



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Departamento de Matemática

IPEIO

Introdução às Probabilidades e Estatística

Luís Ramos
Pedro Mota

2016/17

Observação:

Estas folhas servem de apoio às aulas de IPEIO, Módulo de Probabilidades e Estatística. Para uma melhor compreensão dos assuntos abordados, aconselha-se a leitura de alguns dos livros indicados nas referências bibliográficas.

Conteúdo

| | | |
|----------|--|-----------|
| 1 | Introdução à Teoria da Probabilidade | 3 |
| 1.1 | Espaço de Resultados e Acontecimentos | 3 |
| 1.2 | Probabilidade | 5 |
| 1.3 | Cálculo Combinatório | 7 |
| 1.4 | Probabilidade Condicional e Independência | 7 |
| 2 | Variáveis aleatórias | 11 |
| 2.1 | Variáveis aleatórias | 11 |
| 2.2 | Função de distribuição | 12 |
| 2.3 | Classificação das variáveis aleatórias | 13 |
| 2.4 | Momentos | 16 |
| 2.5 | Outros parâmetros relevantes | 18 |
| 2.6 | Covariância | 21 |
| 3 | Principais Distribuições | 23 |
| 3.1 | Distribuições discretas | 23 |
| 3.1.1 | Distribuição de Bernoulli | 23 |
| 3.1.2 | Distribuição Binomial | 23 |
| 3.1.3 | Distribuição Hipergeométrica | 26 |
| 3.1.4 | Distribuição de Poisson | 27 |
| 3.2 | Distribuições Contínuas | 29 |
| 3.2.1 | Distribuição Exponencial | 29 |
| 3.2.2 | Distribuição Normal | 31 |
| 3.2.3 | Distribuição do Qui Quadrado | 32 |
| 3.2.4 | Distribuição t de Student | 33 |
| 4 | Teorema Limite Central | 35 |
| 5 | Estimação Pontual | 37 |
| 5.1 | Alguns conceitos importantes | 37 |
| 5.2 | Propriedades dos estimadores | 38 |
| 5.3 | Distribuições por Amostragem | 40 |
| 5.3.1 | Distribuição por amostragem da média amostral, \bar{X} | 40 |
| 5.3.2 | Distribuição por amostragem da variância amostral, S^2 | 41 |
| 5.3.3 | Distribuição por amostragem da proporção, \hat{P} | 42 |

| | | |
|----------|--|-----------|
| 6 | Estimação por Intervalo de Confiança | 45 |
| 6.1 | Intervalo de Confiança para a média da população, μ | 46 |
| 6.1.1 | População Normal com variância conhecida | 46 |
| 6.1.2 | População Normal com variância desconhecida | 49 |
| 6.1.3 | População desconhecida com variância conhecida e $n > 30$ | 50 |
| 6.1.4 | População desconhecida com variância desconhecida e $n > 30$ | 51 |
| 6.2 | Intervalo de Confiança para a variância populacional, σ^2 , e para o desvio padrão populacional, σ | 52 |
| 6.3 | Intervalo de Confiança para proporção populacional, p | 54 |
| 7 | Teste de Hipóteses | 57 |
| 7.1 | Introdução | 57 |
| 7.2 | Teste de Hipóteses para a média da população | 59 |
| 7.2.1 | Teste bilateral | 60 |
| 7.2.2 | Teste unilateral direito | 62 |
| 7.2.3 | Teste unilateral esquerdo | 63 |
| 7.3 | Teste de Hipóteses para a variância, σ^2 , de uma população Normal | 63 |
| 7.4 | Teste de Hipóteses para a proporção p de uma população | 64 |
| 8 | Regressão Linear | 67 |
| 8.1 | Introdução | 67 |
| 8.2 | Estimadores dos Mínimos Quadrados de β_0 e β_1 | 68 |
| 8.3 | Estimação de σ^2 e Qualidade do Ajuste | 69 |
| 8.4 | Propriedades dos estimadores dos mínimos quadrados | 69 |
| 8.4.1 | Distribuição por amostragem de $\hat{\sigma}^2$ | 69 |
| 8.4.2 | Distribuição por amostragem de $\hat{\beta}_0$ e $\hat{\beta}_1$ | 70 |
| 8.5 | Inferência sobre os parâmetros do Modelo de Regressão | 71 |
| 8.5.1 | Intervalo de Confiança e Teste de Hipóteses para β_1 | 71 |
| 8.5.2 | Intervalo de Confiança e Teste de Hipóteses para β_0 | 72 |
| 8.5.3 | Intervalo de Confiança e Teste de Hipóteses para σ^2 | 73 |
| 8.6 | Estimação do valor esperado de Y para uma observação x_0 da variável controlada | 77 |
| 8.7 | Previsão do valor da variável resposta Y para um novo valor x_0 da variável controlada | 77 |

Capítulo 1

Introdução à Teoria da Probabilidade

A teoria das probabilidades tem como objetivo a formulação de modelos de fenómenos em que intervém o acaso. Fenómenos aleatórios são os fenómenos sujeitos à influência do acaso, não sendo por isso controláveis pelo homem.

1.1 Espaço de Resultados e Acontecimentos

Definição 1.1 (Experiência aleatória). *Uma experiência aleatória é uma experiência cujo resultado é desconhecido (antes da sua realização), apesar de se conhecerem todos os possíveis resultados.*

Exemplo 1.2 (Experiência aleatória). *Considere os seguintes exemplos:*

- E_1 : Lançamento de uma moeda e observação da face voltada para cima;
- E_2 : Lançamento de um dado e observação da face voltada para cima;
- E_3 : Tempo de "vida" de uma lâmpada.
- E_4 : Tempo de vida de uma pessoa em anos.

Definição 1.3 (Espaço de resultados ou universo). *Chamamos espaço de resultados ou universo, e representamos por Ω , ao conjunto de todos os possíveis resultados de uma experiência aleatória.*

Observação: Diz-se que o espaço de resultados, Ω , é discreto se tem um número finito ou infinito numerável de elementos. Se Ω contém um intervalo (finito ou infinito) de números reais, então o espaço de resultados é contínuo.

Exemplo 1.4 (Espaço de resultados). *Considere novamente as experiências aleatórias do Exemplo 1.2. Temos:*

- E_1 : $\Omega = \{Cara, Coroa\}$;

- $E_2 : \Omega = \{1, 2, 3, 4, 5, 6\}$;
- $E_3 : \Omega = \mathbb{R}^+$;
- $E_4 : \Omega = \{1, 2, \dots\} = \mathbb{N}$;

Exemplo 1.5 (Espaço de resultados). Na experiência aleatória que consiste em lançar um dado, numerado de 1 a 6, e observar a face voltada para cima, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Se forem lançados dois dados, o espaço de resultados é,

$$\Omega = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \dots, (6, 5), (6, 6)\},$$

ou seja, $\Omega = \{(i, j) : i = 1, \dots, 6; j = 1, \dots, 6\}$.

Definição 1.6 (Acontecimento e Acontecimento elementar). Um acontecimento é um subconjunto do espaço de resultados, Ω . Cada acontecimento formado por apenas um ponto amostral é designado por **acontecimento elementar** ou **simples**.

Observação: Ao conjunto \emptyset chamamos acontecimento impossível e a Ω acontecimento certo.

Definição 1.7 (Sub-acontecimento). A é sub-acontecimento de B , e escreve-se $A \subset B$, se e só se a realização de A implica a realização de B .

Observação: Podemos aplicar as operações usuais sobre conjuntos de modo a obter outros acontecimentos de interesse. As operações mais usuais são:

- A **união** de dois acontecimentos A e B , e representa-se por $A \cup B$;
- A **intersecção** de dois acontecimentos A e B , e representa-se por $A \cap B$;
- O **complementar** do acontecimento A e representa-se por \bar{A} ;
- A **diferença** dos acontecimentos A e B e representa-se por $A - B (= A \cap \bar{B})$;

Algumas propriedades importantes:

1. Distributiva: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ e $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
2. Leis de **De Morgan**: $\overline{A \cap B} = \bar{A} \cup \bar{B}$ e $\overline{A \cup B} = \bar{A} \cap \bar{B}$.

Definição 1.8 (Acontecimentos disjuntos ou mutuamente exclusivos). Dois acontecimentos A e B dizem-se disjuntos se não têm elementos em comum, ou seja, se $A \cap B = \emptyset$.

1.2 Probabilidade

Em muitas experiências aleatórias estamos interessados em medir a possibilidade de ocorrer um determinado acontecimento. A probabilidade permite-nos quantificar essa possibilidade.

Definição 1.9 (Definição Clássica ou de Laplace de Probabilidade). *Se uma experiência aleatória tem a si associado um número finito N de resultados, mutuamente exclusivos e igualmente prováveis, então a probabilidade de qualquer acontecimento A , $P(A)$, é dada por:*

$$P(A) = \frac{N_A}{N} = \frac{n^\circ \text{ de resultados favoráveis a } A}{n^\circ \text{ de resultados possíveis}}.$$

Exemplo 1.10. *A probabilidade de sair face ímpar, num lançamento de um dado equilibrado é $P(\text{"Sair face ímpar"}) = \frac{3}{6} = \frac{1}{2}$.*

Definição 1.11 (Definição Frequencista de Probabilidade). *A probabilidade de um acontecimento A é dada pelo limite da frequência relativa com que se observou A , isto é,*

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n},$$

onde n_A representa o número de observações de A , e n o número de realizações da experiência aleatória. Para valores elevados de n , podemos assumir que $P(A) \approx \frac{n_A}{n}$.

Definição 1.12 (Definição Axiomática de Probabilidade). *A Probabilidade é uma função, que a cada acontecimento A faz corresponder um valor real, $P(A)$, e que verifica as seguintes condições ou axiomas:*

1. $P(A) \geq 0$, qualquer que seja o acontecimento A ;
2. $P(\Omega) = 1$;
3. Se A_1, A_2, \dots são acontecimentos disjuntos dois a dois, então $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Proposição 1.13. *Sejam A e B dois acontecimentos. Os seguintes resultados são consequência imediata dos axiomas da definição 1.12:*

1. $P(\emptyset) = 0$;
2. Se $A \subseteq B$ então $P(A) \leq P(B)$;
3. $P(\bar{A}) = 1 - P(A)$;
4. $P(A) \in [0, 1]$;
5. $P(A - B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$;

$$6. P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demonstração.

1. Como \emptyset e Ω são acontecimentos disjuntos e $P(\emptyset \cup \Omega) = P(\Omega) = 1$, resulta pelo 3º axioma que $P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega)$, ou seja, $P(\emptyset) = 0$.

2. Sejam A e B dois acontecimentos tais que $A \subseteq B$. Então $B = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A})$. Como A e $B \cap \bar{A}$ são acontecimentos disjuntos, podemos utilizar o 3º axioma, resultando,

$$P(B) = P(A \cup (B \cap \bar{A})) = P(A) + P(B \cap \bar{A}).$$

Usando o 1º axioma, podemos garantir que $P(B \cap \bar{A}) \geq 0$ e conseqüentemente $P(B) \geq P(A)$.

3. Como A e \bar{A} são acontecimentos disjuntos, podemos utilizar o 3º axioma. Assim,

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

ou seja, $P(\bar{A}) = 1 - P(A)$.

4. Pelo 1º axioma, para qualquer acontecimento A , $P(A) \geq 0$. Logo, basta apenas demonstrar que $P(A) \leq 1$. Como $A \subseteq \Omega$, resulta que $P(A) \leq P(\Omega) = 1$.

5. Como $A = (A \cap \bar{B}) \cup (A \cap B) = (A - B) \cup (A \cap B)$, e $(A - B)$ e $(A \cap B)$ são acontecimentos disjuntos, então podemos utilizar o 3º axioma. Assim,

$$P(A) = P(A - B) + P(A \cap B) \quad \Leftrightarrow \quad P(A - B) = P(A) - P(A \cap B).$$

6. Como $A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$ e $(A - B)$, $(B - A)$ e $(A \cap B)$ são acontecimentos disjuntos dois a dois, podemos utilizar o resultado do 3º axioma, obtendo:

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

□

Observação: O último resultado da Proposição 1.13 pode ser generalizado para a união de n acontecimentos ($n \geq 2$). Assim, dados os acontecimentos A_i , $i = 1, \dots, n$,

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(\cap_{i=1}^n A_i);$$

Para $n = 3$ obtemos o caso particular:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Definição 1.14 (Acontecimentos incompatíveis). Dois acontecimentos A e B dizem-se incompatíveis se $P(A \cap B) = 0$.

1.3 Cálculo Combinatório

O cálculo de uma probabilidade, através da definição clássica, depende da contagem do número de casos favoráveis e do número de casos possíveis. Em muitas situações este cálculo pode não ser imediato. O cálculo combinatório é uma ferramenta que nos poderá auxiliar em muitas dessas situações.

Definição 1.15 (Produto Cartesiano). *Seja $A = \{a_1, \dots, a_n\}$ um conjunto com n elementos e $B = \{b_1, \dots, b_m\}$ um conjunto com m elementos. Designa-se por produto cartesiano o conjunto de pares (a_i, b_j) em que o primeiro provém de A e o segundo de B e representa-se por $A \times B$. O número de elementos de $A \times B$ é dados por $\#(A \times B) = n \times m$.*

Considere agora que temos n elementos distintos, e pretendemos seleccionar k . De quantas maneiras distintas é possível seleccionar os k elementos? Como existem várias formas distintas de escolher os k elementos, a resposta à questão anterior é dada pela seguinte tabela:

| Interessa a ordem? | Há repetição? | Designação | Número de maneiras distintas de escolher os k elementos |
|--------------------|---------------|---------------------------|---|
| Sim | Não | Arranjos | ${}^n A_k = \frac{n!}{(n-k)!}, \quad k \leq n$ |
| Sim | Sim | Arranjos com repetição | ${}^n A'_k = n^k$ |
| Não | Não | Combinações | ${}^n C_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}, \quad k \leq n$ |
| Não | Sim | Combinações com repetição | ${}^n C'_k = \frac{(n+k-1)!}{(n-1)!k!}$ |

Observações:

- “!” representa a função factorial (por convenção $0! = 1$);
- No caso particular em que interessa a ordem, não há repetição e estamos a seleccionar todos os elementos disponíveis ($k = n$), é mais usual designarmos **Permutações** de n elementos, P_n , em vez de ${}^n A_n$. É obvio que ${}^n A_n = P_n = n!$.

1.4 Probabilidade Condicional e Independência

Vamos começar por um exemplo que irá introduzir a noção de probabilidade condicional.

Exemplo 1.16. *Uma empresa farmacêutica realizou um ensaio clínico para comparar a eficácia de um novo medicamento (medicamento experimental). Escolheram-se ao acaso 200 doentes com a doença que se pretende curar. Metade desses doentes foram tratados com o novo medicamento e os restantes com um medicamento convencional. Ao fim de 5 dias, os resultados são os seguintes:*

| | Melhorou (M) | Não melhorou (\bar{M}) | Total |
|--|------------------|----------------------------|-------|
| Medicamento Experimental E | 69 | 31 | 100 |
| Medicamento Convencional (\bar{E}) | 58 | 42 | 100 |
| Total | 127 | 73 | 200 |

1. Qual a probabilidade, de um doente escolhido ao acaso,

(a) tomar o medicamento experimental?

Resposta: Usando a regra de Laplace, $P(E) = \frac{100}{200} = \frac{1}{2}$.

(b) tomar o medicamento experimental e melhorar?

Resposta: Usando a regra de Laplace, $P(E \cap M) = \frac{69}{200}$.

2. Qual a probabilidade de um doente, que melhorou, ter tomado o medicamento experimental?

Resposta: $\frac{69}{127}$.

Observação: A solução da pergunta 2, do exemplo anterior, é igual a $\frac{P(E \cap M)}{P(M)}$.

Definição 1.17 (Probabilidade Condicional). Sejam A e B dois acontecimentos. A probabilidade condicional de A dado B é

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{se } P(B) > 0.$$

Teorema 1.18 (Teorema da Probabilidade Composta). Sejam A e B dois acontecimentos tais que $P(B) > 0$. Então, resulta da definição de Probabilidade Condicional,

$$P(A \cap B) = P(A|B)P(B).$$

Observação: Nalguns casos, a probabilidade condicional $P(A|B)$ pode ser igual a $P(A)$, ou seja, o conhecimento da ocorrência de B não afecta a probabilidade de A ocorrer.

Definição 1.19 (Acontecimentos Independentes). Dois acontecimentos A e B dizem-se independentes se e só se,

$$P(A \cap B) = P(A)P(B).$$

Definição 1.20 (Partição do espaço de resultados). Dizemos que $\{E_1, \dots, E_n\}$ é uma partição do espaço de resultados Ω quando

$$E_i \cap E_j = \emptyset \quad (i \neq j) \quad \text{e} \quad \bigcup_{i=1}^n E_i = \Omega.$$

Teorema 1.21 (Teorema da Probabilidade Total). *Seja $\{E_1, \dots, E_n\}$ uma partição do espaço de resultados Ω , com $P(E_i) > 0, \forall i$. Dado um qualquer acontecimento A , tem-se,*

$$P(A) = P(A|E_1)P(E_1) + \dots + P(A|E_n)P(E_n).$$

Demonstração. Repare-se que devido a $\{E_1, \dots, E_n\}$ formarem uma partição de Ω , teremos que $(A \cap E_i) \cap (A \cap E_j) = \emptyset$ ($i \neq j$) e aplicando o Teorema 1.18 da Probabilidade Composta:

$$P(A) = P\left(A \cap \left(\bigcup_{i=1}^n E_i\right)\right) = P\left(\bigcup_{i=1}^n (A \cap E_i)\right) = \sum_{i=1}^n P(A \cap E_i) = \sum_{i=1}^n P(A|E_i)P(E_i).$$

□

Teorema 1.22 (Teorema de Bayes). *Seja $\{E_1, \dots, E_n\}$ uma partição do espaço de resultados Ω , com $P(E_i) > 0, \forall i$. Dado um qualquer acontecimento A , com $P(A) > 0$, tem-se*

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{j=1}^n P(A|E_j)P(E_j)}.$$

Demonstração. Aplicando a definição 1.17, de Probabilidade Condicional, depois o Teorema 1.18 da Probabilidade Composta e o Teorema 1.21 da Probabilidade Total,

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum_{j=1}^n P(A|E_j)P(E_j)}.$$

□

Exemplo 1.23 (Teste de P.E. D - 2007/08). *Diga, justificando, se a seguinte afirmação é verdadeira ou falsa:*

Três máquinas A, B e C produzem botões, respectivamente, 15%, 25% e 60% da produção total. As percentagens de botões defeituosos fabricados por estas máquinas são respectivamente 5%, 7% e 4%. Se ao acaso, da produção total de botões, for encontrado um defeituoso, a probabilidade de ele ter sido produzido pela máquina B é de cerca de 36%.

Resolução:

Sejam A, B, C e D os seguintes acontecimentos:

A - O Botão é produzido pela máquina A;

B - O Botão é produzido pela máquina B;

C - O Botão é produzido pela máquina C;

D - O Botão tem defeito;

De acordo com o enunciado, temos as seguintes probabilidades: $P(A) = 0.15$, $P(B) = 0.25$, $P(C) = 0.6$, $P(D|A) = 0.05$, $P(D|B) = 0.07$ e $P(D|C) = 0.04$.

Pretende-se determinar $P(B|D)$. Usando o Teorema de Bayes, obtemos:

$$P(B|D) = \frac{P(D|B)P(B)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)} = \frac{175}{490} \simeq 36\%.$$

Logo a afirmação está correcta, isto é, a probabilidade de um botão defeituoso ter sido produzido pela máquina B é de cerca de 36%.

Capítulo 2

Variáveis aleatórias

2.1 Variáveis aleatórias

Definição 2.1 (Variável aleatória). Uma variável aleatória (v.a.), $X : \Omega \rightarrow \mathbb{R}$, é uma função real e finita, tal que a imagem inversa de $] -\infty; x]$ é um acontecimento, isto é, $A_x = X^{-1}(-\infty; x] = \{\omega \in \Omega : X(\omega) \leq x\}$ com $x \in \mathbb{R}$ é um acontecimento.

Observação: É fácil de verificar que se X é uma variável aleatória e $g : \mathbb{R} \rightarrow \mathbb{R}$ uma função, então $Y = g(X)$ é também uma variável aleatória.

