

# Capítulo 2

## Momentos amostrais

Quando estudadas as ideias e técnicas da teoria das probabilidades fundamentais, criamos um modelo matemático de um ensaio aleatório, associando-o com um espaço de amostragem no qual os acontecimentos aleatórios correspondem a um conjunto de uma certa  $\sigma$ -álgebra. A noção de probabilidade definida nesta  $\sigma$ -álgebra corresponde à noção de incerteza no resultado em qualquer realização do experimento aleatório.

Vamos começar o estudo de alguns problemas de estatística matemática. Suponha que buscamos informações sobre algumas características numéricas de um conjunto de elementos, chamado de uma população. Por razões de custo, de tempo ou simplesmente para não destruir todos os elementos amostrados podemos não querer ou não poder estudar cada elemento da população. Nosso objetivo é tirar conclusões sobre as características desconhecidas da população com base em informações sobre algumas características de uma amostra adequadamente selecionada.

Formalmente, seja  $X$  uma variável aleatória que descreve a população sob investigação e seja  $F$  a função de distribuição de  $X$ . Há duas possibilidades: ou  $X$  tem função de distribuição  $F(\cdot; \theta)$  com uma forma funcional conhecida exceto, talvez, para o parâmetro  $\theta$ , o qual pode ser um vetor ou  $X$  tem uma função de distribuição  $F$  sobre a qual não sabemos nada, exceto talvez que  $F$  seja, digamos, absolutamente contínua. No primeiro caso, seja  $\Theta$  o conjunto dos possíveis valores do parâmetro desconhecido  $\theta$ . Seguidamente, o trabalho de um estatístico é decidir com base em uma amostra selecionada adequadamente que membro ou membros da família  $\{F(\cdot; \theta) : \theta \in \Theta\}$  pode representar a função de distribuição de  $X$ . Problemas desse tipo são chamados de problemas de inferência estatística paramétrica e o espaço estatístico é dado por  $(\mathbb{R}, \{F(\cdot; \theta) : \theta \in \Theta\}, \mathcal{B}(\mathbb{R}))$ , sendo  $\mathbb{R}$  o conjunto dos reais na reta,  $\{F(\cdot; \theta) : \theta \in \Theta\}$  a família de distribuições de  $X$  e  $\mathcal{B}(\mathbb{R})$  a  $\sigma$ -álgebra dos Borelianos na reta. O caso em que nada se sabe sobre a forma funcional da função de distribuição  $F$  de  $X$  é claramente muito mais difícil. Problemas de inferência deste tipo de são o domínio de estudo da estatística não paramétrica. Neste livro abordamos os problemas da estatística paramétrica.

### 2.1 Amostras aleatórias

Considere-se um experimento estatístico que culmina em desfechos  $x$ , que são os valores assumidos por uma variável aleatória  $X$ . Seja  $F$  a função de distribuição de  $X$ . Na prática,  $F$  não será completamente conhecida, isto é, um ou mais parâmetros associados com  $F$  serão desconhecidos. O trabalho de um estatístico é estimar esses parâmetros desconhecidos ou testar a validade de certas afirmações sobre eles. Ele pode, por exemplo, obter  $n$  observações independentes de  $X$ . Isso significa que ele observa  $n$  valores  $x_1, x_2, \dots, x_n$  assumidos da variável aleatória  $X$ . Cada  $x_i$  pode ser considerado como o valor assumido pela variável aleatória  $X_i$ ,  $i = 1, 2, \dots, n$  onde  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes com distribuição comum  $F$ . Os valores observados  $(x_1, x_2, \dots, x_n)$  são então valores assumidos por  $(X_1, X_2, \dots, X_n)$ . O conjunto  $(X_1, X_2, \dots, X_n)$  é, então, uma amostra de tamanho  $n$  da distribuição da população  $F$ . O conjunto de  $n$  valores  $x_1, x_2, \dots, x_n$  é chamado de uma realização ou estimativa da amostra. Note-se que os possíveis valores do vector aleatório  $(X_1, X_2, \dots, X_n)$  podem ser olhados como pontos em  $\mathbb{R}^n$ , os quais podem ser chamados de elementos do espaço amostral. Na prática podemos não observar  $x_1, x_2, \dots, x_n$

mas alguma função  $g(x_1, x_2, \dots, x_n)$ . Então  $g(x_1, x_2, \dots, x_n)$  serão considerados os valores assumidos pela variável aleatória  $g(X)$ .

