sqrt{\frac{9.47671}{1489.6}}} \\
&= 4.437 \\
&\in W = (1.86, +\infty).
\end{aligned}$$

  Conclui-se que devemos rejeitar a hipótese de as despesas básicas anuais não serem influenciadas pelo rendimento anual a favor de uma relação ser do tipo linear positivo entre estas variáveis, ao n.s. de 5% bem como a qualquer outro n.s. maior que 5%. •

Exemplo 9.52 — Exame de 25 de Janeiro de 2007

Num processo de fabrico suspeita-se que o número de artigos defeituosos produzidos por uma máquina (Y) dependa da velocidade a que essa mesma máquina está a operar (x). Abaixo encontram-se 12 registos do número de defeituosos e da velocidade associada.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	200	400	300	400	200	300	300	400	200	400	200	300
y_i	28	75	37	53	22	58	40	96	46	52	30	69

$\sum_{i=1}^{12} x_i = 3600$, $\sum_{i=1}^{12} x_i^2 = 1160000$, $\sum_{i=1}^{12} y_i = 606$, $\sum_{i=1}^{12} y_i^2 = 35732$, $\sum_{i=1}^{12} x_i y_i = 196800$.

(a) Admitindo a validade do modelo de regressão linear simples ($Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, 12$), obtenha estimativas pontuais de β_0 e do aumento esperado no número de artigos defeituosos produzidos caso a velocidade duplique de 200 para 400 unidades.

- **Estimativa de β_0**

  Como $n = 12$ e

  $\sum_{i=1}^n x_i = 3600$
  $\bar{x} = 300$
  $\sum_{i=1}^n x_i^2 = 1160000$
  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 80000$

  $\sum_{i=1}^n y_i = 606$
  $\bar{y} = 50.5$
  $\sum_{i=1}^n y_i^2 = 35732$
  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 5129$

  $\sum_{i=1}^n x_i y_i = 196800,$
  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 15000,$
  $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{15000}{80000} = 0.1875$

  conclui-se que

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \\ &= 50.5 - 0.1875 \times 300 \\ &= -5.75.\end{aligned}$$

- **Estimativa do aumento esperado quando...**

  O aumento esperado no número de artigos defeituosos produzidos caso a velocidade duplique de 200 para 400 unidades é dado por

$$\begin{aligned}E(Y|x = 400) - E(Y|x = 200) &= (\beta_0 + \beta_1 \times 400) - (\beta_0 + \beta_1 \times 200) \\ &= 200\,\beta_1.\end{aligned}$$

  Assim sendo a estimativa deste aumento é igual a

$$\begin{aligned}200\,\hat{\beta}_1 &= 200 \times 0.1875 \\ &= 37.5 \text{ artigos defeituosos.}\end{aligned}$$

(b) Determine um intervalo de confiança a 95% para o valor esperado do número de artigos defeituosos produzidos quando a velocidade da máquina é de 450 unidades. Comente esta estimativa intervalar.

- **Estimativa de** $E(Y|x_0)$

  Ao considerar-se que a velocidade da máquina é de 450 unidades, estima-se que o valor esperado do número de artigos defeituosos produzidos seja igual a

$$\begin{aligned}\hat{E}(Y|x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\ &= -5.75 + 0.1875 \times 450 \\ &= 78.625 \text{ artigos defeituosos.}\end{aligned}$$

- **IC a** 95% **para** $E(Y|x_0)$

  - **Passo 1 — Selecção da v.a. fulcral para** $E(Y|x_0) = \beta_0 + \beta_1 x_0$

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}$$

  - **Passo 2 — Obtenção dos quantis de probabilidade**

    Tendo em conta que $(1-\alpha) \times 100\% = 95\%$ temos $\alpha = 0.05$ e usaremos com os quantis simétricos:

$$\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) = \pm F^{-1}_{t_{(12-2)}}(1-0.05/2) \stackrel{tabela}{=} 2.228.$$

  - **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

    Omitimos este passo...