Exemplo 2.2 (Variável aleatória). Considere a experiência aleatória que consiste no lançamento de 2 moedas equilibradas, e registo da face voltada para cima. O espaço de resultados é

$$\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}.$$

Podemos, **por exemplo**, atribuir a cada um dos acontecimentos elementares de Ω , os seguintes valores:

| ω | (Ca, Ca) | (Ca, Co) | (Co, Ca) | (Co, Co) |
|-------------|------------|------------|------------|------------|
| $X(\omega)$ | 2 | 1 | 1 | 0 |

Repare que

$$A_x = X^{-1}(-\infty; x] = \begin{cases} \emptyset, & x < 0 \\ \{(Co, Co)\} & 0 \leq x < 1 \\ \{(Co, Co), (Ca, Co), (Co, Ca)\} & 1 \leq x < 2 \\ \Omega & x \geq 2 \end{cases}$$

Como todas as imagens inversas, $X^{-1}(-\infty; x]$, são acontecimentos de Ω , então de acordo com a definição 2.1, X é uma variável aleatória.

Observação: Relativamente ao Exemplo 2.2, X é a aplicação que atribui a cada acontecimento de Ω o número de caras.

2.2 Função de distribuição

Definição 2.3 (Função de distribuição). A função de distribuição da v.a. X é:

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}), \quad \forall x \in \mathbb{R}.$$

Exemplo 2.4. Considere novamente o Exemplo 2.2. A função de distribuição desta v.a. é:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \leq x < 1 \\ \frac{3}{4}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

Observação: Como $F_X(x) = P(X \leq x)$, conclui-se que a função de distribuição existe sempre. Quando não existir mais do que uma v.a., pode-se representar a função de distribuição simplesmente por F .

Propriedades da função de distribuição:

1. $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow +\infty} F(x) = 1$;
2. F é contínua à direita, isto é, $\lim_{x \rightarrow a^+} F(x) = F(a)$;
3. F é não decrescente, isto é, se $x < y$, então $F(x) \leq F(y)$.

Teorema 2.5. Qualquer função F é uma função de distribuição se e só se verificar as três propriedades anteriores.

Proposição 2.6. Seja X uma v.a. com função de distribuição F . Tem-se:

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-), \quad \forall x \in \mathbb{R},$$

onde $F(x^-) = \lim_{t \rightarrow x^-} F(t)$.

Definição 2.7 (Variáveis aleatórias identicamente distribuídas). Duas variáveis aleatórias X e Y dizem-se identicamente distribuídas, se têm a mesma função de distribuição, isto é, se $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$.

Definição 2.8 (Variáveis aleatórias independentes). Duas variáveis aleatórias X e Y dizem-se independentes, se e só se,

$$P(X \leq x \cap Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall (x, y) \in \mathbb{R}^2$$

2.3 Classificação das variáveis aleatórias

A função de distribuição não é necessariamente contínua em todos os valores $x \in \mathbb{R}$. Podemos por isso classificar as variáveis aleatórias em função da continuidade da respectiva função de distribuição. Considere o conjunto de pontos de descontinuidade da função de distribuição F ,

$$D = \{a \in \mathbb{R} : P(X = a) > 0\}. \quad (2.1)$$

Definição 2.9 (Variável aleatória discreta). Uma v.a. X diz-se do tipo discreto ou simplesmente discreta se o conjunto D é quanto muito numerável, e se $P(X \in D) = 1$.

Definição 2.10 (Função de probabilidade). Seja X uma v.a. discreta. Chama-se função de probabilidade (f.p.), ou função massa de probabilidade, de X à função definida pelo conjunto dos valores de D e pelas respectivas probabilidades, isto é, por (x_i, p_i) onde $x_i \in D$ e $p_i = P(X = x_i)$. Uma representação usual para a função de probabilidade da v.a. X , é:

$$X = \begin{cases} x_1 & x_2 & \dots & x_i & \dots \\ P(X = x_1) & P(X = x_2) & \dots & P(X = x_i) & \dots \end{cases}$$


Propriedades da função de probabilidade:

1. $P(X = x_i) = f(x_i) = p_i \geq 0$;
2. $\sum_{i=1}^{\infty} p_i = 1$.

Observação: Para qualquer subconjunto real I , $P(X \in I) = \sum_{x_i \in I \cap D} P(X = x_i)$.

Exemplo 2.11. Considere novamente o Exemplo 2.2. O conjunto de pontos de descontinuidade da função de distribuição é $D = \{0, 1, 2\}$. Como $P(X \in D) = 1$, conclui-se que X é uma v.a. discreta com função de probabilidade,

$$X \begin{cases} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{cases}$$

Podemos representar a informação disponibilizada pela função de probabilidade através de um gráfico de barras que pode ser gerado em  utilizando, por exemplo, as linhas de código:

```
> x<-0:2; fx<-c(1/4,1/2,1/4); x; fx
[1] 0 1 2
[1] 0.25 0.50 0.25
> plot(x,fx, type="h", lwd=10, ylab="P(X=x)", xlab="x",
      main="função de probabilidade de X",ylim=c(0,0.5),xaxp=c(0,2,2))
```

função de probabilidade de X

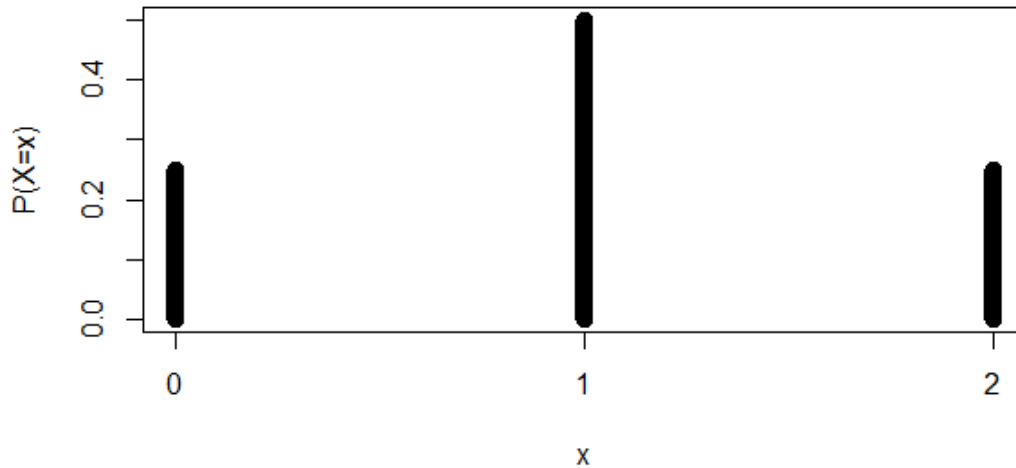


Figura 2.1: Gráfico da função de probabilidade da v.a. X .

Quer da análise da função de probabilidade ou do seu gráfico, teremos por exemplo,

$$P(X \geq 1) = P(X \in \{1, 2\}) = P(X = 1) + P(X = 2) = \frac{3}{4}.$$

Comandos **R**: "<-", "=", c() e plot().

Definição 2.12 (Variável aleatória contínua). Uma v.a. X diz-se do tipo contínuo ou simplesmente contínua se $D = \emptyset$ e se existe uma função não negativa, f , tal que para $I \subseteq \mathbb{R}$,

$$P(X \in I) = \int_I f(x) dx.$$

À função f chamamos **função densidade probabilidade** ou **função densidade**.

Propriedades da função densidade probabilidade:

1. $f(x) \geq 0, \quad \forall x \in \mathbb{R};$
2. $\int_{-\infty}^{+\infty} f(x) dx = 1.$

Observação: Como $\int_I f(x) dx$ é um integral de uma função não negativa e é sempre convergente, então a $P(X \in I)$, corresponde ao valor da área entre o eixo das abcissas e o gráfico da função f no conjunto I considerado. Consequentemente $P(X = x) = 0, \forall x \in \mathbb{R}$ e

$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2), \quad \forall x_1 \leq x_2.$$

Observação: Por definição, $F'(x) = f(x)$, nos pontos onde existe derivada. Se não existir derivada, $f(x) = 0$.

Exemplo 2.13 (Variável aleatória contínua). *A proporção de ratos, de uma população, infectada com Leptospirose é uma variável aleatória com função densidade:*

$$f(x) = \begin{cases} k(1-x) & 0 < x < 1; \\ 0, & \text{outros valores de } x. \end{cases}$$

1. Mostre que $k = 2$.
2. Sabendo que menos de metade de uma população de ratos está infectada com Leptospirose, qual a probabilidade da proporção de ratos infectados ser superior a 0.3?
3. Determine o valor t que verifica a equação $P(X < t) = 0.5$.

Consegue-se representar graficamente a função densidade de probabilidade e calcular, por exemplo, $P(0.3 < X < 0.4)$ em **R**:

```
> fx<-function(x){ifelse(0<x & x<1,2*(1-x),0)}
> integrate(fx,0.3,0.4)$value
[1] 0.13
> plot(fx,-1,2)
```

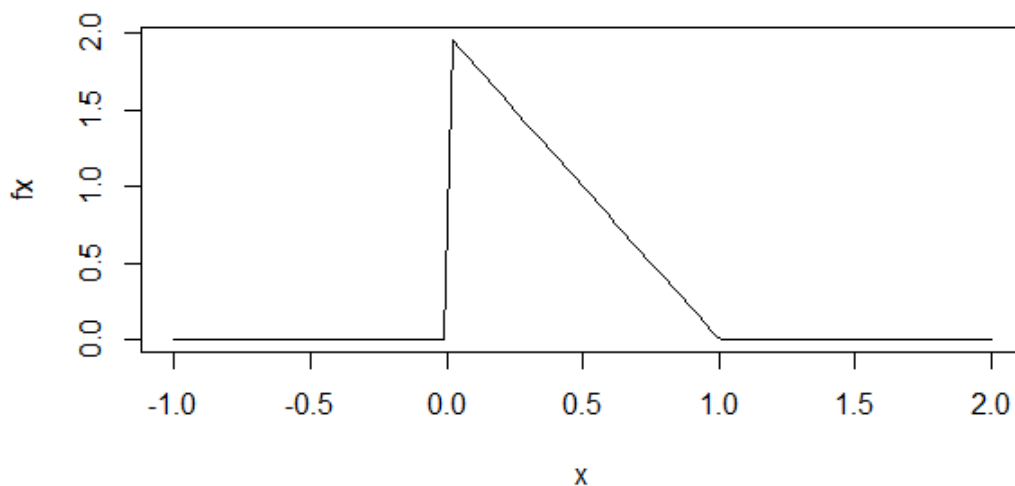


Figura 2.2: Gráfico da função densidade de probabilidade da v.a. X .

Comandos **R**: `function()`, `ifelse()` e `integrate()`.

2.4 Momentos

Definição 2.14 (Valor médio). O valor médio, valor esperado ou simplesmente média da v.a. X é dado por,

$$\mu = E(X) = \begin{cases} \sum_{i=1}^{\infty} x_i P(X = x_i) & \text{se } X \text{ é uma v.a. discreta;} \\ \int_{-\infty}^{+\infty} x f(x) dx & \text{se } X \text{ é uma v.a. contínua;} \end{cases}$$

desde que a série/integral seja absolutamente convergente.

Definição 2.15 (Valor médio de uma função de uma variável aleatória). Seja X uma v.a. e g uma função real de variável real contínua com quanto muito um conjunto numerável de pontos de descontinuidade. Então o valor médio de $Y = g(X)$ é dado por:

$$E(g(X)) = \begin{cases} \sum_{i=1}^{\infty} g(x_i) P(X = x_i) & \text{se } X \text{ é uma v.a. discreta;} \\ \int_{-\infty}^{+\infty} g(x) f(x) dx & \text{se } X \text{ é uma v.a. contínua;} \end{cases}$$

desde que a série/integral seja absolutamente convergente.

Exemplo 2.16. Considere a variável aleatória introduzida no Exemplo 2.2. Os valores médios de X e $g(X) = X^2$, são respectivamente:

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1,$$

$$E(g(X)) = E(X^2) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}.$$

Ou em **R**:

```
> x<-0:2
> fx<-c(1/4,1/2,1/4)
> x
[1] 0 1 2
> fx
[1] 0.25 0.50 0.25
> m<-sum(x*fx); m
[1] 1
> m2<-sum(x^2*fx); m2
[1] 1.5
```

Comandos **R**: `sum()`

Seja X uma v.a. para a qual existe valor esperado.

Propriedades do valor esperado:

1. Se a é uma constante, $E(a) = a$;
2. Se a e b são constantes, $E(aX + b) = aE(X) + b$.
3. Se existirem $E(g_1(X))$ e $E(g_2(X))$, então

$$E(g_1(X) + g_2(X)) = E(g_1(X)) + E(g_2(X)).$$

Definição 2.17 (Momentos de ordem k). *Seja X uma variável aleatória. Definem-se momentos de ordem k em torno da origem por:*

$$m_k = E(X^k),$$

e os momentos centrais de ordem k de X por:

$$\mu_k = E((X - \mu)^k),$$

desde que os valores esperados existam.

Definição 2.18 (Variância e desvio padrão). *A variância da v.a. X , σ^2 ou $V(X)$, é o momento central de ordem dois, isto é,*

$$\sigma^2 = V(X) = E((X - \mu)^2),$$

desde que exista o valor esperado de $(X - \mu)^2$. À sua raiz quadrada positiva, $\sigma = \sqrt{V(X)}$, chamamos desvio padrão da v.a. X .

Proposição 2.19. *Se X é uma v.a., para a qual existe variância, então $V(X) = E(X^2) - E^2(X)$.*

Seja X uma v.a. para a qual existe variância.

Propriedades da Variância:

1. Se a é uma constante, $V(a) = 0$;
2. Se a e b são constantes, $V(aX + b) = a^2V(X)$.

Exemplo 2.20. *Considere a variável aleatória introduzida no Exemplo 2.2. A variância de X é:*

$$V(X) = E((X - 1)^2) = (0 - 1)^2 \times \frac{1}{4} + (1 - 1)^2 \times \frac{1}{2} + (2 - 1)^2 \times \frac{1}{4} = \frac{1}{2}.$$

Nota: *A variância também podia ser calculada através do resultado da Proposição 2.19.*

Em :

```
> v<-sum((x-m)^2*fx); v
[1] 0.5
> m2-m^2
[1] 0.5
```

Exemplo 2.21 (Momentos de uma variável aleatória contínua). *Considere-se novamente a variável aleatória, que representa a proporção de ratos infectados, do exemplo 2.13. Os valores médios de X e $g(X) = X^2$, são respectivamente:*

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^1 2x(1-x) dx = 1/3 = 0.33333,$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 2x^2(1-x) dx = 1/6 = 0.16667.$$

Ou em **R**:

```
> m<-integrate(function(x) x*fx(x),-Inf,Inf)$value; m
[1] 0.3333333
> m2<-integrate(function(x) x^2*fx(x),-Inf,Inf)$value; m2
[1] 0.1666667
```

A variância de X é:

$$\begin{aligned} V(X) &= E((X - 1/3)^2) = \int_{-\infty}^{+\infty} (x - 1/3)^2 f(x) dx \\ &= \int_0^1 2(x - 1/3)^2(1-x) dx = 1/18 = 0.05556. \end{aligned}$$

Em **R**:

```
> v<-integrate(function(x) (x-m)^2*fx(x),-Inf,Inf)$value; v
[1] 0.0555556
```

Nota: A variância também podia ser calculada através da fórmula,

$$V(X) = E(X^2) - E^2(X)$$

2.5 Outros parâmetros relevantes

Definição 2.22 (Coeficiente de variação). *Seja X uma v.a. com suporte não negativo. O Coeficiente de variação de X é,*

$$CV = \frac{\sigma}{\mu} \times 100\%.$$

Definição 2.23 (Coeficiente de Simetria). *O Coeficiente de simetria, de uma v.a. X , é definido por*

$$\beta_1 = \frac{\mu_3}{\sigma^3}.$$

Definição 2.24 (Coeficiente de achatamento ou Kurtosis). Define-se o coeficiente de achatamento ou kurtosis como

$$\beta_2 = \frac{\mu_4}{\sigma^4} - 3.$$

Definição 2.25 (Quantil). O quantil de ordem p , χ_p , da v.a. X é a solução da equação:

$$F(\chi_p) = p \Leftrightarrow \int_{-\infty}^{\chi_p} f(x)dx = p, \quad 0 < p < 1.$$

Se X é uma v.a. discreta, a equação $F(\chi_p) = p$ pode não ter solução exacta. Nesse caso

$$\chi_p = \min\{x : F(x) \geq p\}.$$

Definição 2.26 (Mediana). Trata-se do quantil de ordem $p = 1/2$. Costuma-se representar a mediana, da v.a. X , por $\text{med}(X)$.

Definição 2.27 (Moda). A Moda, representada por m_o , é o valor que maximiza a função de probabilidade ou a função densidade probabilidade, desde que seja único.

Exemplo 2.28 (Variável aleatória discreta). No caso da v.a. discreta introduzida no exemplo 2.2:

$$CV = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{1/2}}{1} \times 100\% = \frac{\sqrt{2}}{2} \times 100\% = 70.71\%$$

$$\beta_1 = \frac{E((X-1)^3)}{\left(\frac{1}{2}\right)^{3/2}} = \frac{(0-1)^3 \times \frac{1}{4} + (1-1)^3 \times \frac{1}{2} + (2-1)^3 \times \frac{1}{4}}{\left(\frac{1}{2}\right)^{3/2}} = 0$$

$$\beta_2 = \frac{E((X-1)^4)}{\left(\frac{1}{2}\right)^2} = \frac{(0-1)^4 \times \frac{1}{4} + (1-1)^4 \times \frac{1}{2} + (2-1)^4 \times \frac{1}{4}}{\frac{1}{4}} = -1$$

Em :

```
> mu<-function(k){sum((x-m)^k*fx)}
> mu(2)
[1] 0.5
> sigma<-sqrt(mu(2)); sigma
[1] 0.7071068
> CV<-sigma/m*100; CV
[1] 70.71068
> beta1<-mu(3)/sigma^3; beta1
[1] 0
> beta2<-mu(4)/sigma^4 -3; beta2
[1] -1
```


Exemplo 2.29 (Variável aleatória contínua). No caso da v.a. contínua introduzida no exemplo 2.13:


$$CV = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{1/18}}{1/3} \times 100\% = \frac{\sqrt{2}}{2} \times 100\% = 70.71\%$$

$$\beta_1 = \frac{E((X - m)^3)}{\sigma^3} = \frac{\int_0^1 2(x - 1/3)^3(1 - x)dx}{\sigma^3} = 0.5657$$


$$\beta_2 = \frac{E((X - m)^4)}{\sigma^4} = \frac{\int_0^1 2(x - 1/3)^4(1 - x)dx}{\sigma^4} = -0.6$$

Em :

```
mu<-function(k)(integrate(function(x)(x-m)^k*fx(x),-Inf,Inf)$value)
> mu(2)
[1] 0.05555556
> sigma<-sqrt(mu(2)); sigma
[1] 0.2357023
> CV<-sigma/m*100; CV
[1] 70.71068
> beta1<-mu(3)/sigma^3; beta1
[1] 0.5656854
> beta2<-mu(4)/sigma^4 -3; beta2
[1] -0.6
```

Em , é possível introduzir informação sobre variáveis aleatórias discretas ou absolutamente contínuas. O "package"`distrEx` pode ser instalado e depois ativado com o comando `library()`.

```
> library(distrEx)
Loading required package: distr
```

Podemos usar os comandos do , `DiscreteDistribution()` ou `AbscontDistribution()`, para introduzir a informação sobre a função de probabilidade ou densidade, respectivamente. E, utilizando funções definidas no próprio programa, obter alguns dos resultados já referidos anteriormente.

| | |
|---|---|
| <pre>> X<-DiscreteDistribution(supp=0:2, prob=c(0.25,0.5,0.25)) > E(X) [1] 1 > var(X) [1] 0.5 > sd(X) [1] 0.7071068 > skewness(X) [1] 0 > kurtosis(X) [1] -1</pre> | <pre>> fx<-function(x){ifelse(0<x & x<1,2*(1-x),0)}; > X<-AbscontDistribution(d=fx) > E(X) [1] 0.333333 > var(X) [1] 0.05555569 > sd(X) [1] 0.2357026 > skewness(X) [1] 0.5656795 > kurtosis(X) [1] -0.6000013</pre> |
|---|---|

Nota: Skewness corresponde ao coeficiente de simetria.