Vamos agora formalizar esses conceitos.

### Definição 2.1

Seja  $X$  uma variável aleatória com função de distribuição  $F$  e  $X_1, \dots, X_n$  variáveis aleatórias independentes com distribuição comum  $F$ . Chamaremos a coleção  $X_1, \dots, X_n$  de uma amostra aleatória de tamanho  $n$  de  $F$  ou simplesmente como  $n$  observações independentes de  $X$ .

Se  $X_1, X_2, \dots, X_n$  é uma amostra aleatória de  $F(\cdot; \theta)$ , a função de distribuição conjunta é dada por

$$F(x_1, \dots, x_n; \theta) = \prod_{i=1}^n F(x_i; \theta). \quad (2.1)$$

### Definição 2.2

Sejam  $X_1, X_2, \dots, X_n$   $n$  observações independentes da variável aleatória  $X$  e seja  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  uma função real derivável. Então a variável aleatória  $g(X_1, X_2, \dots, X_n)$  é chamada de estatística amostral ou simplesmente estatística, desde que não dependa de parâmetros desconhecidos.

Segundo esta definição cada uma das variáveis na amostra isoladamente é uma estatística assim como funções destas que, eventualmente, podem não fornecer informações úteis. Duas das estatísticas mais comumente utilizadas são mostradas no exemplo a seguir.

### Definição 2.3

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória da função de distribuição  $F$ . A estatística

$$\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}, \quad (2.2)$$

é chamada de média amostral e a estatística

$$S_n^2 = \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{n-1}, \quad (2.3)$$

é chamada de variância amostral.

Deve-se lembrar que as estatísticas amostrais apresentadas neste exemplo  $\bar{X}_n$ ,  $S_n^2$  e outras que irão definir-se posteriormente são variáveis aleatórias, com todas as consequências que isso implica, enquanto os parâmetros populacionais  $\mu$ ,  $\sigma^2$  e assim por diante são constantes fixas, que podem ser desconhecidas.

### Exemplo 2.1

Seja  $X \sim \text{Bernoulli}(p)$ , onde  $p$  é desconhecido. A função de distribuição de  $X$ , mostrada na Figura 2.1, é dada

por

$$F(x) = p\delta(x - 1) + (1 - p)\delta(x), \quad x \in \mathbb{R},$$

onde a função  $\delta(\cdot)$  foi definida em (1.2) como  $\delta(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases}$ , chamada de função delta.

Suponha que cinco observações independentes de  $X$  sejam 0, 1, 1, 1, 0. Então 0, 1, 1, 1, 0 é uma realização da amostra  $X_1, X_2, \dots, X_5$ . A estimativa da média amostral é

$$\bar{x}_5 = \frac{0 + 1 + 1 + 1 + 0}{5} = 0,6$$

o qual é o valor assumido pela variável aleatória  $\bar{X}_n$ . A estimativa da variância amostral é

$$s_5^2 = \sum_{i=1}^5 \frac{(x_i - \bar{x}_5)^2}{5 - 1} = \frac{2 \times (0,6)^2 + 3 \times (0,4)^2}{4} = 0,3$$

sendo este o valor assumido pela variável aleatória  $S_n^2$ .

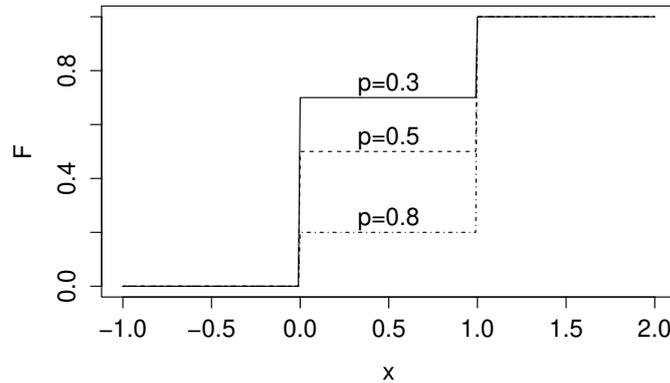


Figura 2.1: Representação da função de distribuição Bernoulli para três valores do parâmetro  $p = 0.3, 0.5$  e  $0.8$ . Observe que nesta curva a reta no intervalo  $(0, 1)$  depende de  $1 - p$ , isso porque a função  $\delta$  é sempre zero para  $x - 1$  nesse intervalo.