  - **Passo 4 — Concretização**

    Como a expressão geral do IC a $(1-\alpha) \times 100\%$ para $\beta_0 + \beta_1 x_0$

$$IC_{(1-\alpha)\times 100\%}(\beta_0 + \beta_1 x_0)$$
$$= \left[\ (\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}\ \right]$$

    e

$$n = 10$$
$$x_0 = 450$$
$$\hat{\beta}_0 + \hat{\beta}_1 x_0 = 78.625$$
$$F^{-1}_{t_{(n-2)}}(1-\alpha/2) = 2.228$$
$$\begin{aligned}\hat{\sigma}^2 &= \tfrac{1}{n-2}\left[\left(\textstyle\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\ \left(\textstyle\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\ &= \tfrac{1}{12-2}(5129 - 0.1875^2 \times 80000) \\ &= 231.65\end{aligned}$$

$$\bar{x} = 300$$

tem-se

$$= \left[78.625 \pm 2.228 \times \sqrt{231.65 \times \left[\frac{1}{12} + \frac{(450-300)^2}{80000}\right]}\right]$$
$$= [78.625 \pm 2.228 \times 9.18998]$$
$$= [58.1497,\ 99.1003].$$

- **Comentário** — Como

$$x_0 = 450 \not\in [x_{(1)},\ x_{(n)}] = [\min_{i=1,\ldots,n} x_i,\ \max_{i=1,\ldots,n} x_i] = [200,\ 400],$$

i.e., $x_0 = 450$ não pertence à gama de valores da variável explicativa, estamos a cometer um ERRO DE EXTRAPOLAÇÃO, pelo que as estimativas pontual e intervalar de $E(Y|x_0 = 450)$ devem ser usadas com muita cautela.

(c) Calcule o coeficiente de determinação e comente o valor obtido.

- **Coeficiente de determinação**

Tirando mais uma vez partido dos cálculos auxiliares da alínea (a) obtém-se:

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{15000^2}{80000 \times 5129} \\ &\simeq 0.548. \end{aligned}$$

Apenas 54.8% da variação total no número de artigos defeituosos (Y) é explicada pela velocidade a que a máquina está a operar (x), pelo que a recta estimada se ajusta de forma nada razoável ao conjunto de dados.

Mais uma razão para ter cautela com as estimativas pontual e intervalar obtidas na alínea (b). •

Exemplo 9.53 (Exercício 9.1, colectânea de exercícios) — **Regressão linear simples; estimação pontual de MQ; recta de regressão; coeficiente de determinação; teste à significância da regressão; intervalo de confiança para o valor esperado da resposta; erro de extrapolação**

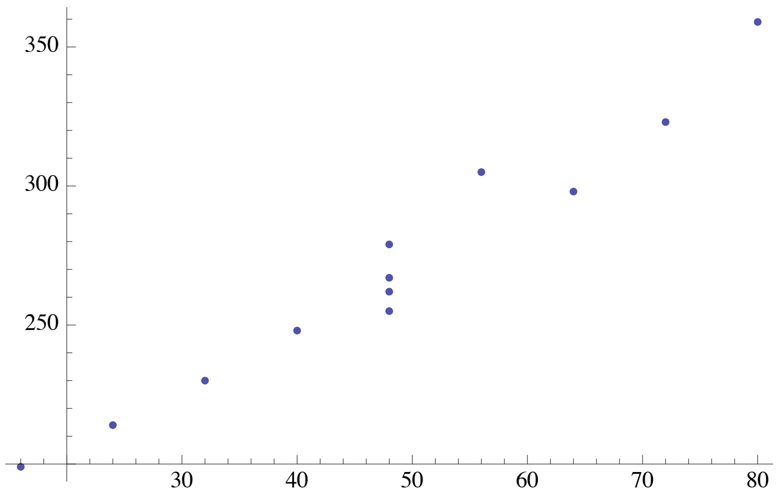
Interessa estudar a relação entre a resistência de um determinado tipo de plástico (Y) e o tempo que decorre a partir da conclusão do processo de moldagem até ao momento de medição da resistência (x) (em horas). As observações que se seguem foram efectuadas em 12 peças construídas com este plástico, escolhidas aleatoriamente.