2.6 Covariância

Definição 2.30 (Covariância). Dado duas v.a. X e Y com $\mu_X = E(X)$ e $\mu_Y = E(Y)$, define-se a covariância entre as v.a.'s X e Y por:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Proposição 2.31. Caso exista a covariância entre X e Y , esta pode ser calculada através da fórmula:

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Outras propriedades do valor médio e variância:

1. $E(X \pm Y) = E(X) \pm E(Y)$;
2. $V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X, Y)$.

Proposição 2.32. Se X e Y são independentes, então $E(XY) = E(X)E(Y)$, e consequentemente $Cov(X, Y) = 0$.

Propriedades da Covariância: Sejam X , Y , e Z v.a.'s, a , b e c constantes reais. Então:

1. $Cov(X, Y) = Cov(Y, X)$;
2. $Cov(X, X) = V(X)$;
3. $Cov(a + bX, c + dY) = bdCov(X, Y)$;
4. $Cov(aX + bY, cZ) = acCov(X, Z) + bcCov(Y, Z)$.

Capítulo 3

Principais Distribuições

3.1 Distribuições discretas

3.1.1 Distribuição de Bernoulli

Definição 3.1 (Prova de Bernoulli). Trata-se de um experiência aleatória com apenas dois resultados possíveis (que se costumam designar por “Sucesso” ou “Insucesso”).

Definição 3.2 (Distribuição de Bernoulli). É sempre possível definir uma variável aleatória X que toma o valor 1 se o resultado da experiência é “Sucesso” e 0 se é “Insucesso”. Denotando $p = P(\text{“Sucesso”}) > 0$, então a função de probabilidade de X é dada por:

$$X \begin{cases} 0 & 1 \\ 1-p & p \end{cases} \quad \text{ou} \quad P(X = x) = p^x(1-p)^{1-x}, \quad x = 0, 1, \quad 0 < p < 1.$$

Dizemos que a v.a. X segue uma distribuição de Bernoulli, de parâmetro p , e escrevemos $X \sim \text{Ber}(p)$.

Proposição 3.3. Seja a v.a. $X \sim \text{Ber}(p)$. Então

$$E(X) = p \quad \text{e} \quad V(X) = p(1-p).$$

3.1.2 Distribuição Binomial

Definição 3.4 (Distribuição Binomial). Considere-se uma sucessão de provas de Bernoulli independentes, onde em cada prova a probabilidade de “sucesso”, p , é constante. A v.a. $X =$ “número de sucessos em n provas de Bernoulli” segue uma distribuição Binomial de parâmetros n e p , e escrevemos $X \sim B(n, p)$. A função de probabilidade é:

$$P(X = x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, \dots, n, \quad 0 < p < 1.$$

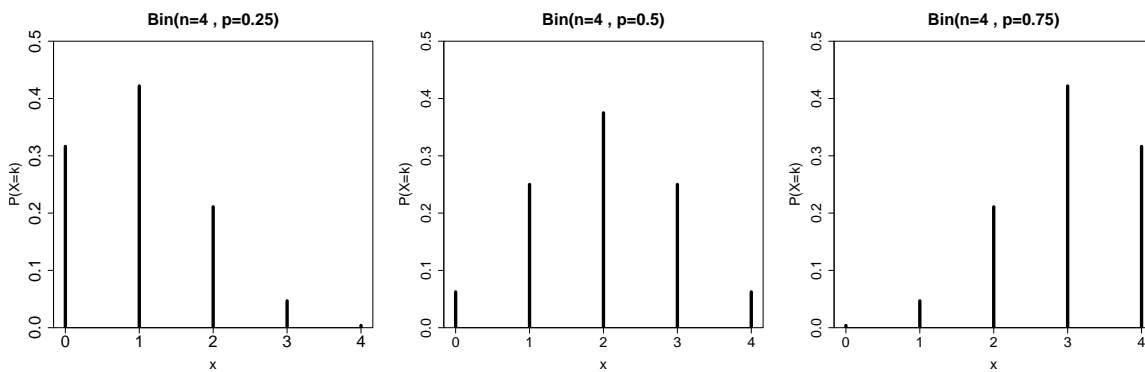


Figura 3.1: Gráficos da função de probabilidade de uma v.a. $B(4, p)$, para alguns valores de p .

Observação: Pela definição anterior, temos que $X = I_1 + I_2 + \dots + I_n$, onde I_i , $i = 1, \dots, n$ são v.a.'s independentes com distribuição $Ber(p)$.

Proposição 3.5. *Seja X uma variável aleatória com distribuição $B(n, p)$. Então v.a. $Y = n - X$ tem distribuição $B(n, 1 - p)$.*

Proposição 3.6 (Valor médio e Variância). *Considere a v.a. $X \sim B(n, p)$. Então,*

$$E(X) = np \quad e \quad V(X) = np(1 - p).$$

Demonstração. A demonstração torna-se mais simples se usarmos a representação $X = I_1 + I_2 + \dots + I_n$, introduzida na última observação. Assim,

$$E(X) = E(I_1 + I_2 + \dots + I_n) = E(I_1) + E(I_2) + \dots + E(I_n) = p + p + \dots + p = np.$$

Atendendo à independência das variáveis I_i ,

$$V(X) = V(I_1 + I_2 + \dots + I_n) = V(I_1) + V(I_2) + \dots + V(I_n) = np(1 - p).$$

□

Exemplo 3.7 (Exame de P.E. D - 2007/08). *Num concurso de televisão o apresentador propõe ao concorrente o seguinte jogo: atiram-se ao ar 3 moedas, em simultâneo, e se todos os lançamentos resultarem em caras o apresentador dá 10 € ao concorrente; Se todos os lançamentos resultarem em coroas o apresentador dá igualmente ao concorrente 10 €. Mas se os lançamentos resultarem em 2 caras e 1 coroa ou em 2 coroas e 1 cara, o concorrente tem de dar ao apresentador 5 €.*

- Represente X a quantidade de dinheiro ganha pelo concorrente. Determine a sua função de probabilidade.
- Baseado no valor esperado de X , diga se o concorrente deve aceitar jogar este jogo.

Resolução:

(a) Considere a v.a. Y : "número de caras obtidas em 3 lançamentos de uma moeda (equilibrada)". Então como em cada lançamento o resultado é cara (sucesso) ou coroa (insucesso) e os resultados dos lançamentos são mutuamente independentes, $Y \sim B(3, 1/2)$.


Como $P(X = -5) = P(Y = 1) + P(Y = 2) = 3/4$ e $P(X = 10) = P(Y = 0) + P(Y = 3) = 1/4$, resulta a seguinte função de probabilidade:

$$X \begin{cases} -5 & 10 \\ 3/4 & 1/4 \end{cases}$$


(b) Como $E(X) = -5/4 < 0$, o concorrente não deve jogar.

Proposição 3.8 (Aditividade). Sejam X_i , $i = 1, \dots, m$, m v.a.'s independentes tais que $X_i \sim B(n_i, p)$. Então a sua soma tem também distribuição Binomial, isto é,

$$S_m = \sum_{i=1}^m X_i \sim B(n_1 + \dots + n_m, p).$$

Sendo uma das distribuições discretas conhecidas, no  existem alguns comandos pré-definidos para esta distribuição:

- `dbinom(x, n, p)` - função densidade de probabilidade em x , isto é $P(X = x)$;
- `pbinom(x, n, p)` - função distribuição em x , isto é $F(x) = P(X \leq x)$;
- `qbinom(q, n, p)` - quantil de probabilidade p ;
- `rbinom(nn, n, p)` - geração de nn números aleatórios com a distribuição considerada.

Se considerarmos uma v.a. $X \sim B(5, 0.6)$, podemos obter no , por exemplo:

```
> dbinom(2,5,0.6)
[1] 0.2304
> dbinom(0,5,0.6)+dbinom(1,5,0.6)
[1] 0.08704
> pbinom(1,5,0.6)
[1] 0.08704
> rbinom(10,5,0.6)
[1] 2 2 1 2 4 3 4 3 1 3
```

$P(X = 2) = \binom{5}{2} 0.6^2 0.4^3 = 0.2304$ e $P(X \leq 1) = \binom{5}{0} 0.6^0 0.4^5 + \binom{5}{1} 0.6^1 0.4^4 = 0.08704$.

Nota: O último comando, apresentado na figura, permite gerar uma sequência pseudo-aleatórias de 10 observações para a distribuição $B(5, 0.6)$.

3.1.3 Distribuição Hipergeométrica

Definição 3.9 (Distribuição Hipergeométrica). Considere-se uma população de N elementos, dos quais M possuem determinada característica e os restantes $(N - M)$ não a possuem (dicotomia). Considere-se a experiência aleatória que consiste em seleccionar ao acaso e sem reposição n elementos (amostra). Associada a esta experiência aleatória, defina-se a v.a. X - nº de elementos com a característica, entre os seleccionados sem reposição. Esta v.a. X tem uma função de probabilidade,

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \quad \max(0, M + n - N) \leq x \leq \min(M, n),$$

e diz-se ter distribuição Hipergeométrica de parâmetros (N, M, n) (pode ser escrito abreviadamente $X \sim H(N, M, n)$).

Proposição 3.10 (Valor médio e Variância). Seja a v.a. $X \sim H(N, M, n)$. Então:

$$E(X) = n \frac{M}{N} \quad \text{e} \quad V(X) = n \frac{M}{N^2(N-1)} (N - M)(N - n).$$


Exemplo 3.11. Num aquário existem 9 peixes, dos quais 5 estão saudáveis (S) e os restantes 4 estão doentes (D). Considere-se a experiência aleatória: extracção ao acaso e sem reposição de 3 peixes e registo do seu estado de saúde. Associada a esta experiência, considere-se a v.a. X - número de peixes saudáveis na amostra extraída de 3 peixes. Quantos peixes saudáveis esperamos encontrar em cada extracção?

Resposta: Como $X \sim H(9, 5, 3)$, o número de peixes saudáveis, que esperamos encontrar em cada extracção de 3 peixes, é $E(X) = 5/3$.

Nota: Em situações em que se conhece totalmente a composição da população e há apenas dois resultados possíveis, a distribuição Binomial caracteriza extracções com reposição. Se não houver reposição, a distribuição adequada é a Hipergeométrica.

Também existem comandos pré-definidos no :

- `dhyper(x, M, N - M, n)` - função densidade de probabilidade em x , isto é $P(X = x)$;
- `phyper(x, M, N - M, n)` - função distribuição em x , isto é $F(x) = P(X \leq x)$;
- `qhyper(p, M, N - M, n)` - quantil de probabilidade p ;
- `rhyper(nn, M, N - M, n)` - geração de nn números aleatórios com a distribuição considerada.

Nota: Apesar da notação adoptada ser $H(N, M, n)$ repare-se que no , os parâmetros a serem introduzidos são M , $N - M$ e n .

Exemplo 3.12. No exemplo dos peixes em que $X \sim H(N = 9, M = 5, n = 3)$, se o pretendido for $P(X = 2)$, $P(X \leq 1)$ ou gerar 10 valores pseudo-aleatórios com a distribuição $H(N = 9, M = 5, n = 3)$, no **R** (com $M = 5, N - M = 4$ e $n = 3$):

```
> dhyper(2,5,4,3)
[1] 0.4761905
> dhyper(0,5,4,3)+dhyper(1,5,4,3)
[1] 0.4047619
> phyper(1,5,4,3)
[1] 0.4047619
> rhyper(10,5,4,3)
[1] 2 1 3 2 1 2 1 3 1 1
```

E, apesar de não estarem definidas as funções $E()$, $var()$, $sd()$, $skewness()$ ou $kurtosis()$, podemos usar as funcionalidades do suplemento **distrEx** para esses cálculos:

```
> Xmnk<-function(m,n,k){DiscreteDistribution(supp=0:k,prob=dhyper(0:k,m,n,k))}
> m<-5; n<-4; k<-3
> E(Xmnk(m,n,k))
[1] 1.666667
> var(Xmnk(m,n,k))
[1] 0.5555556
> skewness(Xmnk(m,n,k))
[1] -0.06388766
> kurtosis(Xmnk(m,n,k))
[1] -0.3428571
```

3.1.4 Distribuição de Poisson

Definição 3.13 (Processo de Poisson). Suponha que estamos interessados em estudar a variável aleatória X que conta o número de ocorrências de um acontecimento num dado intervalo de tempo¹ de duração t (por exemplo, o número de acidentes rodoviários ocorridos num dia ou o número de clientes que entram numa loja durante 1 hora). Temos um processo de Poisson de parâmetro $\lambda > 0$, quando se verificam as seguintes condições:

1. A probabilidade p de ocorrer exactamente um acontecimento num intervalo de amplitude arbitrariamente pequena d é proporcional à sua duração, isto é, $p = \lambda d$;
2. A probabilidade de ocorrer mais do que um acontecimento num intervalo de amplitude arbitrariamente pequena é aproximadamente igual a zero;
3. O número de acontecimentos que ocorrem em dois intervalos disjuntos são independentes.
4. O número de ocorrências em dois intervalos com a mesma duração, têm a mesma distribuição.

¹Note que podemos também considerar uma área, um volume, etc.

Para deduzir a função de probabilidade, vamos considerar um intervalo unitário ($t = 1$), dividido em n sub-intervalos, todos com amplitude $d = 1/n$, com n suficientemente grande. Nas condições acima indicadas, o número de ocorrências em cada sub-intervalo é bem aproximado por uma v.a. $Ber(p)$, com $p = \lambda/n$. Então X tem aproximadamente distribuição $B(n, \lambda/n)$, isto é,

$$P(X = x) \approx \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, \quad x = 0, 1, \dots, n.$$

Se $n \rightarrow \infty$,

$$P(X = x) = \lim_{n \rightarrow \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots, n.$$

Definição 3.14 (Distribuição de Poisson). Dizemos que a variável aleatória X segue uma distribuição de Poisson de parâmetro λ , e escrevemos $X \sim P(\lambda)$, se a função de probabilidade de X é:

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

Observação: Se num processo de Poisson, os acontecimentos ocorrem a uma taxa média λ , por unidade de tempo, então o número de ocorrências num intervalo de amplitude $t > 0$ tem distribuição de Poisson de parâmetro λt .

Por exemplo, se durante a hora de almoço (das 12 às 14 horas) a chegada de automóveis a um parque se processa a uma taxa de 180 automóveis por hora e tem distribuição de Poisson, então a distribuição do número de automóveis que chegam em 15 minutos é Poisson com parâmetro $\lambda t = 180 \times \frac{1}{4} = 45$. A distribuição do número de automóveis que chegam durante a hora do almoço é Poisson de parâmetro $\lambda t = 180 \times 2 = 360$.

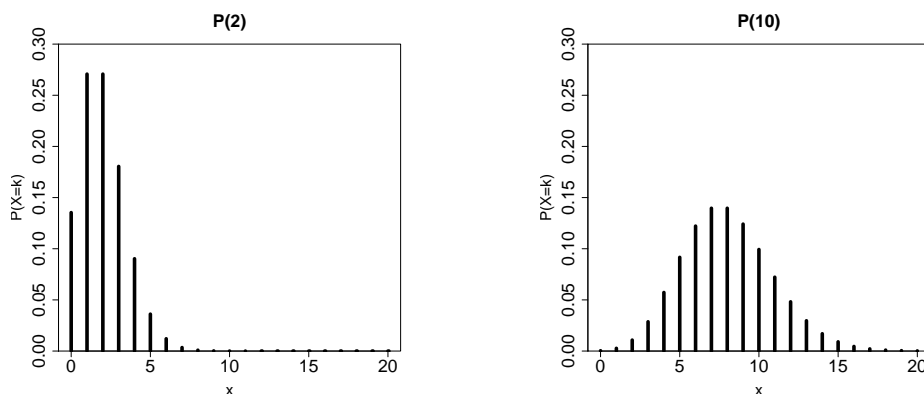


Figura 3.2: Função de probabilidade de uma v.a. $P(\lambda)$, para alguns valores de λ .

Proposição 3.15 (Valor médio e Variância). *Seja X uma v.a. com distribuição $P(\lambda)$. Então,*

$$E(X) = \lambda \quad e \quad V(X) = \lambda.$$

Teorema 3.16 (Aditividade). *Sejam X_1, X_2, \dots, X_m variáveis aleatórias independentes com $X_i \sim P(\lambda_i)$, $i = 1, \dots, m$. Então,*

$$S_m = \sum_{i=1}^m X_i \sim P(\lambda_1 + \dots + \lambda_m).$$

Em \mathbb{R} temos comandos correspondentes à distribuição de Poisson ($P(\lambda)$): `dpois(x, λ)`, `ppois(x, λ)`, `qpois(q, λ)` e `rpois(nn, λ)`.

3.2 Distribuições Contínuas

3.2.1 Distribuição Exponencial

Começamos por introduzir a função Gama, presente em muitos livros de Análise Matemática.

A função Gama corresponde ao integral:

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx, \quad a > 0 \quad (3.1)$$

Propriedades da função Gama:

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$;
2. $\Gamma(n) = (n - 1)!$, $n \in \mathbb{N}$
3. $\Gamma(1/2) = \sqrt{\pi}$
4. $\int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$.

Definição 3.17 (Distribuição Exponencial). *Uma variável aleatória X diz-se seguir uma distribuição Exponencial de parâmetro λ , e escrevemos $X \sim \text{Exp}(\lambda)$, se a sua função densidade de probabilidade for dada por:*

$$f(x) = \begin{cases} 0, & x \leq 0; \\ \lambda e^{-\lambda x}, & x > 0; \end{cases} \quad \lambda > 0.$$

A sua função de distribuição é dada por:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

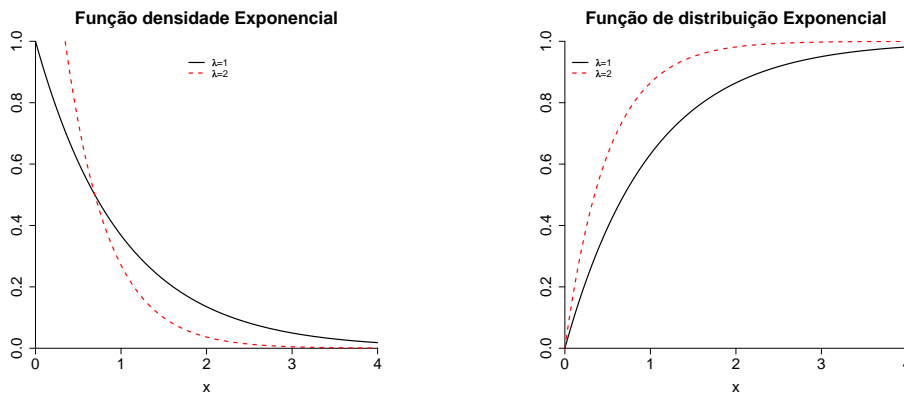


Figura 3.3: Função densidade (esquerda) e função de distribuição (direita) de uma v.a. $Exp(\lambda)$.

Proposição 3.18 (Valor médio e Variância). *Considere a v.a. $X \sim Exp(\lambda)$. Então,*

$$E(X) = \frac{1}{\lambda} \quad e \quad V(X) = \frac{1}{\lambda^2}.$$

Demonstração. Vamos utilizar as propriedades da função Gama para calcular o valor médio. Assim,

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} x^{2-1} e^{-\lambda x} dx = \lambda \frac{\Gamma(2)}{\lambda^2} = \frac{1}{\lambda}.$$

De modo análogo se calcula $E(X^2)$ e se verifica que $V(X) = \frac{1}{\lambda^2}$. □

Proposição 3.19 (Relação entre a distribuição Exponencial e Poisson). *Considere um acontecimento que ocorre de acordo com um Processo de Poisson de parâmetro λ , por unidade de tempo. Então, o tempo até à primeira ocorrência e o tempo entre duas ocorrências consecutivas tem distribuição $Exp(\lambda)$.*

Exemplo 3.20. *Admita que o número de avarias de uma fotocopiadora é um processo de Poisson com taxa $\lambda = 5/\text{ano}$. Calcule a probabilidade do tempo entre avarias consecutivas ser inferior a um mês.*

Resolução: O tempo X entre avarias consecutivas tem distribuição $Exp(5)$. Assim, a probabilidade pedida é:

$$P(X < 1/12) = F_X(1/12) = 1 - e^{-\lambda/12} = 1 - e^{-5/12} = 0.3408.$$

Teorema 3.21 (Falta de memória da distribuição exponencial). *Seja $X \sim Exp(\lambda)$. Então:*

$$P(X \geq x + y | X \geq y) = P(X \geq x).$$

Em **R** temos os seguintes comandos para a distribuição Exponencial ($Exp(\lambda)$): `dexp(x, λ)`, `pexp(x, λ)`, `qexp(q, λ)` e `rexp(nn, λ)`.