### Exemplo 2.2

Seja  $X \sim N(\mu, \sigma^2)$ ,  $\mu$  conhecida,  $\sigma^2$  desconhecido e  $X_1, \dots, X_n$  uma amostra aleatória dessa distribuição. De acordo com nossa definição, a função  $\sum_{i=1}^n X_i / \sigma^2$  não é uma estatística. Suponha que cinco observações de  $X$  são  $-0,864; 0,561; 2,355; 0,582$  e  $-0,774$ . Então, a estimativa da média amostral é  $0,372$  e a estimativa da variância amostral é  $1,648$ .

## 2.2 Estatísticas de ordem

Seja  $(X_1, X_2, \dots, X_n)$  um vetor aleatório  $n$ -dimensional e  $(x_1, x_2, \dots, x_n)$  uma  $n$ -tupla assumida por  $(X_1, X_2, \dots, X_n)$ . Vamos organizar  $x_1, x_2, \dots, x_n$  em ordem crescente de magnitude, para que

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

onde  $x_{(1)} = \min(x_1, x_2, \dots, x_n)$ ,  $x_{(2)}$  é o segundo menor valor em  $x_1, \dots, x_n$  e assim por diante,  $x_{(n)} = \max(x_1, \dots, x_n)$ . Se quaisquer dois  $x_i, x_j$  forem iguais, a ordem não importa.

**Definição 2.4**

A função  $X_{(k)}$  de  $(X_1, \dots, X_n)$  que assume o valor  $x_{(k)}$  em cada possível sequência  $(x_1, x_2, \dots, x_n)$  de valores assumidos por  $(X_1, X_2, \dots, X_n)$  é conhecida como a  $k$ -ésima estatística de ordem ou a estatística de ordem  $k$ . O conjunto  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  é chamado de estatísticas de ordem para  $(X_1, X_2, \dots, X_n)$ .

**Exemplo 2.3**

Consideremos  $X_1, X_2, X_3$  três variáveis aleatórias discretas de maneira que,  $X_1$  e  $X_3$  sejam tais que assumam somente valores 0, 1 e que  $X_2$  assumam valores 1, 2, 3. O vetor aleatório  $(X_1, X_2, X_3)$  assume os valores:  $(0, 1, 0)$ ,  $(0, 2, 0)$ ,  $(0, 3, 0)$ ,  $(0, 1, 1)$ ,  $(0, 2, 1)$ ,  $(0, 3, 1)$ ,  $(1, 1, 0)$ ,  $(1, 2, 0)$ ,  $(1, 3, 0)$ ,  $(1, 1, 1)$ ,  $(1, 2, 1)$  e  $(1, 3, 1)$ . Então  $X_{(1)}$  assume somente valores 0 ou 1;  $X_{(2)}$  assume somente valores 0 ou 1 e  $X_{(3)}$  assume somente valores 1, 2 ou 3.

**Teorema 2.1**

Seja  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  um vetor aleatório de dimensão  $n$  e seja  $X_{(k)}$ ,  $1 \leq k \leq n$ , a  $k$ -ésima estatística de ordem. Então  $X_{(k)}$  é também uma variável aleatória.

**Demonstração:** Exercício. ■

Na apresentação dos resultados a seguir assumiremos que  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes e igualmente distribuídas contínuas com função de densidade  $f$ . Seja  $\{X_{(1)}, X_{(2)}, \dots, X_{(n)}\}$  o conjunto das estatísticas de ordem para  $X_1, X_2, \dots, X_n$ . Dado que todas as  $X_i$  são contínuas segue que, com probabilidade 1

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

**2.2.1 Propriedades das estatísticas de ordem**

Começaremos o estudo das propriedades encontrando a função de densidade conjunta de  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ .

**Teorema 2.2**

Sejam  $X_1, \dots, X_n$  variáveis aleatórias contínuas independentes, igualmente distribuídas com densidade  $f$ . A função de densidade conjunta de  $(X_{(1)}, \dots, X_{(n)})$  é dada por

$$f(x_{(1)}, \dots, x_{(n)}) = \begin{cases} n! \prod_{i=1}^n f(x_{(i)}), & x_{(1)} < x_{(2)} < \dots < x_{(n)} \\ 0, & \text{caso contrário} \end{cases} \quad (2.4)$$

**Demonstração:** A transformação de  $(X_1, \dots, X_n)$  a  $(X_{(1)}, \dots, X_{(n)})$  não é biunívoca. De fato, existem um total de  $n!$  possíveis arranjos de  $x_1, \dots, x_n$  em ordem crescente de magnitude. Assim, existem  $n!$  inversas para a transformação. Por exemplo, uma das  $n!$  permutações pode ser

$$x_4 < x_1 < x_{n-1} < x_3 < \dots < x_n < x_2.$$

A inversa correspondente é

$$x_4 = x_{(1)}, x_1 = x_{(2)}, x_{n-1} = x_{(3)}, x_3 = x_{(4)} \cdots x_n = x_{(n-1)}, x_2 = x_{(n)}.$$

O determinante Jacobiano desta transformação é a matriz  $n \times n$  identidade com as colunas reorganizadas, isto devido a que cada  $x_{(i)}$  é igual a uma, e somente uma, das  $x_1, x_2, \dots, x_n$ . Portanto  $J = \pm 1$  e

$$f(x_{(2)}, x_{(n)}, x_{(4)}, x_{(1)}, \dots, x_{(3)}, x_{(n-1)})|J| = \prod_{i=1}^n f(x_{(i)}),$$

quando  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ . A mesma expressão é válida para cada um dos  $n!$  arranjos. Segue então que

$$\begin{aligned} f(x_{(1)}, \dots, x_{(n)}) &= \sum_{\substack{\text{Todas as } n! \\ \text{permutações}}} \prod_{i=1}^n f(x_{(i)}) \\ &= \begin{cases} n! \prod_{i=1}^n f(x_{(i)}), & \text{caso } x_{(1)} < x_{(2)} < \dots < x_{(n)} \\ 0, & \text{caso contrário} \end{cases} \end{aligned}$$

■

**Exemplo 2.4**

Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes com função de densidade comum

$$f(x) = \begin{cases} 1, & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Então a função de densidade conjunta de  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  é

$$f(x_{(1)}, \dots, x_{(n)}) = \begin{cases} n!, & 0 < x_{(1)} < x_{(2)} < \dots < x_{(n)} \\ 0, & \text{caso contrário} \end{cases} \tag{2.5}$$

Estamos confiados que como resultado do Teorema 2.2 temos funções de densidade. Vejamos neste exemplo se isso é realmente acontece. Consideremos, para simplificar, o caso  $n = 3$  e verifiquemos se a integral da função de densidade em (2.5) é 1. Então

$$\begin{aligned} \iiint_{\mathbb{R}} f(x_{(1)}, x_{(2)}, x_{(3)}) dx_{(1)} dx_{(2)} dx_{(3)} &= 6 \int_0^1 \left[ \int_{x_{(1)}}^1 \left( \int_{x_{(2)}}^1 dx_{(3)} \right) dx_{(2)} \right] dx_{(1)} \\ &= 6 \int_0^1 \left[ \int_{x_{(1)}}^1 (1 - x_{(2)}) dx_{(2)} \right] dx_{(1)} \\ &= 6 \int_0^1 \left[ \frac{1}{2} - x_{(1)} + \frac{x_{(1)}^2}{2} \right] dx_{(1)} = 1. \end{aligned}$$

Um detalhe interessante é que esta e outras propriedades demonstradas aqui somente são válidas quando as variáveis aleatórias são contínuas. Isso não significa que estatísticas de ordem não possam ser definidas no caso discreto. O que estamos dizendo é que estas propriedades somente podem ser demonstradas no caso contínuo.