i	1	2	3	4	5	6	7	8	9	10	11	12
Tempo (x_i)	16	24	32	40	48	48	48	48	56	64	72	80
Resistência (y_i)	199	214	230	248	255	262	279	267	305	298	323	359

$$\sum_i x_i = 576,\ \sum_i x_i^2 = 31488,\ \sum_i y_i = 3239,\ \sum_i y_i^2 = 897639,\ \sum_i x_i y_i = 164752.$$

(a) Represente graficamente as observações e desenhe a recta que, no seu entender, melhor se a justa às observações.

- **Representação gráfica dos dados**



Apesar de nenhuma curva simples passar exactamente por todos os pontos, há forte indicação no sentido de os pontos do gráfico se dispersarem aleatoriamente em torno de uma recta.

(b) Considere um modelo de regressão linear simples para explicar as observações. Obtenha a estimativa dos mínimos quadrados dos coeficientes da recta de regressão e desenhe-a no gráfico.

- **Modelo de RLS**

  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

  $Y_i =$ resistência do plástico da peça i

  $\beta_0 =$ ordenada na origem (constante DESCONHECIDA)

  $\beta_1 =$ declive (constante DESCONHECIDA)

  $x_i =$ tempo decorrido entre a conclusão da moldagem e a medição da resistência d peça i

  $\epsilon_i =$ erro aleatório associado à medição da resistência da peça i

- **Hipóteses de trabalho**

  ϵ_i são v.a.

  - não correlacionadas[12]

  tais que

  - $E(\epsilon_i) = 0$
  - $V(\epsilon_i) = \sigma^2$ (constante DESCONHECIDA).

- **Estimativas de β_0 e β_1**

  Uma vez que $n = 12$ e

  - $\sum_{i=1}^n x_i = 576$
    $\bar{x} = 48$
    $\sum_{i=1}^n x_i^2 = 31488$
    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 3840$

  - $\sum_{i=1}^n y_i = 3239$
    $\bar{y} = 269.9166667$
    $\sum_{i=1}^n y_i^2 = 897639$
    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 23378.91667$

  - $\sum_{i=1}^n x_i y_i = 164752$
    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 9280,$

[12] I.e., $corr(\epsilon_i, \epsilon_j) = 0,\ i \neq j$.

tem-se

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\
&= \frac{164752 - 12 \times 48 \times 269.9166667}{31488 - 12 \times 48^2} \\
&= \frac{9280}{3840} \\
&= 2.416666667
\end{aligned}$$

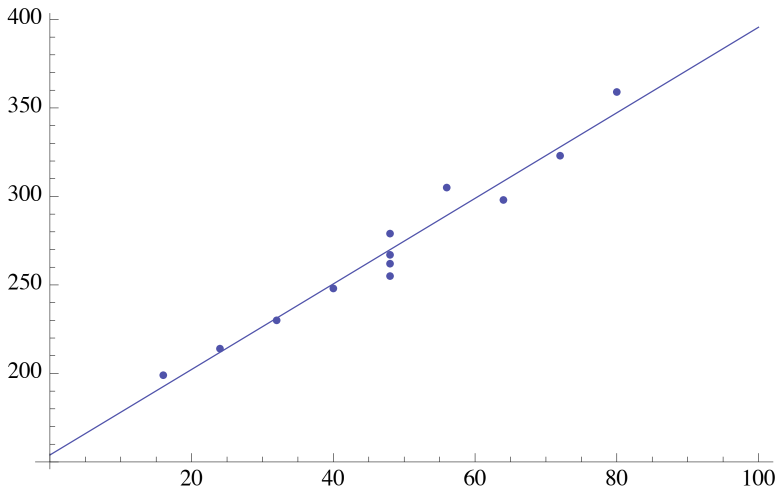
$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 269.9166667 - 2.416666667 \times 48 \\
&= 153.9166667
\end{aligned}$$