3.2.2 Distribuição Normal

Definição 3.22 (Distribuição Normal). Uma variável aleatória X diz-se seguir uma distribuição Normal de parâmetros μ e σ^2 , e escrevemos $X \sim N(\mu, \sigma^2)$, se a sua função densidade de probabilidade for dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$$

A função de distribuição é dada pelo integral:

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

para o qual não existe solução analítica. É assim necessário recorrer a métodos numéricos para obter os valores desta função.

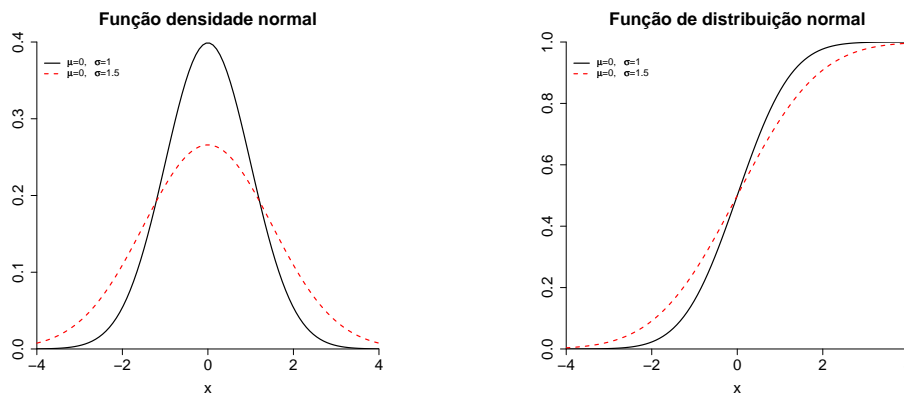


Figura 3.4: Função densidade (esquerda) e função de distribuição (direita) de uma v.a. $N(\mu, \sigma)$.

Observações:

- Esta distribuição é também conhecida pelo nome de **Gaussiana** ou **distribuição de Gauss**.
- Quando $\mu = 0$ e $\sigma = 1$, a v.a. toma o nome de **Normal reduzida**. Neste caso é costume representar por ϕ e Φ , respectivamente, a função densidade e função de distribuição.
- A distribuição Normal é simétrica em torno de μ .

Proposição 3.23 (Valor médio e Variância). Seja a v.a. $X \sim N(\mu, \sigma^2)$. Então

$$E(X) = \mu \quad e \quad V(X) = \sigma^2.$$

Teorema 3.24. Seja $X \sim N(\mu, \sigma^2)$. Resulta que,

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Teorema 3.25. Se $X \sim N(\mu, \sigma^2)$ e a, b são constantes reais, com $a \neq 0$, então


$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Teorema 3.26. Sejam X_1, X_2, \dots, X_n , n variáveis aleatórias independentes com distribuições $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$. Considerando as constantes reais a_1, a_2, \dots, a_n , com algum $a_i \neq 0$, temos que:

$$Y = a_1X_1 + \dots + a_nX_n \sim N\left(\underbrace{a_1\mu_1 + \dots + a_n\mu_n}_{=\mu_Y}, \underbrace{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2}_{=\sigma_Y^2}\right).$$

Note que:

$$\begin{aligned}\mu_Y &= E(Y) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) = \sum_{i=1}^n a_i \mu_i \\ \sigma_Y^2 &= V(Y) = V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i) = \sum_{i=1}^n a_i^2 \sigma_i^2\end{aligned}$$

Em  temos os comandos correspondentes à distribuição Normal ($N(\mu, \sigma)$): `dnorm(x, mu, sigma)`, `pnorm(x, mu, sigma)`, `qnorm(q, mu, sigma)` e `rnorm(nn, mu, sigma)`.

3.2.3 Distribuição do Qui Quadrado

Definição 3.27 (Distribuição do Qui Quadrado). Uma variável aleatória X diz-se seguir uma distribuição Qui-quadrado com n graus de liberdade, e escrevemos $X \sim \chi_n^2$, se a sua função densidade probabilidade for dada por:

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)2^{n/2}} e^{-x/2} x^{n/2-1}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

onde Γ representa a função Gama, introduzida em (3.1).

Proposição 3.28. Considere a v.a. $X \sim \chi_n^2$. Então,

$$E(X) = n, \quad e \quad V(X) = 2n.$$

Teorema 3.29. Sejam X_1, X_2, \dots, X_n v.a.'s independentes com distribuição Normal Reduzida. Então,

$$X_i^2 \sim \chi_1^2,$$

e

$$Y = X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi_n^2.$$

Em  temos os comandos: `dchisq(x, df)`, `pchisq(x, df)`, `qchisq(q, df)` e `rchisq(nn, df)`.

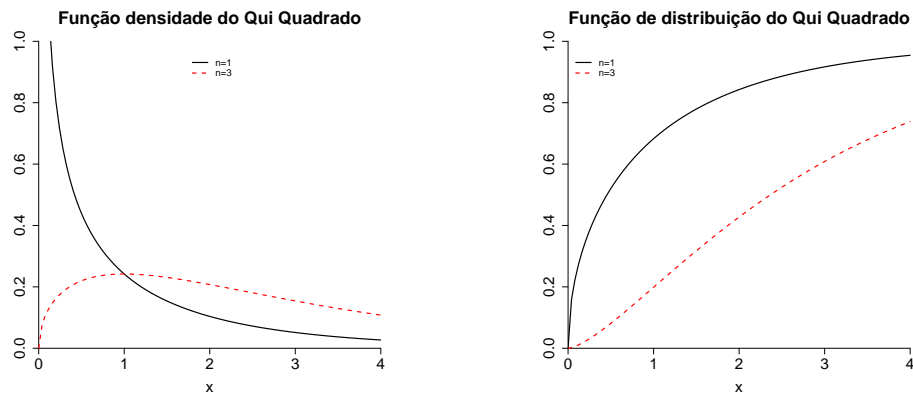


Figura 3.5: Função densidade (esquerda) e função de distribuição (direita) de uma v.a. χ_n^2 .

3.2.4 Distribuição t de Student

Definição 3.30 (Distribuição t de Student). Uma v.a. T diz-se ter distribuição t de Student com n graus de liberdade, e escreve-se $T \sim t_n$, se a sua função densidade probabilidade é dada por:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{(n+1)}{2}}, \quad t \in \mathbb{R}.$$

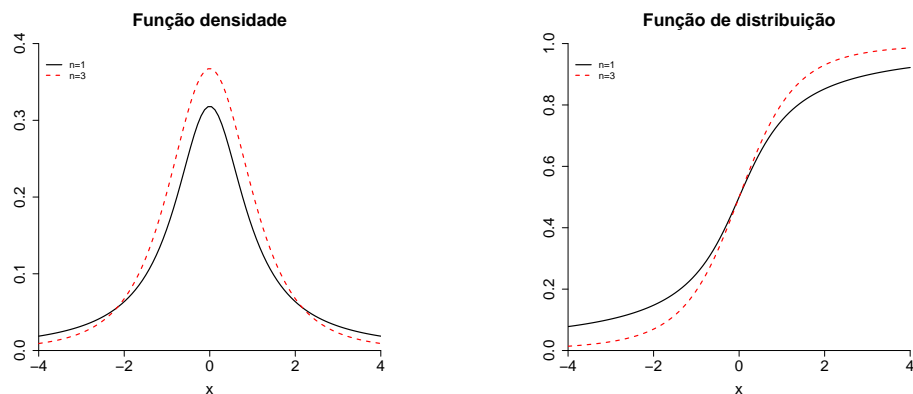


Figura 3.6: Função densidade (esquerda) e função de distribuição (direita) de uma v.a. t_n .


Proposição 3.31 (Valor médio e Variância). Seja $X \sim t_n$. Então,

$$E(X) = 0, \quad n > 1, \quad e \quad V(X) = \frac{n}{n-2}, \quad n > 2.$$

Teorema 3.32. Sejam $X \sim N(0, 1)$ e $Y \sim \chi_n^2$, com X e Y independentes. Então a variável aleatória,

$$T = \frac{X}{\sqrt{Y/n}},$$

tem distribuição *t* de Student com n graus de liberdade.

Em  temos os comandos: `dt(x, df)`, `pt(x, df)`, `qt(q, df)` e `rt(nn, df)`.

Capítulo 4

Teorema Limite Central

Apresentamos neste capítulo, um dos mais importantes resultados da teoria das probabilidades e da estatística, o Teorema Limite Central. Este teorema dá-nos a distribuição aproximada da soma de n variáveis aleatórias independentes e identicamente distribuídas.

Teorema 4.1 (Teorema Limite Central). *Seja X_1, X_2, \dots , uma sucessão de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), com valor médio μ e variância $\sigma^2 \neq 0$, finitos. Considere as variáveis aleatórias S_n e Z_n , definidas por $S_n = \sum_{i=1}^n X_i$ e,*

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}. \quad (4.1)$$

Então a distribuição de Z_n converge para uma distribuição Normal reduzida, quando $n \rightarrow +\infty$, isto é,

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} \underset{a}{\approx} N(0, 1).$$

Observação: Se no quociente da equação (4.1), que define Z_n , dividirmos tanto o numerador como o denominador por n , obtemos

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma},$$

onde \bar{X}_n representa a média S_n/n . O Teorema Limite Central pode assim também ser enunciado em relação à média das variáveis aleatórias X_i , em vez da soma, S_n .

Observação: O Teorema Limite Central não indica nada sobre a velocidade de convergência de Z_n para a distribuição $N(0, 1)$. Essa velocidade de convergência depende da distribuição das v.a.'s X_i . Na prática, este teorema usa-se muitas vezes quando $n \geq 30$ (embora este valor nem sempre garanta uma boa aproximação).

Exemplo 4.2. *Num estudo sobre vendas num hipermercado, concluiu-se que a procura diária de arroz (em Kg) é uma v.a. com valor médio 40Kg e desvio-padrão 5Kg. Tendo sido encomendado 14.500Kg de arroz para venda venda no próximo ano, qual a probabilidade deste stock cobrir a procura de arroz nesse período? (Considere-se um ano com 364 dias).*

Resolução: Seja $X_i =$ procura de arroz no dia i , $i = 1, 2, \dots, 364$ e admitamos que estas v.a.'s são i.i.d.. Sabemos que:

$$E(X_i) = 40Kg, \quad V(X_i) = 25Kg^2, \quad i = 1, 2, \dots, 364.$$

A procura de arroz durante um ano será $S_{364} = \sum_{i=1}^{364} X_i$ e queremos calcular $P(S_{364} \leq 14.500)$. Ignoramos qual a distribuição de S_{364} , mas como se trata de uma soma de um grande número de v.a.'s i.i.d. ($364 > 30$), então pelo T.L.C.,

$$\frac{S_{364} - 364 \times 40}{\sqrt{364 \times 5}} = \frac{S_{364} - 14.560}{\sqrt{364 \times 5}} \underset{a}{\approx} N(0, 1).$$

Assim,

$$\begin{aligned} P(S_{364} \leq 14.500) &= P\left(\frac{S_{364} - 14.560}{\sqrt{364 \times 5}} \leq \frac{14.500 - 14.560}{\sqrt{364 \times 5}}\right) \approx \\ &\approx P(Z \leq -0.63) = \Phi(-0.63) = 1 - \Phi(0.63) = 1 - 0.7357 = 0.2643. \end{aligned}$$

Conclusão: "É recomendável comprar mais arroz!"

Capítulo 5

Estimação Pontual

5.1 Alguns conceitos importantes

Definição 5.1 (População). *Uma população consiste no conjunto de elementos sobre o qual incide o estudo estatístico.*

Definição 5.2 (Característica Estatística ou Atributo). *Característica Estatística ou Atributo é a característica que se observa nos elementos da população.*

Observação: Numa população podemos ter mais que uma **característica estatística** ou **atributo**.

Exemplo 5.3. *Um estudo consiste em estudar a altura dos alunos da FCT num certo ano lectivo. A população em estudo é constituída por todos os alunos da FCT nesse ano lectivo. A característica estatística ou atributo é a altura.*

Definição 5.4 (Amostra). *Uma amostra é um subconjunto da população.*

Observação: Nos métodos estatísticos, que iremos estudar, a amostra recolhida deve ser representativa da população. Caso isso não aconteça, podemos retirar conclusões erradas. É assim conveniente escolher os elementos da amostra de forma aleatória, ou seja, trabalhar com uma amostra aleatória.

Definição 5.5 (Amostra aleatória). *Vamos admitir que cada valor observado x_i é a realização da variável aleatória X_i , com função de distribuição F . O vector (X_1, X_2, \dots, X_n) constitui uma amostra aleatória se e só se as n variáveis aleatórias são independentes e têm todas a mesma distribuição. Os valores que se obtêm por concretização da amostra aleatória são representados por (x_1, x_2, \dots, x_n) .*

Definição 5.6 (Estatística). *Uma estatística é uma qualquer função da amostra aleatória, (X_1, X_2, \dots, X_n) , que não depende de qualquer parâmetro desconhecido.*

Observação: Da definição anterior, conclui-se que uma estatística é uma variável aleatória. Logo qualquer estatística tem função de distribuição. A essa função de distribuição dá-se o nome de distribuição por amostragem da estatística.

Exemplo 5.7 (Estatística). Dada uma amostra aleatória (X_1, X_2, \dots, X_n) , de dimensão n , são estatísticas: A média amostral (\bar{X}) , a variância amostral (S^2) , o mínimo da amostra, o máximo da amostra, a mediana, os quartis ou a própria amostra.

Definição 5.8 (Estimador pontual e estimativa pontual). Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão n dum população com função de distribuição $F(x|\theta)$, com parâmetro desconhecido θ . Um **estimador pontual** de θ é uma estatística $\hat{\Theta} = h(X_1, X_2, \dots, X_n)$ destinada a estimar o parâmetro θ . Depois da amostra ter sido recolhida, o valor particular de $\hat{\theta} = h(x_1, x_2, \dots, x_n)$, é designado por **estimativa pontual** de θ .

Tabela 5.1: Alguns dos parâmetros populacionais que interessam estimar e respectivos estimadores pontuais.

| Parâmetro Populacional | Estimador Pontual |
|--|---|
| Média populacional μ | Média amostral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ |
| Variância populacional σ^2 | Variância amostral $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ |
| Desvio padrão populacional σ | Desvio padrão amostral $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ |
| Proporção populacional p | Proporção amostral $\hat{P} = \frac{X}{n}$ |

5.2 Propriedades dos estimadores

Um dos principais objectivos da Estatística é a estimação de parâmetros desconhecidos, como por exemplo a média da população, a partir de uma amostra. Como muitas vezes temos vários estimadores para o mesmo parâmetro, qual devemos utilizar? É aconselhável a escolha do estimador que melhor satisfaça um critério de eficiência. Para definir o critério de eficiência que iremos utilizar, precisamos das seguintes definições:

Definição 5.9 (Estimador centrado e assintoticamente centrado). Um estimador pontual, $\hat{\Theta}$, diz-se centrado para o parâmetro θ se e só se

$$E(\hat{\Theta}) = \theta.$$

Caso $E(\hat{\Theta}) \neq \theta$, o estimador diz-se enviesado. A diferença $b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$ corresponde ao valor do enviesamento ou viés de $\hat{\Theta}$. Se $E(\hat{\Theta}) \neq \theta$, e $\lim_{n \rightarrow \infty} E(\hat{\Theta}) = \theta$, diz-se que o estimador é assintoticamente centrado.

Definição 5.10 (Erro Padrão de um estimador). Dado um estimador pontual $\hat{\Theta}$, centrado, define-se o seu erro padrão, $SE_{\hat{\Theta}}$, por

$$SE_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}.$$

Caso o erro padrão envolva parâmetros desconhecidos, que possam ser estimados a partir dos valores da amostra, a substituição destes valores estimados no erro padrão produz o chamado **erro padrão estimado**, denotado por $\widehat{SE}_{\hat{\Theta}}$.

Exemplo 5.11 (Cálculo do erro padrão do estimador da média da população - μ). Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma população com valor médio μ e variância σ^2 . Como,

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

concluimos que \bar{X} é estimador centrado do valor médio da população, μ . Temos ainda,

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \quad (X_i \text{ v.a.'s independentes}) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}, \end{aligned}$$

ou seja, $SE_{\bar{X}} = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$.

Definição 5.12 (Erro Quadrático Médio). O erro quadrático médio de um estimador pontual $\hat{\Theta}$ de um parâmetro θ é definido por

$$EQM(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2] = V(\hat{\Theta}) + b^2(\hat{\Theta}),$$

onde $b(\hat{\Theta})$ é o viés de $\hat{\Theta}$, definido em 5.9.

Definição 5.13 (Eficiência). Sejam $\hat{\Theta}_1$ e $\hat{\Theta}_2$ dois estimador pontuais de uma parâmetro θ . Diz-se que $\hat{\Theta}_1$ é mais eficiente que $\hat{\Theta}_2$, se e só se, $EQM(\hat{\Theta}_1) \leq EQM(\hat{\Theta}_2)$.

Definição 5.14 (Estimador consistente). Um estimador pontual $\hat{\Theta}$ de um parâmetro θ , diz-se consistente se

$$\lim_{n \rightarrow \infty} EQM(\hat{\Theta}) = 0.$$

Observação: Se o estimador for centrado, então $EQM(\hat{\Theta}) = V(\hat{\Theta})$, logo será consistente se $\lim_{n \rightarrow \infty} V(\hat{\Theta}) = 0$.

Exemplo 5.15 (Consistência da Média amostral). Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma população com valor médio μ e variância σ^2 . Sabemos que \bar{X} é estimador centrado do valor médio da população, μ e $V(\bar{X}) = \frac{\sigma^2}{n}$. Como ,

$$\lim_{n \rightarrow \infty} EQM(\hat{\Theta}) = \lim_{n \rightarrow \infty} V(\hat{\Theta}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0,$$

concluimos que \bar{X} é consistente para μ .

5.3 Distribuições por Amostragem

Nesta secção vamos estudar a distribuição por amostragem dos estimadores pontuais da Tabela 5.1.

5.3.1 Distribuição por amostragem da média amostral, \bar{X}

Suponhamos que foi seleccionada uma amostra aleatória de dimensão n , (X_1, X_2, \dots, X_n) , de uma população de média μ e variância σ^2 . A distribuição por amostragem de \bar{X} pode ser obtida sob diversas condições:

1. Suponhamos a **população tem distribuição Normal** e que o **valor da variância da população é conhecido**. Consequentemente, tendo em conta as propriedades da distribuição normal, $\bar{X} \sim N(\mu, \sigma^2/n)$, ou seja,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

2. Suponhamos a **população tem distribuição Normal** e que o **valor da variância da população é desconhecido**. Vamos aqui usar S^2 para estimar σ^2 . Nestas condições,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{e} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Como a população tem distribuição Normal, podemos assegurar que Z e S^2 são v.a. independentes (demonstração fora do âmbito desta disciplina). Logo, usando o Teorema 3.32,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{(n-1)}}} \sim t_{n-1}.$$

3. Suponhamos que **não se conhece a distribuição da população** e que o valor da **variância da população é conhecida**, mas a dimensão da amostra, n , é superior ou igual a 30. Neste caso, a distribuição por amostragem da média amostral pode ser aproximada pela distribuição Normal reduzida, justificado através do Teorema Limite Central:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0, 1).$$

4. Finalmente, consideremos que seleccionámos uma amostra aleatória de uma **população com distribuição desconhecida**, com **variância da população desconhecida** e que temos um tamanho de amostra n superior ou igual a 30. Tal como no caso anterior,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0, 1).$$

Como σ^2 não é conhecido, mas a dimensão da amostra é grande então $S \simeq \sigma$, e podemos substituir, na expressão anterior, σ por S (desvio padrão), isto é,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1).$$

5.3.2 Distribuição por amostragem da variância amostral, S^2

Suponhamos que foi seleccionada uma amostra aleatória de dimensão n , (X_1, X_2, \dots, X_n) , de uma **população Normal** de **média μ , desconhecida**, e **variância σ^2** . Neste contexto, a distribuição por amostragem de $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ é dada por:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

5.3.3 Distribuição por amostragem da proporção, \hat{P}

Admita que os elementos de determinada população possuem uma dada característica, com uma certa probabilidade p desconhecida, independentemente uns dos outros. Suponhamos que se selecciona uma amostra aleatória de n elementos desta população. Se X denotar o número de elementos da amostra aleatória que possuem a referida característica, sabemos que $X \sim B(n, p)$. Se o tamanho da amostra for suficientemente grande, o Teorema Limite Central justifica que:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \stackrel{a}{\sim} N(0, 1).$$


Como p pode ser pontualmente estimado pela proporção de elementos que na amostra possuem a referida característica, $\hat{P} = \frac{X}{n}$, a **distribuição por amostragem aproximada de \hat{P}** é

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim} N(0, 1).$$

Tabela 5.2: Distribuições por amostragem

| Estimador | População | | Distribuição |
|-----------|--|-------------------------|--|
| \bar{X} | Normal de média μ | σ^2 conhecida | $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ |
| | | σ^2 desconhecida | $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ |
| | Pop. desconhecida de média μ e $n \geq 30$ | σ^2 conhecida | $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$ |
| | | σ^2 desconhecida | $Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$ |
| \hat{P} | Qualquer população e n grande | | $Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim} N(0, 1)$ |
| S^2 | Normal de média μ desconhecida | | $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ |

Estatística Descritiva em

Em  podemos utilizar as seguintes instruções para o estudo descritivo de um conjunto de *dados*.