**Exemplo 2.5**

Consideremos a situação em que temos somente três variáveis aleatórias independentes  $X_1, X_2$  e  $X_3$  com

distribuição geométrica de parâmetro  $p$ , isto é,

$$P(X = x; p) = (1 - p)p^x, \quad x = 0, 1, 2, \dots$$

Encontremos  $P(X_{(1)} < X_{(2)} < X_{(3)})$ .

Nesta situação a probabilidade requerida pode ser escrita como:

$$\begin{aligned} P(X_{(1)} < X_{(2)} < X_{(3)}) &= 1 - P(X_1 = X_2 \neq X_3) - P(X_1 = X_3 \neq X_2) \\ &\quad - P(X_2 = X_3 \neq X_1) - P(X_1 = X_2 = X_3) \end{aligned}$$

a qual pode ser escrita como

$$\begin{aligned} P(X_{(1)} < X_{(2)} < X_{(3)}) &= 1 - 3P(X_1 = X_2 \neq X_3) - P(X_1 = X_2 = X_3) \\ &= 1 - 3[P(X_1 = X_2) - P(X_1 = X_2 = X_3)] - P(X_1 = X_2 = X_3) \\ &= 1 - 3P(X_1 = X_2) + 2P(X_1 = X_2 = X_3). \end{aligned}$$

Não é difícil perceber que

$$P(X_1 = X_2) = \frac{(1 - p)^2}{1 - p^2},$$

e que

$$P(X_1 = X_2 = X_3) = \frac{(1 - p)^3}{1 - p^3},$$

do qual obtemos que

$$P(X_{(1)} < X_{(2)} < X_{(3)}) = \frac{6p^3}{(1 - p)(1 + p + p^2)}.$$

As propriedades das estatísticas de ordem que serão demonstradas valerão somente caso as variáveis sejam contínuas. Isto deve-se a que, caso as variáveis sejam discretas, a probabilidade

$$P(X_{(1)} = X_{(2)} = \dots = X_{(n)}) \neq 0,$$

como vai ser mostrado no seguinte exemplo. Acontece que o fato da probabilidade das estatística de ordem poderem coincidir, com probabilidade diferente de zero, altera a estrutura da demonstração e não nos permite obtermos estes resultados para o caso discreto.

### Exemplo 2.6

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes assumindo somente 0 e 1 com probabilidade 1/2. Observemos que

$$\begin{aligned} P(X_{(1)} = X_{(2)} = \dots = X_{(n)}) &= \prod_{k=1}^n P(X_{(k)} = 0) + \prod_{k=1}^n P(X_{(k)} = 1) \\ &= \prod_{k=1}^n P(X_k = 0) + \prod_{k=1}^n P(X_k = 1) = \frac{1}{2^{n-1}}. \end{aligned}$$

Estudemos agora o comportamento marginal, ou seja, nos interessa agora encontrar a função de distribuição marginal de cada estatística de ordem.

**Teorema 2.3**

Sejam  $X_1, \dots, X_n$  variáveis aleatórias contínuas independentes e igualmente distribuídas e  $(X_{(1)}, \dots, X_{(n)})$  as estatísticas de ordem. A função de densidade marginal de  $X_{(r)}$  é dada por

$$f_r(x_{(r)}) = \frac{n!}{(r-1)!(n-r)!} [F(x_{(r)})]^{r-1} [1 - F(x_{(r)})]^{n-r} f(x_{(r)}), \quad (2.6)$$

onde  $F$  é a função de distribuição comum de  $X_1, \dots, X_n$ .