- **Recta de regressão**

  A estimativa do valor esperado da resistência de uma peça de plástico que seja testada x horas após a sua moldagem é igual a

$$\begin{aligned}
\hat{y} &= \hat{E}(Y|x) \\
&= \hat{\beta}_0 + \hat{\beta}_1 \times x \\
&= 153.9166667 + 2.416666667 \times x.
\end{aligned}$$

- **Representação gráfica dos dados e da recta de regressão**



Muito bom ajuste!

(c) Calcule o coeficiente de determinação e comente o valor obtido.

- **Coeficiente de determinação**

  Tirando partido dos valores obtidos anteriormente, tem-se

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{(9280)^2}{3840 \times 23378.91667} \\
&= 0.959269.
\end{aligned}
$$

  Logo pode afirmar-se cerca de 95.93% da variação total dos dados referentes à variável resposta Y é explicada pelo modelo de regressão, o que indicia um muito bom ajustamento da recta estimada ao nosso conjunto de dados, tal como já se tinha constatado na alínea anterior.

(d) Proceda ao teste da hipótese "O coeficiente angular é nulo", indicando e comentando o valor-p do teste. Qual o interesse desta hipótese? Relacione-o com o resultado obtido em (c).

- **Hipótese de trabalho**

  $\epsilon_i \sim_{i.i.d.} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$ (hipótese de trabalho)

  $\beta_0 =$ ordenada na origem de $E(Y_i|x_i) = \beta_0 + \beta_1 x_i$ (DESCONHECIDO)

  $\beta_1 =$ declive de $E(Y_i|x_i) = \beta_0 + \beta_1 x_i$ (DESCONHECIDO)

  σ^2 (DESCONHECIDO)

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ vs. $H_1 : \beta_1 \neq \beta_{1,0}$

- **Estatística de teste**

$$
T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}
$$

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), pelo que a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base em intervalo para o p-value)**

  Atendendo a que a estimativa de σ^2 é igual a

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{12-2}\left(23378.91667 - 2.416666667^2 \times 3840\right) \\
&= 95.225
\end{aligned}$$

  e tendo em conta que

  $\hat{\beta}_1 = 2.416666667$

  $\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 3840,$

  o valor observado da estatística de teste é dado por

$$\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{2.416666667 - 0}{\sqrt{\frac{95.225}{3840}}} \\
&\simeq 15.346413.
\end{aligned}$$

  Uma vez que este teste é bilateral[13] temos:

$$\begin{aligned}
p - value &= P(|T| > |t| \mid H_0) \\
&= 2 \times P(T > |t| \mid H_0) \\
&= 2 \times P(T > 15.346413 \mid H_0) \\
&= 2 \times \left[1 - F_{t_{(10)}}(15.346413)\right].
\end{aligned}$$

  Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o *p-value* deste teste. Com efeito, ao enquadrarmos convenientemente $|t| = 15.346413$, obtemos sucessivamente

$$\begin{aligned}
F^{-1}_{t_{(10)}}(0.9995) = 4.587 &< 15.346413 \\
1 - F_{t_{(10)}}(15.35) &< 1 - 0.9995 \\
2 \times \left[1 - F_{t_{(10)}}(15.35)\right] &< 2 \times (1 - 0.9995) \\
p - value &< 0.001.
\end{aligned}$$

  Consequentemente, devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 0.1\%$, por exemplo, a qualquer dos níveis usuais de significância $(1\%, 5\%, 10\%)$, resultado este que se

[13] E a f.d.p. da estatística de teste é, sob H_0, simétrica em relação à origem.

coaduna com o muito bom ajustamento do modelo ao conjunto de dados, como tivemos ocasião de referir na alínea (c).