- `getwd()` e/ou `setwd("C:\\Users\\...")` - definir a directoria de trabalho;
- `read.table(ficheiro, header=TRUE)` - ler um ficheiro de dados com cabeçalho;
- `summary(dados)` - indicadores estatísticos básicos (min, max, mediana, média e 1º e 3º quartis);

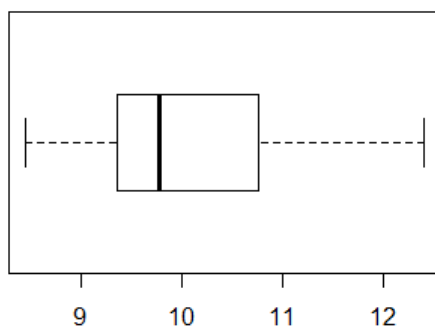
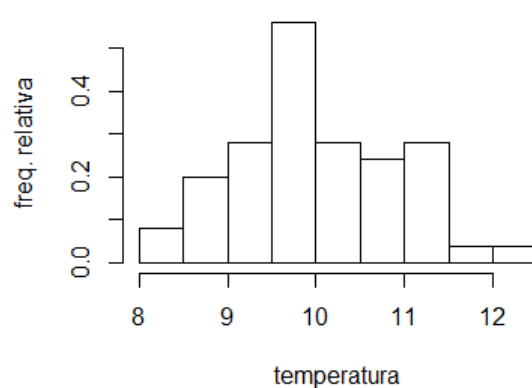
- `var(dados)` `sd(dados)` - variância e desvio padrão;
- `boxplot(dados)` - diagrama de caixa e bigodes;
- `hist(dados)` - histograma.

Exemplo 5.16. No ficheiro `dados1.txt`, podemos encontrar a informação referente a 50 valores de temperaturas mínimas.

```
> setwd("C:/Users/.../IPEIO/Rfiles")
> dados<-read.table("dados1.txt",header=TRUE)
> summary(dados)

temperatura
  Min.   : 8.440
  1st Qu.: 9.377
  Median : 9.775
  Mean   :10.024
  3rd Qu.:10.752
  Max.   :12.400

> var(dados[["temperatura"]])
[1] 0.8389011
> sd(dados[["temperatura"]])
[1] 0.9159154
> boxplot(dados,horizontal=TRUE,main="BoxPlot")
> hist(dados[["temperatura"],prob=TRUE,main="Histograma",
xlab="temperatura",ylab="freq. relativa")
```

BoxPlot**Histograma**

Exemplo 5.17. No ficheiro `dados2.txt`, podemos encontrar a informação referente a 50 valores de tempos de espera.


```

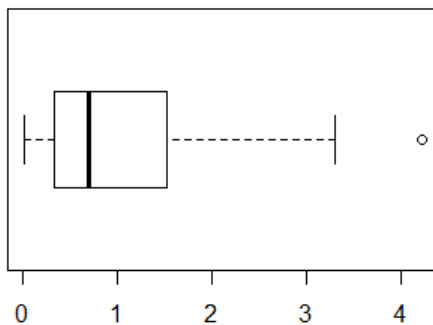
> setwd("C:/Users/.../IPEIO/Rfiles")
> dados<-read.table("dados2.txt",header=TRUE)
> summary(dados)

tempo_de_espera
  Min.   : 0.0100
  1st Qu.: 0.3375
  Median : 0.6950
  Mean   : 1.0238
  3rd Qu.: 1.5150
  Max.   : 4.2200

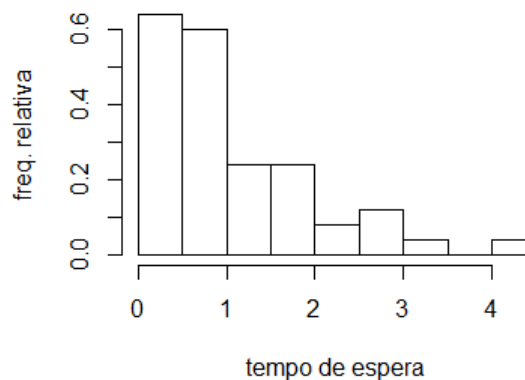
> x<-dados[["tempo_de_espera"]]
> var(x)
[1] 0.8836404
> sd(x)
[1] 0.9400215
> boxplot(x,horizontal=TRUE,main="BoxPlot")
> hist(x,prob=TRUE,main="Histograma",xlab="tempo de espera",ylab="freq. relativa")

```

BoxPlot



Histograma



Repare-se que as estimativas pontuais do valor médio, variância e desvio padrão podem ser obtidas através dos comandos `mean()`, `var()` e `sd()`. Já uma estimativa pontual para a proporção de tempos de espera inferiores a um certo valor pode ser obtida através dos comandos `sum()` e `length()`. Por exemplo:

```

> xbar<-mean(x); xbar           #média de x
[1] 1.0238
> s2<-var(x); s2               #variância de x
[1] 0.8836404
> s<-sd(x); s                 #desvio padrão de x
[1] 0.9400215
> phat<-sum(x<1)/length(x); phat #proporção de elementos em x inferiores a 1
[1] 0.6

```

Capítulo 6

Estimação por Intervalo de Confiança

A indicação de um único valor como estimativa, de um parâmetro θ , não nos dá informação sobre a precisão desse valor. Por isso, em muitas situações, interessa-nos dar uma medida desse erro. Assim, em vez de se indicar a sua estimativa pontual, é preferível indicar que o parâmetro a estimar estará **provavelmente** no intervalo $]t_1, t_2[$, onde os extremos t_1 e t_2 dependem do valor da estimativa pontual desse parâmetro.

Definição 6.1 (Intervalo Aleatório). *Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma população com função de distribuição F . Considere as estatísticas*

$$T_1(X_1, X_2, \dots, X_n) \quad \text{e} \quad T_2(X_1, X_2, \dots, X_n),$$

tais que $P(T_1 < \theta < T_2) = 1 - \alpha$, onde $\alpha \in]0, 1[$ não depende de θ . Então $]T_1, T_2[$ é um intervalo aleatório para θ .

Definição 6.2 (Intervalo de Confiança). *Seja (x_1, x_2, \dots, x_n) uma realização da amostra aleatória e sejam*

$$t_1 = T_1(x_1, x_2, \dots, x_n) \quad \text{e} \quad t_2 = T_2(x_1, x_2, \dots, x_n),$$

os valores das estatísticas T_1 e T_2 (introduzidas na Definição 6.1). Ao intervalo $]t_1, t_2[$ chamamos intervalo de confiança $(1 - \alpha) \times 100\%$ para θ . O valor $(1 - \alpha)$ representa o nível (ou coeficiente) de confiança do intervalo e α o nível de significância. Normalmente são usados níveis de confiança superiores a 90%.

Observações:

- Diferentes amostras produzirão eventuais valores distintos $\hat{\theta}$ e conseqüentemente diferentes extremos t_1 e t_2 .
- Os valores t_1 e t_2 são denominados limites de confiança inferior e superior, respectivamente.

- A amplitude de um intervalo de confiança, $t_2 - t_1$, é uma importante medida da qualidade da informação fornecida através da amostra.

Definição 6.3 (Variável Pivot ou Fulcral). *Seja (X_1, X_2, \dots, X_n) uma amostra aleatória, retirada de uma população com função de distribuição F de parâmetro θ . A função $T(X_1, X_2, \dots, X_n)$ é uma variável pivot, ou fulcral, se a sua distribuição for independente de θ .*

Observação: As variáveis aleatórias Z , T e X^2 , apresentadas na Tabela 5.2, são variáveis Pivot.

Definição 6.4 (Método de determinação de um Intervalo de Confiança a partir de uma variável Pivot). *Seja (X_1, X_2, \dots, X_n) uma amostra aleatória, retirada de uma população com função de distribuição F , com parâmetro θ , e seja T uma variável Pivot.*

- Dado o nível de confiança $(1 - \alpha)$, é necessário determinar os valores c_1 e c_2 tais que

$$P(c_1 < T < c_2) = 1 - \alpha.$$

- Caso se verifique,

$$c_1 < T < c_2 \quad \Leftrightarrow \quad T_1(X_1, X_2, \dots, X_n) < \theta < T_2(X_1, X_2, \dots, X_n),$$

então também se pode garantir que

$$P(T_1(X_1, X_2, \dots, X_n) < \theta < T_2(X_1, X_2, \dots, X_n)) = 1 - \alpha.$$

Logo, o intervalo aleatório para θ é $]T_1(X_1, X_2, \dots, X_n), T_2(X_1, X_2, \dots, X_n)[=]T_1, T_2[$.

- Observada a amostra (x_1, x_2, \dots, x_n) , o intervalo de confiança para θ é dado por $]t_1, t_2[$, onde $t_1 = T_1(x_1, x_2, \dots, x_n)$ e $t_2 = T_2(x_1, x_2, \dots, x_n)$.

6.1 Intervalo de Confiança para a média da população, μ

6.1.1 População Normal com variância conhecida

Suponhamos que seleccionámos uma amostra aleatória (X_1, X_2, \dots, X_n) de uma população Normal, de variância σ^2 conhecida, com a qual pretendemos construir um intervalo de confiança $(1 - \alpha) \times 100\%$ para μ .

- Escolha da estatística pivot:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1);$$

- Determinação de c_1 e c_2 : Seja z_a um valor tal que $P(Z > z_a) = a$. Escolhemos $c_1 = z_{1-\alpha/2} = -z_{\alpha/2}$ e $c_2 = z_{\alpha/2}$, como indicado na Figura 6.1. Esta escolha não é casual. Quando $c_1 = -c_2$ obtemos o intervalo de menor amplitude. O valor $z_{\alpha/2}$ é obtido através da resolução da equação:

$$\begin{aligned} P(-z_{\alpha/2} < Z < z_{\alpha/2}) &= 1 - \alpha \Leftrightarrow P(Z < z_{\alpha/2}) - P(Z \leq -z_{\alpha/2}) = 1 - \alpha \Leftrightarrow \\ \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) &= 1 - \alpha \Leftrightarrow \Phi(z_{\alpha/2}) = 1 - \alpha/2 \Leftrightarrow z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2) \end{aligned} \quad (6.1)$$

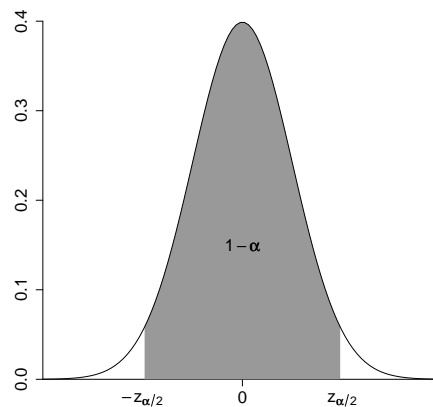


Figura 6.1: Intervalo aleatório da variável pivot Z .

- Determinação dos extremos do intervalo aleatório:

$$\begin{aligned} -z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} &\Leftrightarrow -z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ \Leftrightarrow -z_{\alpha/2} \frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < z_{\alpha/2} \frac{\sigma}{\sqrt{n}} - \bar{X} &\Leftrightarrow \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Logo,

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

- Assim, tendo uma amostra concreta (x_1, x_2, \dots, x_n) , o intervalo de confiança $(1 - \alpha) \times 100$ para μ é:

$$IC_{(1-\alpha) \times 100\%}(\mu) \equiv \left] \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right[.$$

Exemplo 6.5. Considere a população do peso das formigas *Solenopsis*, medido em décimas de grama, que sabemos ter distribuição Normal com média μ e variância $\sigma^2 = 2^2$, $X \sim N(\mu, 2^2)$. Desta população observamos a amostra de 4 pesos, $(8, 13, 9, 8.5)$, a qual usamos para obter uma

estimativa de μ , $\bar{x} = 9.625$ dg. Queremos agora determinar limites inferior e superior de um intervalo de confiança a 95% para μ .

Resolução: Seja \bar{X} a média amostral da amostra de dimensão 4, (X_1, X_2, X_3, X_4) . Como a população tem distribuição Normal, e a variância é conhecida, vamos considerar a estatística pivot $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, cuja distribuição por amostragem foi obtida no capítulo anterior:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{2/\sqrt{4}} \sim N(0, 1).$$

Seja $z_{0.025}$ um valor tal que $P(-z_{0.025} < Z < z_{0.025}) = 0.95$, conforme a Figura 6.2 ilustra. Para

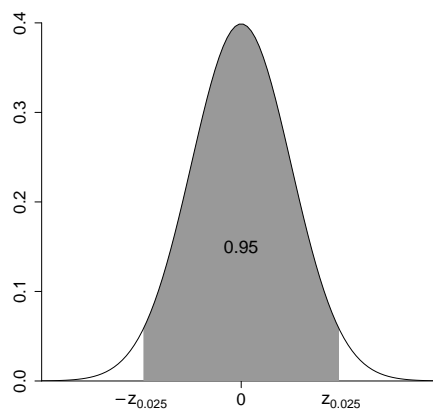


Figura 6.2: Intervalo aleatório da variável pivot Z .

determinar o valor de $z_{0.025}$, é necessário efectuar os seguintes cálculos:

$$\begin{aligned} P(-z_{0.025} < Z < z_{0.025}) = 0.95 &\Leftrightarrow P(Z < z_{0.025}) - P(Z \leq -z_{0.025}) = 0.95 \Leftrightarrow \\ \Phi(z_{0.025}) - \Phi(-z_{0.025}) = 0.95 &\Leftrightarrow \Phi(z_{0.025}) = 0.975 \Leftrightarrow z_{0.025} = \Phi^{-1}(0.975) \approx 1.96 \end{aligned}$$

Assim,

$$\begin{aligned} P(-1.96 < Z < 1.96) = 0.95 &\Leftrightarrow P\left(-1.96 < \frac{\bar{X} - \mu}{2/\sqrt{4}} < 1.96\right) = 0.95 \Leftrightarrow \\ P(-1.96 \times 1 < \bar{X} - \mu < 1.96 \times 1) &= 0.95 \Leftrightarrow \\ P(-\bar{X} - 1.96 < -\mu < -\bar{X} + 1.96) &= 0.95 \Leftrightarrow \\ P(\bar{X} - 1.96 < \mu < \bar{X} + 1.96) &= 0.95 \end{aligned}$$

Logo, o intervalo aleatório, para μ , com 95% de confiança é $]\bar{X} - 1.96; \bar{X} + 1.96[$. Concretizando este intervalo para a amostra observada, $(x_1, x_2, x_3, x_4) = (8, 13, 9, 8.5)$, obtemos o intervalo de confiança a 95% para μ :

$$IC_{95\%}(\mu) =]\bar{x} - 1.96; \bar{x} + 1.96[=]9.625 - 1.96; 9.625 + 1.96[=]7.665; 11.585[.$$

Em **R** também poderemos construir intervalos de confiança e apesar de não haver um comando próprio para esta tarefa, podemos usar alguns dos comando já conhecidos para determinar o intervalo de confiança:

```
> x<-c(8,13,9,8.5)           #amostra
> n<-length(x)              #dimensão da amostra
> xbar<-mean(x)             #média amostral
> sigma<-2                  #desvio padrão populacional (conhecido)
> z0.025<-qnorm(0.975,0,1)  #quantil de prob. 97.5% de N(0,1)
> ICinf<-xbar-z0.025*sigma/sqrt(n); ICinf #extremo inferior do IC
[1] 7.665036
> ICsup<-xbar+z0.025*sigma/sqrt(n); ICsup #extremo superior do IC
[1] 11.58496
```

Observação: Vejamos agora o que sucede aumentando a confiança do intervalo para 99%. Como $Z_{0.005} \approx 2.58$,

$$P(-Z_{0.005} < Z < Z_{0.005}) = 0.99 \Leftrightarrow P\left(-2.58 < \frac{\bar{X}-\mu}{2/\sqrt{4}} < 2.58\right) = 0.99 \Leftrightarrow$$

$$P\left(\bar{X} - 2.58 \times 2/\sqrt{4} < \mu < \bar{X} + 2.58 \times 2/\sqrt{4}\right) = 0.99$$

Assim, $IC_{99\%}(\mu) =]\bar{x} - 2.58; \bar{x} + 2.58[=]9.625 - 2.58; 9.625 + 2.58[=]7.045; 12.205[$.

Concluimos que quando aumentamos o nível de confiança, também aumentamos a sua amplitude!

6.1.2 População Normal com variância desconhecida

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma população Normal(μ, σ^2), de variância σ^2 desconhecida, com a qual pretendemos construir um intervalo de confiança $(1 - \alpha) \times 100\%$ para μ :

- Escolha da estatística pivot:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

- Para um nível de confiança de $(1 - \alpha) \times 100\%$, escolhemos de $c_1 = -t_{n-1, \alpha/2}$ e $c_2 = t_{n-1, \alpha/2}$, como indicado na Figura 6.3.
- Determinação dos extremos do intervalo aleatório:

$$-t_{n-1, \alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1, \alpha/2} \Leftrightarrow$$

$$\Leftrightarrow -t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} < \bar{X} - \mu < t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \Leftrightarrow$$

$$\Leftrightarrow -t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} - \bar{X} < -\mu < t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} - \bar{X} \Leftrightarrow$$

$$\Leftrightarrow \bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}$$

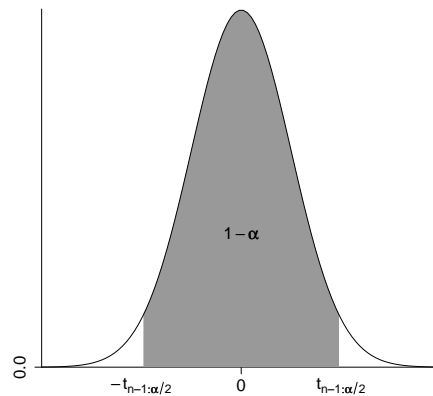



Figura 6.3: Intervalo aleatório da variável pivot T .