**Demonstração:** Partimos da expressão da função de densidade conjunta das estatísticas de ordem obtida no Teorema 2.2. Então,

$$\begin{aligned} f_r(x_{(r)}) &= n! f(x_{(r)}) \int_{-\infty}^{x_{(r)}} \int_{-\infty}^{x_{(r-1)}} \cdots \int_{-\infty}^{x_{(2)}} \int_{x_{(r)}}^{+\infty} \int_{x_{(r+1)}}^{+\infty} \cdots \int_{x_{(n-1)}}^{+\infty} \prod_{i \neq r}^n f(t_i) dt_n \cdots dt_{r+1} dt_1 \cdots dt_{r-1} \\ &= n! f(x_{(r)}) \frac{[1 - F(x_{(r)})]^{n-r}}{(n-r)!} \int_{-\infty}^{x_{(r)}} \cdots \int_{-\infty}^{x_{(2)}} \prod_{i=1}^{r-1} f(t_i) dt_i \\ &= n! f(x_{(r)}) \frac{[1 - F(x_{(r)})]^{n-r}}{(n-r)!} \frac{[F(x_{(r)})]^{r-1}}{(r-1)!}. \quad \blacksquare \end{aligned}$$

Como utilidade deste teorema podemos mencionar o fato de agora podermos encontrar os momentos das estatística de ordem. Faremos isso como consequência do seguinte exemplo.

**Exemplo 2.7**

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes  $U(0, 1)$ . Então

$$f_r(x_{(r)}) = \begin{cases} \frac{n!}{(r-1)!(n-r)!} x_{(r)}^{r-1} (1 - x_{(r)})^{n-r}, & \text{se } 0 < x_{(r)} < 1 \\ 0, & \text{caso contrário} \end{cases} \quad (1 \leq r \leq n).$$

Observemos que, na situação do exemplo acima,

$$X_{(r)} \sim \text{Beta}(r, n - r + 1),$$

logo, valem os resultados da distribuição Beta e, por exemplo,

$$E(X_{(r)}) = r/(n + 1).$$

Para uma densidade qualquer e somente quatro variáveis aleatórias a forma da densidade marginal, de uma qualquer estatística de ordem, é mostrada no seguinte exemplo.

**Exemplo 2.8**

Sejam  $X_1, X_2, X_3, X_4$  variáveis aleatórias independentes com densidade comum  $f$ . A função de densidade conjunta das estatísticas de ordem  $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$  é

$$f(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}) = \begin{cases} 4! f(x_{(1)}) f(x_{(2)}) f(x_{(3)}) f(x_{(4)}), & \text{se } x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)} \\ 0, & \text{caso contrário} \end{cases}.$$

Vamos calcular a função de densidade marginal de  $X_{(2)}$ . Temos que, se  $x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)}$

$$\begin{aligned}
 f_2(x_{(2)}) &= 4! \int \int \int \int f(t_1)f(x_{(2)})f(t_3)f(t_4) dt_1 dt_3 dt_4 \\
 &= 4!f(x_{(2)}) \int_{-\infty}^{x_{(2)}} \int_{x_{(2)}}^{+\infty} \left[ \int_{t_3}^{+\infty} f(t_4) dt_4 \right] f(t_3)f(t_1) dt_3 dt_1 \\
 &= 4!f(x_{(2)}) \int_{-\infty}^{x_{(2)}} \left\{ \int_{x_{(2)}}^{+\infty} [1 - F(t_3)]f(t_3) dt_3 \right\} f(t_1) dt_1 \\
 &= 4!f(x_{(2)}) \int_{-\infty}^{x_{(2)}} \frac{[1 - F(x_{(2)})]^2}{2} f(t_1) dt_1 = 4!f(x_{(2)}) \frac{[1 - F(x_{(2)})]^2}{2!} F(x_{(2)}).
 \end{aligned}$$

Evidentemente, a expressão acima coincide com o resultado apresentado no Teorema 2.3.

#### Teorema 2.4

Sejam  $X_1, \dots, X_n$  variáveis aleatórias contínuas independentes e igualmente distribuídas e  $(X_{(1)}, \dots, X_{(n)})$  as estatísticas de ordem. A função de densidade conjunta de  $X_{(r)}$  e  $X_{(s)}$  é dada por

$$f_{rs}(x_{(r)}, x_{(s)}) = n! \frac{F^{r-1}(x_{(r)})[F(x_{(s)}) - F(x_{(r)})]^{s-r-1}[1 - F(x_{(r)})]^{n-s} f(x_{(r)})f(x_{(s)})}{(r-1)!(n-r)!}, \quad (2.7)$$

caso  $x_{(r)} < x_{(s)}$  e zero noutras situações.