- **Decisão (com base no valor do p-value)**

  Ao recorrer à máquina de calcular obteríamos $p - value \simeq 2.8 \times 10^{-8}$, logo deveríamos rejeitar H_0 a qualquer n.s. $\alpha_0 > 2.8 \times 10^{-6}$, nomeadamente a qualquer dos níveis usuais de significância $(1\%, 5\%, 10\%)$, resultado este, que como se referiu acima, se coaduna com o muito bom ajustamento do modelo ao conjunto de dados.

- **Interesse da hipótese nula**

  Rejeitar $H_0 : \beta_1 = \beta_{1,0} = 0$ é sinónimo de rejeitar a hipótese de inexistência de relação linear entre a resistência da peça (Y) e o tempo (x) que decorre entre a conclusão da moldagem e a medição da resistência da peça, i.e., significa não rejeitar a hipótese de existência de relação linear entre Y e x.

(e) Calcule o intervalo de confiança a 95% para o valor esperado da resistência obtida 48 horas depois de concluída a moldagem. Acha legítimo usar o mesmo procedimento tratando-se de um período de 10 horas em vez de 48 horas? Justifique a sua resposta.

- **Outro parâmetro desconhecido**

  $E(Y|x_0) = \beta_0 + \beta_1 \times x_0$, valor esperado da resistência obtida x_0 horas depois de concluída a moldagem.

- **Passo 1 — V.a. fulcral para $E(Y|x_0) = \beta_0 + \beta_1 x_0$**

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}$$

- **Passo 2 — Quantis de probabilidade**

  Já que $(1 - \alpha) \times 100\% = 95\%$ temos $\alpha = 0.05$ e lidaremos com dois quantis simétricos $a_\alpha = -b_\alpha$ iguais a:

$$\pm F^{-1}_{t_{(n-2)}}(1 - \alpha/2) = \pm F^{-1}_{t_{(12-2)}}(1 - 0.05/2) \stackrel{tabela}{=} \pm 2.228.$$

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$\ldots$$

$$P\left\{(\hat\beta_0+\hat\beta_1 x_0) - F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(\bar x-x_0)^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right]}\right.$$
$$\left.\leq \beta_0+\beta_1 x_0 \leq (\hat\beta_0+\hat\beta_1 x_0) + F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(\bar x-x_0)^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right]}\right\} =$$

- **Passo 4 — Concretização**

A expressão geral do intervalo é

$$IC_{(1-\alpha)\times 100\%}(\beta_0+\beta_1 x_0)$$
$$= \left[\ (\hat\beta_0+\hat\beta_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(\bar x-x_0)^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right]}\ \right],$$

e a sua concretização para $(1-\alpha)\times 100\% = 95\%$ e $x_0 = 48$ igual a

$$\begin{aligned} IC_{95\%}(\beta_0+\beta_1\times 48) &\stackrel{(a)}{=} [(153.9166667+2.416666667\times 48) \\ &\qquad \pm 2.228\times\sqrt{95.225\times\left[\frac{1}{12}+\frac{(48-48)^2}{3840}\right]}\,] \\ &= [269.9166667 \pm 2.228\times\sqrt{7.935416}] \\ &= [263.640419,\ 276.192914]. \end{aligned}$$

- **Comentário**

$x_0 = 10$ não pertence à gama de valores da variável explicativa, $[x_{(1)}, x_{(12)}] = [16, 80]$, pelo que não seria legítimo recorrer ao mesmo procedimento para obter uma estimativa intervalar de $E(Y|x_0 = 10)$ — estaríamos a cometer um ERRO DE EXTRAPOLAÇÃO...