- Assim, obtemos o seguinte intervalo de confiança $(1 - \alpha) \times 100\%$ para μ :

$$IC_{(1-\alpha) \times 100\%}(\mu) \equiv \left] \bar{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} ; \bar{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right[.$$

Em , o exemplo 6.5 das formigas com variância desconhecida, obteríamos os extremos do intervalo de confiança:

```

> x<-c(8,13,9,8.5)           #amostra
> n<-length(x)              #dimensão da amostra
> xbar<-mean(x)              #média amostral
> s<-sd(x)                   #desvio padrão amostral
> t0.025<-q(0.975,n-1)      #quantil de prob. 97.5% de t(n-1)
> ICinf<-xbar-t0.025*s/sqrt(n); ICinf #extremo inferior do IC
[1] 5.986291
> ICsup<-xbar+t0.025*s/sqrt(n); ICsup #extremo superior do IC
[1] 13.26371

```

6.1.3 População desconhecida com variância conhecida e $n > 30$

Supondo que seleccionámos uma amostra aleatória de dimensão $n > 30$, (X_1, X_2, \dots, X_n) , de uma população desconhecida com média μ e variância conhecida σ^2 , e com a qual pretendemos construir um intervalo de confiança $(1 - \alpha) \times 100\%$ para μ :

- Escolha da estatística pivot: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$.
- Determinação de c_1 e c_2 : De modo análogo, ao efectuado na página 47, escolhemos $c_1 = -z_{\alpha/2}$ e $c_2 = z_{\alpha/2}$, onde $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$.
- Repetido as contas efectuadas na sub-secção 6.1.1, obtemos:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

- Assim, observada a amostra (x_1, x_2, \dots, x_n) , obtemos o seguinte intervalo de confiança $(1 - \alpha) \times 100\%$ para μ :

$$IC_{(1-\alpha) \times 100\%}(\mu) \equiv \left] \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right[.$$

6.1.4 População desconhecida com variância desconhecida e $n > 30$

Admitindo que seleccionámos uma amostra aleatória de dimensão $n > 30$, (X_1, X_2, \dots, X_n) , de uma população com distribuição desconhecida, com média μ e variância σ^2 , ambos desconhecidos. Pretendemos um intervalo de confiança $(1 - \alpha) \times 100\%$ para μ . Como usamos a estatística pivot:

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1),$$

a determinação do intervalo de confiança para μ é feita de forma análoga ao caso anterior (substituindo σ por S).

Obtemos assim, o seguinte intervalo de confiança (aproximado) $(1 - \alpha) \times 100\%$ para μ :

$$IC_{(1-\alpha) \times 100\%}(\mu) \equiv \left] \bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right[.$$

Exemplo 6.6 (Exame de P.E. D - 2005/06). *Queremos estudar há quanto tempo residem nas suas moradas actuais as pessoas de certa cidade na província. Uma amostra aleatória de 41 famílias revelou uma média de 35 meses de residência e um desvio padrão de 6.3 meses.*

- Qual a sua melhor estimativa do tempo médio de residência da população desta cidade?
- Deduz** um intervalo de confiança a 98% para o verdadeiro tempo médio de residência. Justifique o seu procedimento.

Resolução:

- Para estimar a média da população vamos usar o estimador média da amostra, \bar{X} . Trata-se do estimador da média que possui duas propriedades importantes: é centrado para μ e consistente. Neste exercício, a estimativa do tempo médio de residência da população é $\bar{x} = 35$ meses.
- Para deduzir o intervalo de confiança, vamos admitir que (X_1, X_2, \dots, X_n) é uma amostra aleatória com $n > 30$. Vamos considerar a estatística pivot $Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$, cuja distribuição foi deduzida no capítulo anterior, isto é,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1).$$

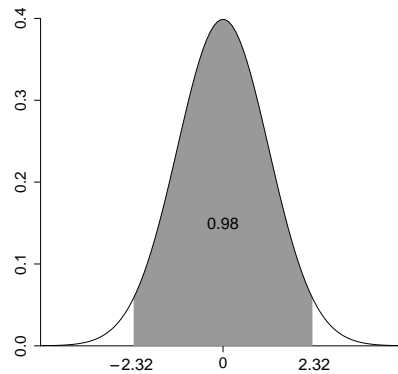


Figura 6.4: Intervalo aleatório da variável pivot Z .

Como $P(-z_{0.01} < Z < z_{0.01}) = 0.98$, onde $z_{0.01} \approx 2.32$, como indicado na Figura 6.4, e

$$\begin{aligned} -z_{0.01} < Z < z_{0.01} &\Leftrightarrow 2.32 < \frac{\bar{X} - \mu}{S/\sqrt{n}} < 2.32 \Leftrightarrow 2.32 \frac{S}{\sqrt{n}} < \bar{X} - \mu < 2.32 \frac{S}{\sqrt{n}} \Leftrightarrow \\ &\Leftrightarrow \bar{X} - 2.32 \frac{S}{\sqrt{n}} < \mu < \bar{X} + 2.32 \frac{S}{\sqrt{n}}, \end{aligned}$$

resulta que

$$P\left(\bar{X} - 2.32 \frac{S}{\sqrt{n}} < \mu < \bar{X} + 2.32 \frac{S}{\sqrt{n}}\right) = 0.98.$$

Logo o intervalo com 98% de confiança para o valor médio da população é

$$IC_{98\%}(\mu) \equiv \left] \bar{x} - 2.32 \frac{s}{\sqrt{n}} ; \bar{x} + 2.32 \frac{s}{\sqrt{n}} \right[.$$

Como da amostra recolhida resultou $\bar{x} = 35$ e $s = 6.3$, o intervalo com 98% de confiança para o valor médio da população é

$$IC_{98\%}(\mu) \equiv \left] 35 - 2.32 \frac{6.3}{\sqrt{41}} ; 35 + 2.32 \frac{6.3}{\sqrt{41}} \right[=]32.72 , 37.28[.$$

6.2 Intervalo de Confiança para a variância populacional, σ^2 , e para o desvio padrão populacional, σ

Nesta secção, vamos deduzir um intervalo de confiança $(1 - \alpha) \times 100\%$ para a variância da população. Consideramos o caso em que temos uma amostra aleatória (X_1, X_2, \dots, X_n) de uma população com distribuição Normal(μ, σ^2), com valor médio μ desconhecido.

- Vamos usar a estatística pivot cuja distribuição por amostragem foi apresentada no capítulo anterior, isto é, a estatística pivot: $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$;

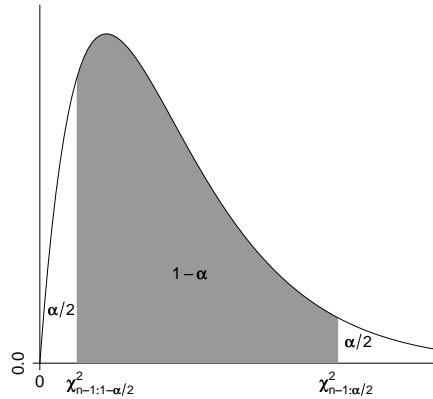


Figura 6.5: Intervalo aleatório da variável pivot X^2 .

- Para um nível de confiança de $(1 - \alpha) \times 100\%$, escolha de c_1 e c_2 : A escolha dos extremos do intervalo aleatório, $c_1 = \chi_{n-1, 1-\alpha/2}^2$ e $c_2 = \chi_{n-1, \alpha/2}^2$, é feita de acordo com a Figura 6.5.

Para determinar estes valores é necessário efectuar as operações:

$$P(X^2 < c_1) = \frac{\alpha}{2} \Leftrightarrow c_1 = F_{\chi_{n-1}^2}^{-1}\left(\frac{\alpha}{2}\right),$$

$$P(X^2 > c_2) = \frac{\alpha}{2} \Leftrightarrow P(X^2 \leq c_2) = 1 - \frac{\alpha}{2} \Leftrightarrow c_2 = F_{\chi_{n-1}^2}^{-1}\left(1 - \frac{\alpha}{2}\right),$$

onde $F_{\chi_{n-1}^2}^{-1}\left(\frac{\alpha}{2}\right)$ e $F_{\chi_{n-1}^2}^{-1}\left(1 - \frac{\alpha}{2}\right)$ podem ser obtidos numa tabela de quantis da distribuição Qui Quadrado.

- Determinação dos extremos do intervalo de confiança: Como,

$$\chi_{n-1, 1-\alpha/2}^2 < X^2 < \chi_{n-1, \alpha/2}^2 \Leftrightarrow \frac{\chi_{n-1, 1-\alpha/2}^2}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi_{n-1, \alpha/2}^2}{(n-1)S^2} \Leftrightarrow$$

$$\frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2},$$

concluimos que

$$P\left(\chi_{n-1, 1-\alpha/2}^2 < X^2 < \chi_{n-1, \alpha/2}^2\right) = P\left(\frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2}\right) = 1 - \alpha.$$

- Assim, observada a amostra (x_1, x_2, \dots, x_n) , e calculada a respectiva variância amostral, s^2 , o intervalo de confiança para σ^2 é:

$$IC_{(1-\alpha) \times 100\%}(\sigma^2) = \left] \frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2} ; \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2} \right[.$$

Observação: Como

$$P\left(\frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2}\right) = P\left(\sqrt{\frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2}}\right) = 1 - \alpha.$$

podemos assim apresentar o seguinte intervalo de confiança para σ :

$$IC_{(1-\alpha)\times 100\%}(\sigma) \equiv \left[\sqrt{\frac{(n-1)s^2}{\chi_{n-1,\alpha/2}^2}} ; \sqrt{\frac{(n-1)s^2}{\chi_{n-1,1-\alpha/2}^2}} \right].$$

Em **R**, no exemplo 6.5 das formigas, obteríamos os extremos dos intervalos de confiança para a variância e para o desvio padrão:

```
> x<-c(8,13,9,8.5) #amostra
> n<-length(x) #dimensão da amostra
> s2<-var(x) #variância amostral
> chisq0.975<-qchisq(0.025,n-1) #quantil de prob. 2.5% de chisq(n-1)
> chisq0.025<-qchisq(0.975,n-1) #quantil de prob. 97.5% de chisq(n-1)
> ICinf<-(n-1)*s2/chisq0.025; ICinf #extremo inferior do IC
[1] 1.678094
> ICsup<-(n-1)*s2/chisq0.975; ICsup #extremo superior do IC
[1] 72.69621
> ICsigma<-c(sqrt(ICinf),sqrt(ICsup)); ICsigma #extremos do IC para o desvio padrão
[1] 1.295413 8.526207
```

Observação: A exagerada amplitude do intervalo de confiança obtido, deve-se em parte à reduzida dimensão da amostra considerada.

6.3 Intervalo de Confiança para proporção populacional, p

Vamos deduzir nesta secção um intervalo de confiança $(1 - \alpha) \times 100\%$ para a proporção populacional p . Consideramos a situação em que estamos interessados em estimar a proporção dos elementos que, na população, possuem determinada característica, através da correspondente proporção amostral \hat{P} , referente a uma amostra de dimensão suficientemente grande. Podemos assim usar a seguinte estatística pivot, cuja distribuição por amostragem foi considerada no capítulo anterior:

- Escolha da estatística pivot:

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim} N(0, 1);$$

- Para um nível de confiança de $(1 - \alpha) \times 100\%$, escolhemos $c_1 = -z_{\alpha/2}$ e $c_2 = z_{\alpha/2}$ tais que $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$. De acordo com os cálculos apresentados na página 47, $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$.

- Determinação dos extremos do intervalo aleatório:

$$-z_{\alpha/2} < Z < z_{\alpha/2} \Leftrightarrow -z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2} \quad (6.2)$$

A resolução das inequações anteriores, em ordem a p , torna-se muito mais simples se substituirmos $\sqrt{p(1-p)/n}$, pela sua estimativa, $\sqrt{\hat{P}(1-\hat{P})/n}$. Se a dimensão da amostra for elevada esta substituição não deverá afectar muito a precisão do intervalo. Assim, efectuando a substituição,

$$\begin{aligned} -z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{\hat{P}(1-\hat{P})/n}} < z_{\alpha/2} &\Leftrightarrow \\ -z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} < \hat{P} - p < z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} &\Leftrightarrow \\ -z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} - \hat{P} < -p < z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} - \hat{P} &\Leftrightarrow \\ \hat{P} - z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} < p < \hat{P} + z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} \end{aligned}$$

Então,


$$\begin{aligned} P \left(-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2} \right) &\simeq \\ P \left(\hat{P} - z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} < p < \hat{P} + z_{\alpha/2} \sqrt{\hat{P}(1-\hat{P})/n} \right) &\simeq 1 - \alpha \end{aligned}$$

- Assim, observada a amostra e calculada a respectiva proporção \hat{p} , obtemos o seguinte intervalo de confiança aproximado:

$$IC_{(1-\alpha) \times 100\%}(p) = \left] \hat{p} - z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} ; \hat{p} + z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right[.$$

Exemplo 6.7. Num inquérito destinado a estimar a proporção p da população que tem TV por cabo, foram inquiridas 200 pessoas, das quais 78 afirmaram ter este serviço. Temos a estimativa pontual da proporção da população com TV por cabo $\hat{p} = 0.39$. Como $n = 200 > 30$ e $z_{0.025} = 1.96$, o intervalo de 95% de confiança para a proporção p é:

$$\left] 0.39 - 1.96 \sqrt{0.39(1-0.39)/200} ; 0.39 + 1.96 \sqrt{0.39(1-0.39)/200} \right[=]0.322, 0.458[.$$

Em  obteríamos os extremos do intervalo:

```
> phat<-78/200 #proporção amostral
> n<-200 #dimensão da amostra
> z0.025<-qnorm(0.975,0,1) #quantil de probabilidade 97.5% de N(0,1)
> ICinf<-phat-z0.025*sqrt(phat*(1-phat)/n);
ICinf #extremo inferior do IC
[1] 0.3224025
> ICsup<-phat+z0.025*sqrt(phat*(1-phat)/n);
ICsup #extremo superior do IC
[1] 0.4575975
```

Observação: Tal como já foi referido, também é possível resolver as inequações em (6.2) em ordem a p sem a substituição de $\sqrt{p(1-p)/n}$, pela sua estimativa, $\sqrt{\hat{P}(1-\hat{P})/n}$. Esta resolução, embora tenha muito mais cálculos, conduz-nos à inequação:

$$\left(1 + \frac{z_{\alpha/2}^2}{n}\right)p^2 - \left(2\hat{P} + \frac{z_{\alpha/2}^2}{n}\right)p + \hat{P}^2 < 0,$$

Os extremos Inferior e superior do intervalo de confiança são, respectivamente:

$$\frac{\hat{P} + z_{\alpha/2}^2/n - z_{\alpha/2}\sqrt{\hat{P}(1-\hat{P})/n + z_{\alpha/2}^2/(4n^2)}}{1 + z_{\alpha/2}^2/n}, \quad \frac{\hat{P} + z_{\alpha/2}^2/n + z_{\alpha/2}\sqrt{\hat{P}(1-\hat{P})/n + z_{\alpha/2}^2/(4n^2)}}{1 + z_{\alpha/2}^2/n}.$$

Como ilustração, apresentamos o intervalo de confiança a 95% para a proporção p , do Exemplo 6.7: $IC_{95\%}(p) =]0.325, 0.459[$.

Capítulo 7

Teste de Hipóteses

7.1 Introdução

Vamos começar por introduzir alguns conceitos importantes e alguma notação.

Definição 7.1 (Hipótese Estatística). *Uma hipótese estatística é uma conjectura acerca da distribuição de uma ou mais variáveis aleatórias. Para cada hipótese que se faça, designada por hipótese nula e denotada por H_0 , há sempre outra hipótese, designada por hipótese alternativa e denotada por H_1 . Se a hipótese estatística H_0 especifica completamente a distribuição é chamada de hipótese simples. Caso contrário é chamada de hipótese composta.*

Uma hipótese estatística pode ser, ou não ser, verdadeira. A verdade ou falsidade nunca pode ser confirmada, a menos que observássemos toda a população, o que nalguns casos é impraticável (quando a população é muito grande) ou até mesmo impossível (no caso de populações infinitas, ou quando característica em estudo leva à destruição dos elementos observados).

Exemplo 7.2 (Hipótese Estatística). *Seja (X_1, X_2, \dots, X_n) uma amostra aleatória da população dos pesos das formigas *Solenopsis* anteriormente considerada. A hipótese estatística de que o peso médio desta população toma o valor δdg denota-se por:*

$$H_0 : \mu = 8 \quad \textit{versus} \quad H_1 : \mu \neq 8 \quad (\textit{Hipótese simples})$$

É usual abreviar a palavra “versus” para “vs”:

$$H_0 : \mu = 8 \quad \textit{vs} \quad H_1 : \mu \neq 8$$

A hipótese estatística de que o peso médio desta população é menor ou igual a δdg denota-se por:

$$H_0 : \mu \leq 8 \quad \textit{vs} \quad H_1 : \mu > 8 \quad (\textit{Hipótese composta}).$$

Ao testarmos uma hipótese nula contra uma hipótese alternativa, a nossa atitude deverá ser admitir H_0 como verdadeira até que os dados fornecidos pela amostra “testemunhem” fortemente contra ela; nesse caso, H_0 deverá ser rejeitada a favor de H_1 .

Definição 7.3 (Teste de Hipóteses). *Um teste de hipóteses é uma regra que nos permite decidir se devemos, ou não, rejeitar H_0 . Esta regra é baseada no valor que a estatística de teste W assume. Assim se,*

- $W(x_1, x_2, \dots, x_n) \in R$, rejeita-se H_0 (e aceita-se H_1 como verdadeira);
- $W(x_1, x_2, \dots, x_n) \notin R$, não se rejeita H_0 .

O conjunto R representa a **região crítica** ou **região de rejeição**.

Definição 7.4 (Erros de tipo I e de tipo II). *Quando realizamos um Teste de Hipóteses podemos cometer um dos seguintes erros:*

- A rejeição de H_0 quando ela é verdadeira (erro de tipo I);
- A não rejeição de H_0 quando esta é falsa (erro de tipo II).

Representamos por α e β , respectivamente, a probabilidade de ocorrer um erro de tipo I ou II, isto é,

- $\alpha = P(\text{erro de tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira});$
- $\beta = P(\text{erro de tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ é falsa}).$

Chamamos ainda **nível de significância** a α e **potência do teste** a $1 - \beta$. Os níveis de significância mais usuais são $\alpha = 0.01$, $\alpha = 0.05$ ou $\alpha = 0.1$.

Observação: O teste ideal é aquele em que estas as probabilidades α e β têm valor mínimo. Contudo, é impossível minimizá-las simultaneamente. De facto, quando α diminui, β aumenta e vice-versa. O procedimento usual consiste em fixar o nível de significância α e escolher a região de rejeição que minimiza β , isto é, que maximize a potência do teste.

Exemplo 7.5. *Seja (X_1, X_2, \dots, X_n) uma amostra aleatória da população dos pesos das formigas *Solenopsis*, isto é, da população $X \sim N(\mu, 2^2)$. Um teste possível para testar:*

$$H_0 : \mu \leq 8 \quad \text{vs} \quad H_1 : \mu > 8,$$

é rejeitar H_0 se $\frac{\bar{X} - 8}{2/\sqrt{n}} > 1.64$.

Definição 7.6 (Valor- p ou “ p -value”). De um modo informal, podemos definir o valor- p ou “ p -value” como o mais pequeno nível de significância que leva à rejeição de H_0 . Assim,

- um valor- p pequeno é desfavorável a H_0 .
- um valor- p elevado indica que as observações são consistentes com H_0 .

Nota: Geralmente o software estatístico apenas apresenta o valor- p do teste. Cabe ao utilizador tomar a decisão ao nível de significância α . Quanto menor for o valor- p , menor é a consistência entre os dados e H_0 . Assim, se valor- $p < \alpha$, devemos rejeitar H_0 , ao nível de significância α .

Regra de cálculo do valor- p :

Seja (x_1, x_2, \dots, x_n) a concretização da amostra aleatória e

$$w_{obs} = W(x_1, x_2, \dots, x_n),$$

o valor observado da estatística de teste W . O valor- p corresponde à probabilidade de se observar um valor igual ou mais “extremo” do que o observado, w_{obs} , se a hipótese nula é verdadeira. O cálculo desta probabilidade depende do tipo de região de rejeição da hipótese H_0 , conforme indicado na seguinte tabela:

| Região de rejeição | valor- p |
|--|---|
| $] -\infty, -c[\cup]c, +\infty[$ ou $]0, b[\cup]c, +\infty[$ | $2 \times \min(P(W < w_{obs} H_0), P(W > w_{obs} H_0))$ |
| $] -\infty, c[$ ou $]0, c[$ | $P(W < w_{obs} H_0)$ |
| $]c, +\infty[$ | $P(W > w_{obs} H_0)$ |

7.2 Teste de Hipóteses para a média da população

De modo análogo, ao efectuado no capítulo anterior, a dedução do teste de hipóteses para o valor médio da população será feito admitindo um dos seguintes pressupostos:

1. População Normal e Variância conhecida;
2. População Normal e Variância desconhecida;
3. População com distribuição desconhecida e Variância conhecida;
4. População com distribuição desconhecida e Variância desconhecida.