#### Demonstração:

$$\begin{aligned}
 f_{rs}(x_{(r)}, x_{(s)}) &= \int_{-\infty}^{x_{(r)}} \dots \int_{-\infty}^{t_2} \int_{x_{(r)}}^{x_{(s)}} \dots \int_{t_{s-2}}^{x_{(s)}} \int_{x_{(s)}}^{+\infty} \dots \int_{t_{n-1}}^{+\infty} \\
 &\quad n! f(t_1) \dots f(t_n) dt_n \dots dt_{s+1} dt_{s-1} \dots dt_{r+1} dt_1 \dots dt_{r-1} \\
 &= n! \int_{-\infty}^{x_{(r)}} \dots \int_{-\infty}^{t_2} \int_{x_{(r)}}^{x_{(s)}} \dots \int_{t_{s-2}}^{x_{(s)}} \frac{[1 - F(x_{(s)})]^{n-s}}{(n-s)!} \times \\
 &\quad \times f(t_1)f(t_2) \dots f(x_{(s)}) dt_{s-1} \dots dt_{r+1} dt_1 \dots dt_{r-1} \\
 &= n! \frac{[1 - F(x_{(s)})]^{n-s}}{(n-s)!} f(x_{(s)}) \int_{-\infty}^{x_{(r)}} \dots \int_{-\infty}^{t_2} f(t_1) \dots f(x_{(r)}) \times \\
 &\quad \times \frac{[F(x_{(s)}) - F(x_{(r)})]^{s-r-1}}{(s-r-1)!} dt_1 \dots dt_{r-1} \\
 &= \frac{n!}{(n-s)!(s-r-1)!} [1 - F(x_{(s)})]^{n-s} \times \\
 &\quad \times [F(x_{(s)}) - F(x_{(r)})]^{s-r-1} f(x_{(s)})f(x_{(r)}) \frac{[F(x_{(r)})]^{r-1}}{(r-1)!},
 \end{aligned}$$

caso  $x_{(r)} < x_{(s)}$ . ■

De modo semelhante, podemos mostrar que a função de densidade conjunta de  $X_{(k_1)}, \dots, X_{(k_m)}$  se  $1 \leq k_1 <$

$k_2 < \dots < k_m \leq n$ ,  $1 \leq m \leq n$ , é dada por

$$f_{k_1 k_2 \dots k_m}(x_{(k_1)}, x_{(k_2)}, \dots, x_{(k_m)}) = \frac{n!}{(k_1 - 1)!(k_2 - k_1 - 1)! \times \dots \times (n - k_m)!} \times \\ \times F^{k_1 - 1}(x_{(k_1)}) f(x_{(k_1)}) [F(x_{(k_2)}) - F(x_{(k_1)})]^{k_2 - k_1 - 1} f(x_{(k_2)}) \times \dots \times \\ \times [F(x_{(k_{m-1})}) - F(x_{(k_{m-2})})]^{k_{m-1} - k_{m-2} - 1} f(x_{(k_{m-1})}) [1 - F(x_{(k_m)})]^{n - k_m} f(x_{(k_m)}),$$

caso  $x_{(k_1)} < x_{(k_2)} < \dots < x_{(k_m)}$  e zero noutras situações.

**Exemplo 2.9 (Continuação do Exemplo 2.7)**

Sabemos que as variáveis aleatórias  $X_1, X_2, \dots, X_n$  são independentes e tem como função de densidade comum

$$f(x) = \begin{cases} 1, & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Então, a função de densidade conjunta de  $X_{(r)}$  e  $X_{(s)}$  é dada por

$$f_{rs}(x_{(r)}, x_{(s)}) = \begin{cases} n! \frac{x_{(r)}^{r-1} (x_{(r)} - x_{(s)})^{s-r-1} (1 - x_{(s)})^{n-s}}{(r-1)!(s-r-1)!(n-s)!}, & \text{se } x_{(r)} < x_{(s)} \\ 0, & \text{caso contrário} \end{cases}.$$

onde  $1 \leq r < s \leq n$ .

Uma situação mais complexa é trabalharmos com funções de estatísticas de ordem. Não temos um resultado simples para o caso de qualquer funções destas estatísticas. Mas, no exemplo a seguir, podemos encontrar um resultado interessante para o comportamento da diferença de estatísticas de ordem.