Exemplo/Exercício 9.54 (Exercício 9.2, colectânea de exercícios) — **Regressão linear simples; estimativa pontual (e intervalar) de valor esperado da resposta; coeficiente de determinação; teste à significância da regressão**

Um estudo sobre a influência da velocidade do vento (x, em km/h), na quantidade de água (Y) que se evapora por dia, em centenas de litros, na albufeira de certa barragem, a temperaturas constantes, conduziu a:

x_i	20	50	30	100	70
y_i	3	5	3	10	8

(a) Adoptando um modelo de regressão linear simples, estime a recta de regressão de Y sobre x e obtenha uma estimativa da quantidade média de água evaporada quando a velocidade do vento é igual a $90km/h$. Faça uso dos seguintes valores: $\sum_i x_i = 270$, $\sum_i x_i^2 = 18700$, $\sum_i y_i = 29$, $\sum_i y_i^2 = 207$, $\sum_i x_i y_i = 1960$.

- **Modelo de RLS**

  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

  $Y_i =$ quantidade de água que se evapora no dia i

  $x_i =$ velocidade do vento no dia i

  $\epsilon_i =$ erro aleatório associado à medição da quantidade de água que se evapora no dia i

- **Hipóteses de trabalho**

  $\epsilon_i \sim_{i.i.d.}$ Normal$(0, \sigma^2)$, $i = 1, \ldots, n$ (hipótese de trabalho)

  $\beta_0 =$ ordenada na origem de $E(Y_i|x_i) = \beta_0 + \beta_1 x_i$ (DESCONHECIDO)

  $\beta_1 =$ declive de $E(Y_i|x_i) = \beta_0 + \beta_1 x_i$ (DESCONHECIDO)

  σ^2 (DESCONHECIDO)

- **Estimativas de MV de β_0 e β_1**

  Atendendo a que

  $n = 5$

  $\sum_{i=1}^n x_i = 270$

  $\bar{x} = \frac{270}{5} = 54$

  $\sum_{i=1}^n x_i^2 = 18700$

  $\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 18700 - 5 \times 54^2 = 4120$

$\sum_{i=1}^{n} y_i = 29$

$\bar{y} = \frac{29}{5} = 5.8$

$\sum_{i=1}^{n} y_i^2 = 207$

$\sum_{i=1}^{n} y_i^2 - n\,(\bar{y})^2 = 207 - 5 \times 5.8^2 = 38.8$

$\sum_{i=1}^{n} x_i y_i = 1960$

$\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 1960 - 5 \times 54 \times 5.8 = 394,$

as estimativas de MV para este modelo são iguais a:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2} \\
&= \frac{394}{4120} \\
&= 0.095631 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= 5.8 - 0.095631 \times 54 \\
&= 0.635922
\end{aligned}$$

- **Recta de regressão**

  Para $x \in [x_{(1)}, x_{(n)}] = [20, 100]$, tem-se

  $$\hat{E}(Y|x) = \hat{\beta}_0 + \hat{\beta}_1 \times x = 0.635922 + 0.095631 \times x;$$

  logo a estimativa do valor esperado da quantidade de água evaporada quando a velocidade do vento é igual a $x = 90km/h$ é dada por

  $$\hat{E}(Y|x = 90) = \hat{\beta}_0 + \hat{\beta}_1 \times 90 = 0.635922 + 0.095631 \times 90 = 9.242712.$$

(b) Calcule o coeficiente de determinação do modelo estimado.

- **Coeficiente de determinação**

  Tirando partido dos valores obtidos anteriormente, tem-se

$$\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{(394)^2}{4120 \times 38.8} \\
&= 0.971099,
\end{aligned}$$

i.e., cerca de 97.11% da variação total dos dados referentes à variável resposta Y é explicada pelo modelo de regressão, o que indicia um excelente ajustamento da recta estimada ao nosso conjunto de dados.