7.2.1 Teste bilateral

- Vamos admitir que (X_1, X_2, \dots, X_n) representa uma amostra aleatória de uma população Normal com variância conhecida e que pretendemos testar

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0 \quad (\text{teste bilateral})$$

- Já sabemos que \bar{X} é um estimador centrado de μ . Também já se verificou que $\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$, embora o valor médio, μ , seja desconhecido. Assim, vamos considerar a seguinte estatística de teste:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \underset{\text{sob } H_0}{\sim} N(0, 1).$$

- Considere o nível de significância α . Estamos interessados em rejeitar H_0 quando os valores observados não estiverem de acordo com esta hipótese, isto é, quando a diferença entre \bar{X} e μ_0 for grande. Assim, vamos considerar a região de rejeição $R_\alpha =] -\infty, -z_{\alpha/2}[\cup] z_{\alpha/2}, +\infty[$, indicada na Figura 7.1.

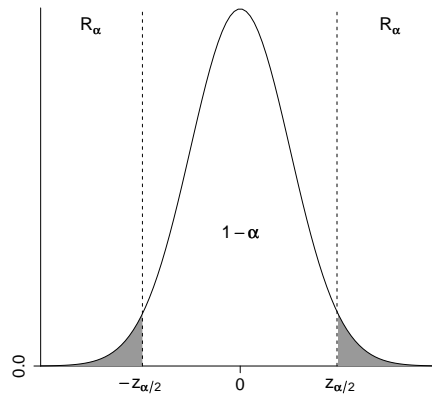


Figura 7.1: Região de rejeição para o teste bilateral para o valor médio.

- A regra de decisão do teste consiste em rejeitar H_0 se $z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \in R_\alpha$, ou seja, se $|z_{obs}| > z_{\alpha/2}$.

Exemplo 7.7. Considere novamente o exemplo da população dos pesos das formigas *Solenopsis*, isto é, a população $X \sim N(\mu, 2^2)$, da qual observámos a amostra aleatória de 4 pesos (8, 13, 9, 8.5). Com base nesta amostra vamos testar, a um nível de significância 5%, a hipótese de que o peso médio populacional μ é igual a 9dg, ou seja vamos testar: $H_0 : \mu = 9$ vs $H_1 : \mu \neq 9$.

- Como a população é normal com variância conhecida, a estatística de teste é:

$$Z = \frac{\bar{X} - 9}{2/\sqrt{4}} \underset{\text{sob } H_0}{\sim} N(0, 1).$$

- Região de rejeição para $\alpha = 0.05$: $R_\alpha =] -\infty, -1.96[\cup] 1.96, +\infty[$.
- Regra de decisão do teste: Rejeitar H_0 ao nível de significância 5% se $z_{obs} = \frac{\bar{x}-9}{2/\sqrt{4}} \in R_{0.05}$.
- Decisão: Como $z_{obs} = \frac{9.625-9}{2/\sqrt{4}} = 0.625 \notin R_{0.05}$, não rejeitamos H_0 ao nível de significância 5%, significando que os dados não vão contra o pressuposto de que o peso médio das formigas é 9dg..

Exemplo 7.8 (Cálculo do valor- p do teste do Exemplo 7.7). Como $Z_{obs} \simeq 0.63$,

$$\begin{aligned} \text{valor-}p &= 2 \min (P(Z < 0.63 \mid H_0), P(Z > 0.63 \mid H_0)) = 2P(Z > 0.63 \mid H_0) = \\ &= 2(1 - P(Z \leq 0.63 \mid H_0)) = 2(1 - \Phi(0.63)) = 0.5286. \end{aligned}$$

Em **R**:

```
x<-c(8,13,9,8.5)           #amostra
> xbar<-mean(x)           #média amostral
> n<-length(x)           #dimensão da amostra
> sigma<-2                #desvio padrão populacional (conhecido)
> mu<-9                   #valor de teste do parâmetro mu
> z.obs<-(xbar-mu)/(sigma/sqrt(n)) #estatística de teste sob H0
> z.obs                    #valor observado da estatística
[1] 0.625
> alpha<-0.05            #nível de significância do teste
> z.alpha<-qnorm(1-alpha/2,0,1) #quantil de probabilidade 1-alpha/2 de N(0,1)
> z.alpha                  #valor crítico
[1] 1.959964
> p.value<-2*min(pnorm(z.obs,0,1),
                  1-pnorm(z.obs,0,1)); p.value #valor-p
[1] 0.5319711
```

Outros testes de hipóteses bilaterais para o valor médio

O teste de hipóteses bilateral, apresentado nesta secção, baseou-se no pressuposto da população ter distribuição Normal e da variância ser conhecida. Noutras condições o teste faz-se de forma análoga, podendo ser necessário alterar a estatística de teste e respectiva região de rejeição, conforme indicado na seguinte tabela:

| População | Variância | Rejeitar H_0 se |
|---|---------------------------|---|
| Pop. Normal de média μ | σ^2 , conhecida | $\left \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \right > z_{\alpha/2}$ |
| | σ^2 , desconhecida | $\left \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \right > t_{n-1,\alpha/2}$ |
| Pop. desconhecida de média μ ($n \geq 30$) | σ^2 , conhecida | $\left \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \right > z_{\alpha/2}$ |
| | σ^2 , desconhecida | $\left \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \right > z_{\alpha/2}$ |

7.2.2 Teste unilateral direito

- Vamos admitir que (X_1, X_2, \dots, X_n) representa uma amostra aleatória de uma população Normal com variância conhecida e pretendemos testar

$$H_0 : \mu \leq \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0 \quad (\text{teste unilateral direito})$$

- De modo análogo, ao apresentado no teste bilateral, vamos considerar a seguinte estatística de teste:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \underset{\text{sob } H_0}{\sim} N(0, 1).$$

- Vamos considerar a região de rejeição $R_\alpha =]z_\alpha, +\infty[$, indicada na Figura 7.2.

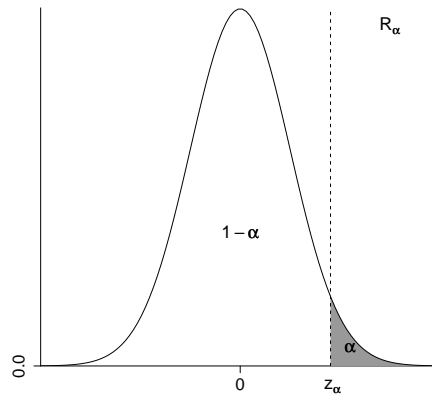


Figura 7.2: Região de rejeição para o teste unilateral direito para o valor médio.

- Regra de decisão: Rejeitar H_0 , ao nível de significância α se $z_{obs} \in R_\alpha$.

Outros testes de hipóteses unilaterais direitos para o valor médio

A estatística de teste e a região de rejeição podem mudar ligeiramente, consoante a população tem, ou não, distribuição Normal e a variância é, ou não é, conhecida. A próxima tabela apresenta, de forma resumida, as alterações que se devem fazer no teste de hipóteses anteriormente deduzido:

| População | Variância | Rejeitar H_0 se |
|---|---------------------------|--|
| Pop. Normal de média μ | σ^2 , conhecida | $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$ |
| | σ^2 , desconhecida | $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{n-1, \alpha}$ |
| Pop. desconhecida de média μ ($n \geq 30$) | σ^2 , conhecida | $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$ |
| | σ^2 , desconhecida | $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > z_\alpha$ |

7.2.3 Teste unilateral esquerdo

O procedimento que deduz o teste unilateral esquerdo, para o valor médio,

$$H_0 : \mu \geq \mu_0 \quad vs \quad H_1 : \mu < \mu_0 \quad (\text{teste unilateral esquerdo}),$$

é análogo ao do teste unilateral direito. Por esta razão apenas se apresentamos a seguinte tabela resumo:

| População | Variância | Rejeitar H_0 se |
|---|---------------------------|--|
| Pop. Normal de média μ | σ^2 , conhecida | $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$ |
| | σ^2 , desconhecida | $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} < -t_{n-1,\alpha}$ |
| Pop. desconhecida de média μ ($n \geq 30$) | σ^2 , conhecida | $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$ |
| | σ^2 , desconhecida | $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} < -z_\alpha$ |

7.3 Teste de Hipóteses para a variância, σ^2 , de uma população Normal

Suponha que observamos uma amostra aleatória (X_1, X_2, \dots, X_n) de uma população $X \sim N(\mu, \sigma^2)$, em que μ é desconhecido. Vamos nesta secção apresentar alguns testes de hipóteses, relativos ao valor da variância da população, σ^2 .

- Testamos uma das três seguintes hipóteses (nula e alternativa):

1. $H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2 \quad (\text{teste bilateral});$
2. $H_0 : \sigma^2 \leq \sigma_0^2 \quad vs \quad H_1 : \sigma^2 > \sigma_0^2 \quad (\text{teste unilateral direito});$
3. $H_0 : \sigma^2 \geq \sigma_0^2 \quad vs \quad H_1 : \sigma^2 < \sigma_0^2 \quad (\text{teste unilateral esquerdo}).$

- Vamos escolher a estatística de teste com base no estimador de σ^2 , S^2 , variância amostral:

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2} \underset{\text{sob } H_0}{\sim} \chi_{n-1}^2.$$

- Definamos a região de rejeição do teste: Para um nível de significância α , pré-especificado, as regiões de rejeição dos três tipos de hipóteses são, respectivamente, indicadas na figura 7.3. Ou seja, a região de rejeição do teste, para um nível de significância α pré-especificado é, respectivamente:

1. $R_\alpha =]0, \chi_{n-1, 1-\alpha/2}^2[\cup]\chi_{n-1, \alpha/2}^2, +\infty[\quad (\text{teste bilateral});$
2. $R_\alpha =]\chi_{n-1, \alpha}^2, +\infty[\quad (\text{teste unilateral direito});$
3. $R_\alpha =]0, \chi_{n-1, 1-\alpha}^2[\quad (\text{teste unilateral esquerdo});$

- Rejeitamos H_0 se $X_{obs}^2 \in R_\alpha$.

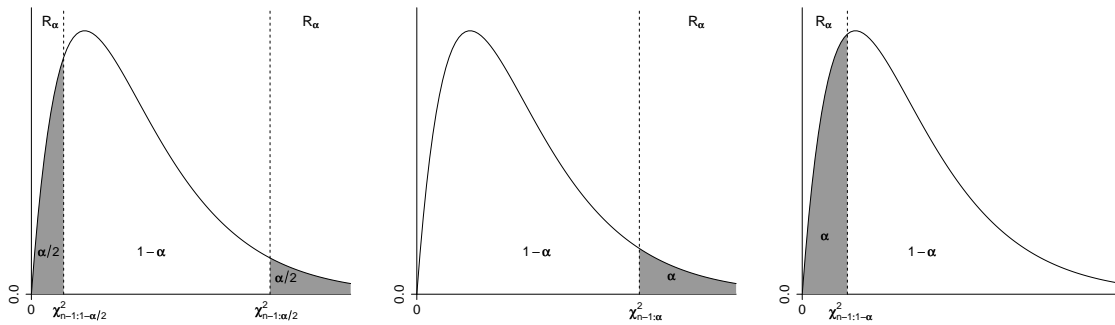


Figura 7.3: Esquerda: Região de rejeição para o teste bilateral. Centro: Região de rejeição para o teste unilateral direito. Direita: Região de rejeição para o teste unilateral esquerdo.

7.4 Teste de Hipóteses para a proporção p de uma população

Admita que temos uma amostra aleatória de dimensão n de uma população, em que determinada proporção desconhecida p dos seus elementos possui certa característica.

- Admita que pretendemos testar uma das seguintes hipóteses (nula e alternativa):

1. $H_0 : p = p_0$ vs $H_1 : p \neq p_0$ (teste bilateral);
2. $H_0 : p \leq p_0$ vs $H_1 : p > p_0$ (teste unilateral direito);
3. $H_0 : p \geq p_0$ vs $H_1 : p < p_0$ (teste unilateral esquerdo);

- Estatística de teste:

$$Z = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}} \underset{\text{sob } H_0}{\overset{a}{\sim}} N(0, 1)$$

- Definamos a região de rejeição do teste: Para um nível de significância α , pré-especificado, as regiões de rejeição dos três tipos de hipóteses são indicadas na figura 7.4.

Sendo, portanto a região de rejeição do teste, para um nível de significância α pré-especificado:

1. $R_\alpha =] - \infty, -z_{\alpha/2}[\cup] z_{\alpha/2}, +\infty[$ (teste bilateral);
2. $R_\alpha =] z_\alpha, +\infty[$ (teste unilateral direito);
3. $R_\alpha =] - \infty, -z_\alpha[$ (teste unilateral esquerdo);

- Regra de decisão do teste: Rejeitar H_0 ao nível de significância α se

$$z_{obs} = \frac{\hat{p}_{obs} - p_0}{\sqrt{p_0(1 - p_0)/n}} \in R_\alpha.$$

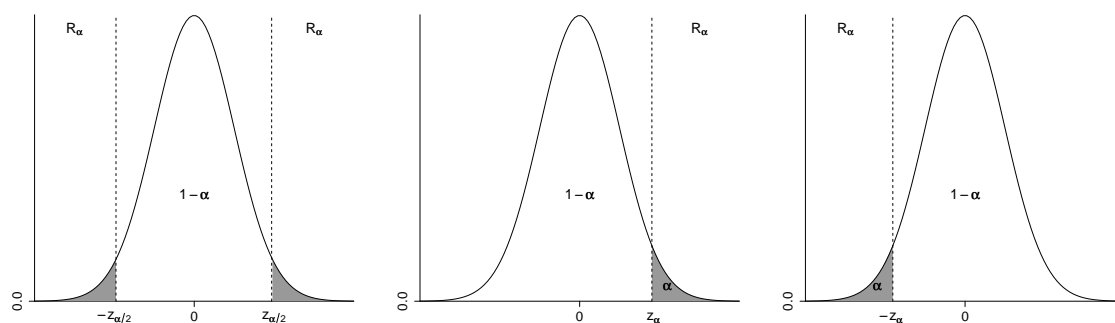


Figura 7.4: Esquerda: Região de rejeição para o teste bilateral. Centro: Região de rejeição para o teste unilateral direito. Direita: Região de rejeição para o teste unilateral esquerdo.

Capítulo 8

Regressão Linear

8.1 Introdução

A regressão é uma técnica estatística que permite estudar a relação entre uma ou mais **variáveis resposta** (também designadas por **variáveis dependentes**) e uma ou mais **variáveis explicativas** (também designadas por **variáveis independentes**). Ao modelo matemático que relaciona as variáveis dá-se o nome de **equação de regressão**.

Estamos apenas interessados no caso em que temos uma variável dependente Y , uma variável independente x e a equação de regressão é linear, isto é,

$$Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

O termo $\beta_0 + \beta_1 x$ é a componente determinística do modelo e ε é o erro aleatório que se pressupõe ter distribuição normal de valor médio nulo e variância σ^2 . Os parâmetros β_0 e β_1 terão de ser estimados a partir dos dados. A este modelo dá-se o nome de equação de **regressão linear simples**. Podemos também usar esta técnica considerando modelos mais complexos como a regressão linear múltipla ou a regressão não linear.

Observações:

1. Y também é uma variável aleatória porque, $Y = \beta_0 + \beta_1 x + \varepsilon$ e $\varepsilon \sim N(0, \sigma^2)$ é uma variável aleatória. Como

$$E(Y|x) = E(\beta_0 + \beta_1 x + \varepsilon|x) = \beta_0 + \beta_1 x + 0 = \beta_0 + \beta_1 x,$$

$$V(Y|x) = V(\beta_0 + \beta_1 x + \varepsilon|x) = V(\varepsilon) = \sigma^2,$$

isto é,

$$Y|x \sim N(\beta_0 + \beta_1 x, \sigma^2).$$

2. O modelo possui o parâmetro adicional, σ^2 , que também terá de ser estimado.

8.2 Estimadores dos Mínimos Quadrados de β_0 e β_1

Suponha que se observam um conjunto de n observações da variável independente e da variável resposta - $(x_1, Y_1), (x_2, Y_2) \dots, (x_n, Y_n)$ - e que se pretendem usar estes valores para estimar os parâmetros de regressão de um modelo de regressão linear simples. Assumimos que os erros aleatórios ε_i , para cada elemento amostral Y_i , são independentes seguindo todos a mesma distribuição $N(0, \sigma^2)$, isto é:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{com } \varepsilon_i \sim N(0, \sigma^2) \text{ independentes.}$$

Assim deveremos encontrar estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$, dos coeficientes da recta de regressão β_0 e β_1 , respectivamente, para obtermos a **recta estimada**,

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

As estimativas pontuais da recta de regressão para as observações x_1, x_2, \dots, x_n serão $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, 2, \dots, n$.

Definição 8.1 (Resíduo). *Embora a variável residual ε não seja observável, é possível calcular os desvios das n observações da amostra.*

$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, 2, \dots, n.$$

A estes desvios damos o nome de *resíduos*.

De entre diversos métodos que existem para a dedução dos estimadores, vamos aqui abordar o método dos mínimos quadrados. Neste método, os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ devem ser obtidos de modo a minimizar a soma do quadrado dos resíduos,

$$SQ_R = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Esta minimização é conseguida resolvendo, em ordem a β_0 e β_1 , o sistema de equações,

$$\begin{cases} \frac{\partial SQ}{\partial \hat{\beta}_0} = 0 \\ \frac{\partial SQ}{\partial \hat{\beta}_1} = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ -2 \sum x_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum Y_i = n \hat{\beta}_0 + \hat{\beta}_1 \sum x_i \\ \sum x_i Y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 \end{cases} \Leftrightarrow \begin{cases} \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum x_i Y_i - n \bar{x} \bar{Y}}{\sum x_i^2 - n \bar{x}^2} \end{cases}$$

Observação 1: Para simplificar a notação, podemos escrever:

$$\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x},$$

com

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2;$$

$$S_{xY} = \sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x}) = \sum_{i=1}^n Y_i(x_i - \bar{x}) = \sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}.$$

Observação 2: A soma dos quadrados dos desvios pode ainda ser escrita da seguinte forma

$$SQ_R = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = S_{YY} - \frac{S_{xY}^2}{S_{xx}} = S_{YY} - \hat{\beta}_1^2 S_{xx},$$

com

$$S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2.$$

8.3 Estimação de σ^2 e Qualidade do Ajuste

Definição 8.2 (Estimador de σ^2). O estimador de σ^2 é:

$$\hat{\sigma}^2 = \frac{SQ_R}{n-2}$$

Definição 8.3 (Coeficiente de Determinação).

$$R^2 = 1 - \frac{SQ_R}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \hat{\beta}_1^2 \frac{S_{xx}}{S_{YY}} = \frac{S_{xY}^2}{S_{xx} S_{YY}}$$

Esta medida compara a soma de quadrados dos resíduos (SQ_R) do modelo de regressão linear simples com a SQ_R do modelo de regressão linear simples com $\beta_1 = 0$. A quantidade R^2 varia entre 0 e 1. Na prática, consideramos que o ajustamento é razoável se $R^2 \geq 0.8$.

8.4 Propriedades dos estimadores dos mínimos quadrados

8.4.1 Distribuição por amostragem de $\hat{\sigma}^2$

Proposição 8.4 (Propriedades de $\hat{\sigma}^2$). No modelo de regressão linear simples,

$$(n-2) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{SQ_R}{\sigma^2} \sim \chi_{n-2}^2.$$

8.4.2 Distribuição por amostragem de $\hat{\beta}_0$ e $\hat{\beta}_1$

Proposição 8.5 (Distribuição por amostragem de $\hat{\beta}_0$ e $\hat{\beta}_1$). *No modelo de regressão linear simples,*

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right), \quad e \quad \hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2\right).$$

Demonstração. Note-se que

$$\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{S_{xx}},$$

isto é, $\hat{\beta}_1$ é uma combinação linear de v.a.'s Y_i independentes, com distribuição Normal. Logo $\hat{\beta}_1$ também tem distribuição Normal. E ainda necessário conhecer os seus parâmetros. O seu valor médio é

$$\begin{aligned} E(\hat{\beta}_1) &= \frac{\sum_{i=1}^n (x_i - \bar{x})E(Y_i)}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{S_{xx}} = \\ &= \frac{\beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i}{S_{xx}} = \frac{\beta_1 S_{xx}}{S_{xx}} = \beta_1 \end{aligned}$$

e a variância,

$$\begin{aligned} V(\hat{\beta}_1) &= V\left(\frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{S_{xx}}\right)_{Y_i's \text{ indep.}} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 V(Y_i)}{S_{xx}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{S_{xx}^2} \\ &= \frac{S_{xx}}{S_{xx}^2} \sigma^2 = \frac{\sigma^2}{S_{xx}}. \end{aligned}$$

Relativamente $\hat{\beta}_0$, recordemos que $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$. Como \bar{Y} e $\hat{\beta}_1$ têm distribuição Normal, então $\hat{\beta}_0$ também tem distribuição normal. O valor médio é

$$E(\hat{\beta}_0) = E(\bar{Y}) - E(\hat{\beta}_1)\bar{x} = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0,$$

e a variância,

$$\begin{aligned} V(\hat{\beta}_0) &= V(\bar{Y}) + \bar{x}^2 V(\hat{\beta}_1) - 2\bar{x} \text{Cov}(\bar{Y}, \hat{\beta}_1) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{S_{xx}} - 0 = \frac{\sigma^2}{n} \left(1 + \frac{n\bar{x}^2}{S_{xx}}\right) \\ &= \frac{\sigma^2}{nS_{xx}} (S_{xx} + n\bar{x}^2) = \frac{\sigma^2}{nS_{xx}} \left(\sum_{i=1}^n x_i^2\right). \end{aligned}$$

Nota: No cálculo de $V(\hat{\beta}_0)$, usou-se o resultado:

$$\begin{aligned} \text{Cov}(\bar{Y}, \hat{\beta}_1) &= \text{Cov}\left(\bar{Y}, \frac{S_{xY}}{S_{xx}}\right) = \frac{1}{S_{xx}} \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n Y_i, \sum_{j=1}^n (x_j - \bar{x})Y_j\right) \\ &= \frac{1}{nS_{xx}} \sum_{i=1}^n (x_i - \bar{x}) V(Y_i) = \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n (x_i - \bar{x}) = 0 \end{aligned}$$

□

Observação: A partir do resultado anterior conclui-se que $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores centrados de β_0 e β_1 , respectivamente.

Consequentemente, querendo fazer inferência sobre os parâmetros β_0 ou β_1 , não podemos usar a distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$, já que elas dependem de σ^2 (geralmente é desconhecido). Como $\hat{\sigma}^2 = \frac{SQ_R}{n-2}$, teremos de usar os seguintes resultados:

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} = \sqrt{S_{xx}} \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}} \sim t_{n-2},$$

$$T = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{nS_{xx}} \sum_{i=1}^n x_i^2}} = \sqrt{\frac{nS_{xx}}{\sum_{i=1}^n x_i^2}} \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}} \sim t_{n-2}.$$

8.5 Inferência sobre os parâmetros do Modelo de Regressão

8.5.1 Intervalo de Confiança e Teste de Hipóteses para β_1

O parâmetro β_1 é o declive da recta de regressão e, como tal mede o grau de crescimento de Y relativamente a x .

Intervalo de confiança a $(1 - \alpha)100\%$ para β_1

- Vamos utilizar a seguinte variável pivot:

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2}$$

- Para um nível de confiança de $(1 - \alpha) \times 100\%$, escolha de c_1 e c_2 - escolhemos $c_1 = -c$ e $c_2 = c$, tal que $P(-c < T < c) = 1 - \alpha$. É fácil de verificar que $c = t_{n-2, \alpha/2}$.
- Determinação dos extremos do intervalo:

$$\begin{aligned} -c < T < c &\Leftrightarrow -c\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} < \hat{\beta}_1 - \beta_1 < c\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \Leftrightarrow \\ -c\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} - \hat{\beta}_1 &< -\beta_1 < c\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} - \hat{\beta}_1 \Leftrightarrow \\ \hat{\beta}_1 - t_{n-2, \alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} &< \beta_1 < \hat{\beta}_1 + t_{n-2, \alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \end{aligned}$$

- Assim, obtemos o seguinte intervalo de confiança:

$$IC_{(1-\alpha) \times 100\%}(\beta_1) = \left] \hat{\beta}_1 - t_{n-2, \alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \hat{\beta}_1 + t_{n-2, \alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \right[.$$

Teste de Hipóteses para β_1

Podemos também realizar um teste de hipóteses sobre o valor do parâmetro β_1 . Embora o teste tanto possa ser bilateral, como unilateral, a primeira opção é a mais frequente. Por isso apenas apresentamos o teste bilateral, embora este possa ser adaptado para o caso unilateral.

- Hipóteses:

$$H_0 : \beta_1 = a \quad vs \quad H_1 : \beta_1 \neq a$$

- Estatística de teste:

$$T = \frac{\sqrt{S_{xx}} \hat{\beta}_1 - a}{\hat{\sigma}} \underset{Sob H_0}{\sim} t_{n-2}$$

- Região de rejeição do teste:

$$R_\alpha =] - \infty; -t_{n-2, \alpha/2}[\cup] t_{n-2, \alpha/2}; +\infty[$$

- Regra de decisão do teste: Rejeitar H_0 ao nível de significância α se

$$t_{obs} \in R_\alpha, \text{ ou seja, se } |t_{obs}| > t_{n-2, \alpha/2}.$$

8.5.2 Intervalo de Confiança e Teste de Hipóteses para β_0

O parâmetro β_0 corresponde ao ponto de intersecção da recta com o eixo das abcissas. A inferência sobre este parâmetro não tem a mesma importância que tem a inferência sobre o declive β_1 da recta de regressão.

Intervalo de Confiança a $(1 - \alpha) \times 100\%$ para β_0

De modo análogo, ao que foi feito para β_1 , mas agora utilizando a variável pivot,

$$T = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{nS_{xx}} \sum_{i=1}^n x_i^2}} = \sqrt{\frac{nS_{xx}}{\sum_{i=1}^n x_i^2}} \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}} \sim t_{n-2},$$

obtemos o intervalo de confiança $(1 - \alpha) \times 100\%$ para β_0 :

$$IC_{(1-\alpha) \times 100\%}(\beta_0) \equiv \left] \hat{\beta}_0 - t_{n-2, \alpha/2} \sqrt{\hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{nS_{xx}}}; \hat{\beta}_0 + t_{n-2, \alpha/2} \sqrt{\hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{nS_{xx}}} \right[.$$

Testes de hipóteses para β_0

Os testes de hipóteses sobre o parâmetro β_0 podem ser tanto bilaterais como unilaterais, sendo

sempre baseados na distribuição por amostragem anteriormente apresentada para $\hat{\beta}_0$. Vamos considerar apenas o teste bilateral para β_0 , ou seja, as hipóteses:

$$H_0 : \beta_0 = a \quad vs \quad H_1 : \beta_0 \neq a$$

O teste realiza-se de modo análogo ao apresentado para β_1 , mudando apenas a estatística de teste que é dada por

$$T = \frac{\hat{\beta}_0 - a}{\sqrt{\frac{\hat{\sigma}^2}{nS_{xx}} \sum_{i=1}^n x_i^2}} \underset{\text{Sob } H_0}{\sim} t_{n-2}$$

8.5.3 Intervalo de Confiança e Teste de Hipóteses para σ^2

Como $\hat{\sigma}^2 = \frac{SQ_R}{n-2}$ é estimador centrado de σ^2 e $\frac{SQ_R}{\sigma^2} \sim \chi_{n-2}^2$, podemos deduzir um intervalo de confiança $(1 - \alpha)$ para a variância σ^2 e para o desvio padrão σ . Seguindo o procedimento adoptado, na secção 6.2, obtemos

$$IC_{(1-\alpha) \times 100\%}(\sigma^2) \equiv \left[\frac{(n-2)\hat{\sigma}^2}{\chi_{n-2; \alpha/2}^2} ; \frac{(n-2)\hat{\sigma}^2}{\chi_{n-2; 1-\alpha/2}^2} \right],$$

e

$$IC_{(1-\alpha) \times 100\%}(\sigma) \equiv \left[\sqrt{\frac{(n-2)\hat{\sigma}^2}{\chi_{n-2; \alpha/2}^2}} ; \sqrt{\frac{(n-2)\hat{\sigma}^2}{\chi_{n-2; 1-\alpha/2}^2}} \right].$$

De modo análogo ao apresentado na secção 7.3, podemos também realizar testes de hipótese (bilaterais e unilaterais) para σ^2 recorrendo à distribuição de $\hat{\sigma}^2$.

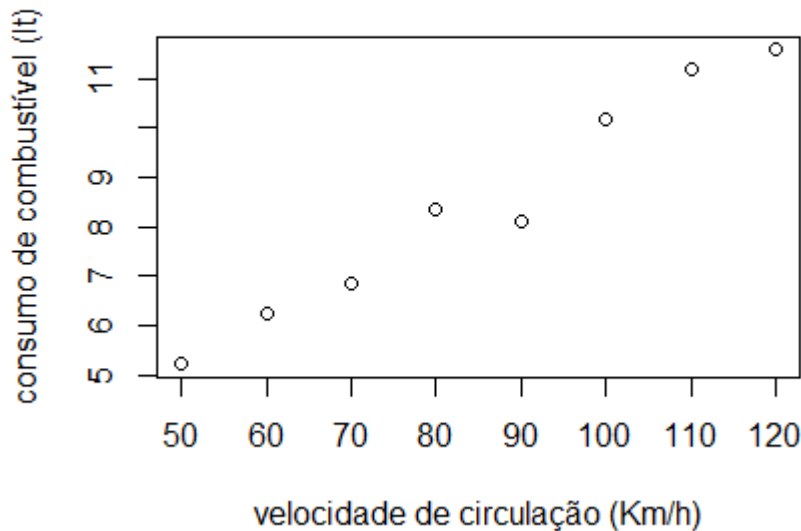
Exemplo 8.6 (Exame de Probabilidades e Estatística C 2005/06). *Pretende-se, se possível, modelar através de uma recta de regressão simples o consumo de combustível, Y , de um automóvel em função da sua velocidade de circulação, x . Para tal registaram-se os valores de consumo de combustível para um mesmo percurso de 100Km, percorrido a diferentes velocidades:*

| | | | | | | | | |
|-------|------|------|------|------|------|-------|-------|-------|
| x_i | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 |
| y_i | 5.22 | 6.25 | 6.85 | 8.36 | 8.09 | 10.16 | 11.17 | 11.57 |

$$\bar{x} = 85, \quad \bar{Y} = 8.46, \quad \sum x_i^2 = 62000, \quad \sum Y_i^2 = 610.43, \quad \sum Y_i x_i = 6145.5, \quad SQ_R = 1.15$$

Podemos fazer uma análise gráfica antes de considerar um modelo de regressão, assim podemos obter o gráfico de dispersão utilizando o comando `plot` em **R**:

```
> x<-c(50,60,70,80,90,100,110,120)           #velocidade de circulação
> Y<-c(5.22,6.25,6.85,8.36,8.09,10.16,11.17,11.57) #consumo de combustível
> plot(x,Y,xlab="velocidade de circulação (Km/h)",
      ylab="consumo de combustível (lt)")      #gráfico de dispersão (x,Y)
```



O gráfico sugere que se tente ajustar um modelo de regressão linear simples. Assim a reta ajustada é $\hat{Y} = 0.494048 + 0.093702x$. As estimativas dos parâmetros da reta podem ser obtidos em [R](#) de várias formas:

Podemos fazer o cálculo utilizando as fórmulas seguintes:

```

> n<-length(x)                #dimensão da amostra
> sxx<-sum(x^2)-n*mean(x)^2    #valor de Sxx
> sxy<-sum(x*Y)-n*mean(x)*mean(Y) #valor de SxY
> beta1est<-sxy/sxx; beta1est  #estimativa de beta1
[1] 0.09370238
> beta0est<-mean(Y)-beta1est*mean(x);beta0est #estimativa de beta0
[1] 0.4940476

```

ou através dos comandos `lm()` e `summary()` em [R](#).

O comando `summary()` fornece muito mais do que o valores das estimativas dos parâmetros. A tabela dos "Coefficients" tem duas linhas de valores, sendo a primeira, "(Intercept)", referente à ordenada na origem (β_0) e a segunda "x" correspondente ao verdadeiro declive da reta (β_1). Na coluna "Estimate" temos as estimativas dos parâmetros da reta, na coluna "Std. Error" temos as estimativas dos erros padrão dos dois estimadores. Na coluna "t value" temos, respetivamente, os valores observados da estatística dos testes de hipóteses:

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$$

e

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0.$$

Por fim, na última coluna " $Pr(> |t|)$ " temos os valores $-p$ para cada um destes testes. Podemos assim verificar que para o teste $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$, o valor observado da estatística

de teste é 13.869 e tem valor $-p = 8.75 \times 10^{-6}$, rejeitando-se hipótese H_0 para um nível de significância, por exemplo, de 1%. Com o comando `summary()` ainda se obtêm outras informações como, por exemplo, o valor do coeficiente de determinação, R^2 , que é o valor 0.9697 apresentado em "Multiple R-squared" e a estimativa do desvio padrão dos erros do modelo, $\hat{\sigma}$, que é o valor 0.4379 apresentado em "Residual standard error".

```
> rlinear<-lm(Y~ x) #regressão linear
> summary(rlinear) #sumário da regressão linear

Call:
lm(formula = Y ~x)

Residuals:
    Min       1Q   Median       3Q      Max
-0.83726 -0.17705  0.08732  0.31396  0.36976

Coefficients:
            Estimate Std. Error t value Pr(> |t|)
(Intercept)  0.494048  0.594784   0.831   0.438
x             0.093702  0.006756  13.869 8.75e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4379 on 6 degrees of freedom
Multiple R-squared:  0.9697, Adjusted R-squared:  0.9647
F-statistic: 192.3 on 1 and 6 DF, p-value: 8.75e-06
```

O valor do coeficiente de determinação, R^2 , podia também ser obtido em [R](#) da seguinte forma:

```
> syy<-sum(Y^2)-n*mean(Y)^2 #valor de SYY
> r2<-sxy^2/(sxx*syy); r2 #valor do coeficiente de determinação
[1] 0.9697499
```

Para explorar melhor o objeto `lm()`, que no nosso exemplo de [R](#) foi atribuído a `rlinear`, podemos utilizar o comando `attributes()` obtendo-se os seguintes atributos do objeto:

```
> attributes(rlinear)
$names
 [1] "coefficients" "residuals" "effects" "rank"
 [5] "fitted.values" "assign" "qr" "df.residual"
 [9] "xlevels" "call" "terms" "model"
$class
 [1] "lm"
```

Com estes atributos é possível extrair os valores contidos no objeto da regressão linear simples:


```

> rlinear$coefficients      #estimativas dos coeficientes da reta de regressão
(Intercept)                x
0.49404762 0.09370238
> rlinear$coefficients[1]  #ordenada na origem
(Intercept)
0.4940476
> rlinear$coefficients[2]  #declive da reta
x
0.09370238

> rlinear$fitted.values    #estimativas de Y
      1      2      3      4      5      6      7      8
5.179167 6.116190 7.053214 7.990238 8.927262 9.864286 10.801310 11.738333

```

Onde temos o vetor das estimativas dos parâmetros da reta dado pelo comando `rlinear$coefficients` e as estimativas de Y dadas pela reta ajustada, \hat{Y}_i , $i = 1, \dots, 8$, obtidas com o comando `rlinear$fitted.values`.

Também podemos obter o vetor dos resíduos, ϵ_i , $i = 1, \dots, 8$, através do comando `rlinear$residuals`:

```

> rlinear$residuals      #resíduos
      1      2      3      4      5      6
0.04083333 0.13380952 -0.20321429 0.36976190 -0.83726190 0.29571429
      7      8
0.36869048 -0.16833333

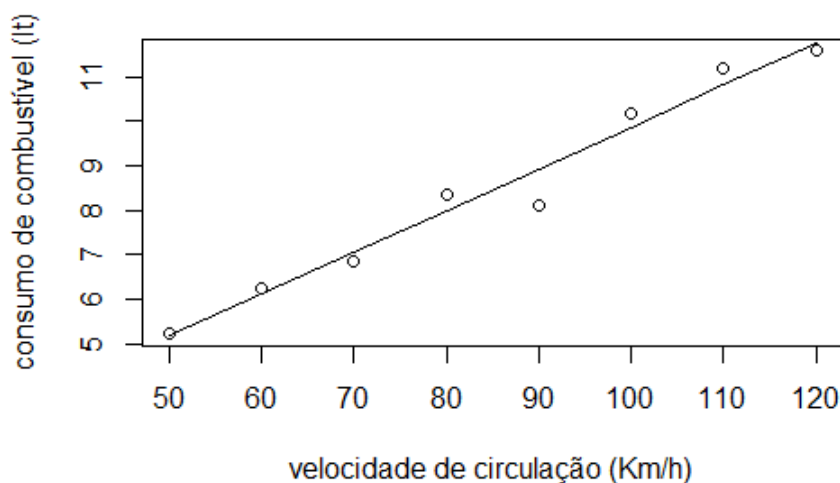
```

Se quisermos adicionar no gráfico de dispersão a reta ajustada podemos utilizar o comando `lines()`.

```

> plot(x, Y, xlab="velocidade de circulação (Km/h)",
      ylab="consumo de combustível (lt)"),      #gráfico de dispersão (x, Y)
> lines(x, rlinear$fitted)                    #reta estimada

```



8.6 Estimação do valor esperado de Y para uma observação x_0 da variável controlada

O valor esperado de Y para uma observação x_0 da variável controlada é

$$\mu_{Y|x_0} = E(Y|x_0) = \beta_0 + \beta_1 x_0.$$

que pode ser estimado por

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0.$$

Caso a variância do erro, σ^2 , não seja conhecida, a distribuição de amostragem de $\hat{\mu}_{Y|x_0}$ é

$$T = \frac{\hat{\mu}_{Y|x_0} - \mu_{Y|x_0}}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}} \sim t_{n-2},$$

o que permite deduzir o intervalo de confiança $(1 - \alpha)$ para $\mu_{Y|x_0}$,

$$\left[\hat{\mu}_{Y|x_0} - t_{n-2; \alpha/2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}, \hat{\mu}_{Y|x_0} + t_{n-2; \alpha/2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \right].$$

Nota: Só devemos fazer estimação de $\mu_{Y|x_0}$ para valores x_0 que estejam dentro do intervalo das observações obtidas para x .

8.7 Previsão do valor da variável resposta Y para um novo valor x_0 da variável controlada

Dada um valor x_0 da variável controlada x , a variável resposta é

$$Y_0 = Y(x_0) = \beta_0 + \beta_1 x_0 + \varepsilon,$$

onde $\varepsilon \sim N(0, \sigma^2)$. O estimador de Y , para um valor x_0 , é $\hat{Y}_0 = \hat{Y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$

O erro de predição, $\varepsilon_p = Y_0 - \hat{Y}_0$, é uma v.a. Normal de valor médio 0. Como Y_0 (observação futura) é independente de \hat{Y}_0 , a variância de ε_p é de dada por

$$V(\varepsilon_p) = V(Y_0 - \hat{Y}_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right).$$

Se σ^2 for estimado por $\hat{\sigma}^2$, então $T = \frac{Y_0 - \hat{Y}_0}{\sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}} \sim t_{n-2}$.

O intervalo de confiança $(1 - \alpha)$ para Y_0 é,

$$\left[\hat{Y}_0 - t_{n-2; \alpha/2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}, \hat{Y}_0 + t_{n-2; \alpha/2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \right].$$

Bibliografia sugerida

- Cordeiro e Magalhães(2004). *Introdução à Estatística. Uma perspectiva química*. Lidel-Edições Técnicas, Lda.
- Guimarães e Cabral(1997). *Estatística*. McGraw-Hill.
- Montgomery e Runger (2002). *Applied Statistics and Probability for Engineers*. Wiley.
- Mood, Graybill e Boes (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.
- Murteira, B., Ribeiro, C., Silva, J. e Pimenta, C. (2007). *Introdução à Estatística*, 2ª edição. McGraw-Hill
- Paulino e Branco (2005). *Exercícios de Probabilidade e Estatística*. Escolar Editora.
- Pestana, D. e Velosa, S. (2002). *Introdução à Probabilidade e à Estatística*. Fundação Calouste Gulbenkian, Lisboa.
- Rohatgi (1976). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley.
- Sokal e Rohlf (1995). *Biometry*. Freeman.
- Tiago de Oliveira (1990). *Probabilidades e Estatística: Conceitos, Métodos e Aplicações*, vol. I, II. McGraw-Hill.