(c) Teste a significância da regressão. Indique o valor-p desse teste e comente o resultado face ao valor obtido na alínea anterior. (Exame 5 Fev 2002)

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$ (i.e. não existe associação linear entre Y e x) vs. $H_1 : \beta_1 \neq \beta_{1,0}$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), pelo que a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base em intervalo para o p-value)**

  Atendendo a que a estimativa de σ^2 é igual a

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{5-2}\left(38.8 - 0.095631^2 \times 4120\right) \\
&= 0.373804
\end{aligned}$$

  e tendo em conta que

  $\hat{\beta}_1 = 0.095631$

  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 4120,$

  o valor observado da estatística de teste é dado por

$$\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{0.095631 - 0}{\sqrt{\frac{0.373804}{4120}}} \\
&= 10.039806.
\end{aligned}$$

Uma vez que este teste é bilateral[14] temos:

$$\begin{aligned} p - value &= P(|T| > |t| \mid H_0) \\ &= 2 \times P(T > |t| \mid H_0) \\ &= 2 \times P(T > 10.039806 \mid H_0) \\ &= 2 \times \left[1 - F_{t_{(3)}}(10.039806)\right]. \end{aligned}$$

Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o *p-value* deste teste. Com efeito, ao enquadrarmos convenientemente $|t| = 10.039806$, obtemos sucessivamente

$$\begin{aligned} F^{-1}_{t_{(8)}}(0.995) = 5.841 &< 10.039806 < 10.21 = F^{-1}_{t_{(3)}}(0.999) \\ 0.002 = 2 \times (1 - 0.999) &< p - value < 2 \times (1 - 0.995) = 0.010. \end{aligned}$$

Consequentemente

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.2\%$;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 1.0\%$, por exemplo, a qualquer dos níveis usuais de significância $(1\%, 5\%, 10\%)$, resultado este que se coaduna com o excelente ajustamento do modela ao conjunto de dados, como tivemos ocasião de referir na alínea anterior.

Obs. — Recorrendo àmáquina de calcular obtemos $p - value = 0.002103$. Deste modo: não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.2\%$; devemos rejeitar H_0 a qualquer n.s. $\alpha_0 > 0.2\%$, nomeadamente a qualquer dos níveis usuais de significância $(1\%, 5\%, 10\%)$.

(a)′ Obtenha um intervalo de confiança a 95% para o valor esperado da quantidade de água evaporada quando a velocidade do vento é igual a $x = 90km/h$.

- **Outro parâmetro desconhecido**

  $E(Y|x_0) = \beta_0 + \beta_1 \times x_0$, valor esperado da quantidade de água evaporada quando a velocidade do vento é igual a $x_0 = 90m/s$.

- **Passo 1 — V.a. fulcral para** $E(Y|x_0) = \beta_0 + \beta_1 x_0$

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}$$

[14]E a f.d.p. da estatística de teste é, sob H_0, simétrica em relação à origem.

- **Passo 2 — Quantis de probabilidade**

  Já que $(1-\alpha)\times 100\% = 95\%$ temos $\alpha = 0.05$ e lidaremos com dois quantis simétricos $a_\alpha = -b_\alpha$ iguais a:

$$\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) = \pm F^{-1}_{t_{(5-2)}}(1-0.05/2) \stackrel{tabela}{=} \pm 3.182.$$

- **Passo 3 — Inversão da desigualdade** $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$\ldots$$

$$P\left\{(\hat\beta_0+\hat\beta_1 x_0) - F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2}\right]}\right.$$
$$\left. \leq \beta_0 + \beta_1 x_0 \leq (\hat\beta_0+\hat\beta_1 x_0) + F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2}\right]}\right\}$$
$$= 1-\alpha$$

- **Passo 4 — Concretização**

  A expressão geral do intervalo é

$$IC_{(1-\alpha)\times 100\%}(\beta_0+\beta_1 x_0)$$
$$= \left[\ (\hat\beta_0+\hat\beta_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2}\right]}\ \right],$$

  e a sua concretização para $(1-\alpha)\times 100\% = 95\%$ e $x_0 = 90$ igual a

$$\begin{aligned} IC_{95\%}(\beta_0+\beta_1\times 90) &\stackrel{(a)}{=} \left[9.242712 \pm 3.182 \times \sqrt{0.373804 \times \left[\frac{1}{5} + \frac{(54-90)^2}{4120}\right.}\right. \\ &= [9.242712 \pm 3.182 \times \sqrt{0.192346}] \\ &= [7.847173, 10.638249]. \end{aligned}$$

•

(c)$'$ Confronte as hipóteses $H_0 : \beta_1 = \beta_{1,0} = 0.2$ e $H_1 : \beta_1 < \beta_{1,0}$ ao n.s. de 5%.

Referências

Casella, G. e Berger, R.L. (2002). *Statistical inference.* Duxbury – Thomson Learning.

Montgomery, D.C. e Runger, G.C. (2003). *Applied Statistics and Probability for Engineers.* John Wiley & Sons, New York. 3a. Edição.

Murteira, B.J.F. (1990). *Probabilidades e à Estatística*, Vols. I e II (2a. Edição). McGraw-Hill, Lisboa.

Paulino, C.D. (1994). *Notas de Análise Preliminar de Dados Univariados.* Reprografia do IST, Lisboa.

Pestana, D.D. e Velosa, S.F. (2002). *Introdução à Probabilidade e à Estatística.* Fundação Calouste Gulbenkian, Lisboa.

Rohatgi, V.K. e Saleh, A.K.Md.E. (2001). *An Introduction to Probability and Statistics.* John Wiley & Sons, Inc.

Formulário

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$$
$$x=0,1,\ldots,n$$
$$E(X)=np \qquad Var(X)=np(1-p)$$

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$$
$$x=0,1,\ldots$$
$$E(X)=Var(X)=\lambda$$

$$P(X=x)=p(1-p)^{x-1}$$
$$x=1,2,\ldots$$
$$E(X)=\frac{1}{p} \qquad Var(X)=\frac{(1-p)}{p^2}$$

$$P(X=x)=\binom{M}{x}\binom{N-M}{n-x}\bigg/\binom{N}{n}$$
$$x=\max\{0,n-N+M\},\ldots,\min\{n,M\}$$
$$E(X)=n\frac{M}{N} \qquad Var(X)=n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$$

$$f_X(x)=\frac{1}{b-a},\ a\le x\le b$$
$$E(X)=\frac{b+a}{2} \qquad Var(X)=\frac{(b-a)^2}{12}$$

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ x\in I\!R$$
$$E(X)=\mu \qquad Var(X)=\sigma^2$$

$$f_X(x)=\lambda e^{-\lambda x},\ x\ge 0$$
$$E(X)=\frac{1}{\lambda} \qquad Var(X)=\frac{1}{\lambda^2}$$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$$

$$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}\overset{a}{\sim} N(0,1)$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

$$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\overset{a}{\sim} N(0,1)$$

$$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{(n_1+n_2-2)}$$

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}$$

$$\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}\overset{a}{\sim}\chi^2_{(k-\beta-1)}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\overset{a}{\sim}\chi^2_{(r-1)(s-1)}$$

$$Y_i=\beta_0+\beta_1x_i+\varepsilon_i$$

$$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1=\frac{\sum\limits_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}$$

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2,\ \hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$$

$$\hat{\sigma}^2=\frac{1}{n-2}\left[\left(\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2\right)-\left(\hat{\beta}_1\right)^2\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\right]$$

$$\frac{\hat{\beta}_0-\beta_0}{\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}}\sim t_{(n-2)}$$

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2-n\bar{x}^2}}}\sim t_{(n-2)}$$

$$\frac{\left(\hat{\beta}_0+\hat{\beta}_1x_0\right)-(\beta_0+\beta_1x_0)}{\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}}\sim t_{(n-2)}$$

$$R^2=\frac{\left(\sum\limits_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}\right)^2}{\left(\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2\right)\times\left(\sum\limits_{i=1}^{n}Y_i^2-n\bar{Y}^2\right)}$$