**Exemplo 2.10**

Sejam  $X_{(1)}, X_{(2)}, X_{(3)}$  as estatísticas de ordem das variáveis aleatórias independentes e igualmente distribuídas  $X_1, X_2, X_3$  com função de densidade comum

$$f(x) = \begin{cases} \beta e^{-x\beta}, & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases}.$$

sendo  $\beta > 0$ . Sejam  $Y_1 = X_{(3)} - X_{(2)}$  e  $Y_2 = X_{(2)}$ . Mostraremos que  $Y_1$  e  $Y_2$  são independentes. Para isso primeiro observemos que a função de densidade conjunta de  $X_{(2)}$  e  $X_{(3)}$  é dada por

$$f_{23}(x, y) = \begin{cases} \frac{3!}{1!0!0!} (1 - e^{-x\beta}) \beta e^{-x\beta} e^{-y\beta}, & \text{se } x \geq y \\ 0, & \text{caso contrário} \end{cases}.$$

A função de densidade conjunta de  $(Y_1, Y_2)$  é então

$$f(y_1, y_2) = 3! \beta^2 (1 - e^{-y_2\beta}) e^{-y_2\beta} e^{-(y_1+y_2)\beta} = \\ = \begin{cases} [3! \beta e^{-2y_2\beta} (1 - e^{-y_2\beta})] [\beta e^{-y_1\beta}], & \text{se } 0 < y_1 < +\infty, 0 < y_2 < +\infty \\ 0, & \text{caso contrário} \end{cases}.$$

Do qual segue que  $Y_1$  e  $Y_2$  são independentes.

Duas estatísticas de ordem importantes são o máximo e mínimo. Nesses casos é possível encontrar, de maneira

analítica, expressões para a função de distribuição. Vejamos no teorema a seguir as expressões da função de distribuição das estatísticas de ordem  $X_{(1)}$  e  $X_{(n)}$ .

### Teorema 2.5

Seja  $X_1, \dots, X_n$  uma amostra aleatória da função de distribuição  $F(\cdot)$  e  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  as estatísticas de ordem correspondentes. A função de distribuição do mínimo assume a forma

$$F(x_{(1)}) = F^n(x_{(1)})$$

e a função de distribuição do máximo é

$$F(x_{(n)}) = 1 - [1 - F(x_{(n)})]^n.$$

**Demonstração:** Exercício. ■

Acerca da função de distribuição de qualquer estatística de ordem temos o resultado a seguir.

### Teorema 2.6

Seja  $X_1, \dots, X_n$  uma amostra aleatória da função de distribuição  $F(\cdot)$  e  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  as estatísticas de ordem correspondentes. A função de distribuição da  $k$ -ésima estatística de ordem  $X_{(k)}$  assume a forma

$$P(X_{(k)} \leq x) = \sum_{i=k}^n \frac{n!}{(n-i)!i!} F^i(x) [1 - F(x)]^{n-i}.$$

**Demonstração:** O evento  $\{X_{(k)} \leq x\}$  ocorre se, e somente se, pelo menos  $k$  dos  $X_1, X_2, \dots, X_n$  são menores ou iguais a  $x$ , por isso o somatório começa em  $k$ . ■

Nos dois teoremas seguintes relacionamos a distribuição condicional de estatísticas de ordem, condicionadas em outra estatística de ordem, com a distribuição de estatísticas de ordem de uma população cuja distribuição é uma forma truncada da função de distribuição da população original  $F$ .

### Teorema 2.7

Seja  $X_1, \dots, X_n$  uma amostra aleatória da função de distribuição absolutamente contínua  $F$  e  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  as estatísticas de ordem correspondentes. A distribuição condicional de  $X_{(j)} | X_{(i)} = x_i$ , para  $i < j$ , coincide com a distribuição da  $(j-i)$ -ésima estatística de ordem obtida de uma amostra de tamanho  $n-i$  de uma população cuja função de distribuição é  $F$  truncada à esquerda em  $x_i$ .

**Demonstração:** A densidade condicional de  $X_{(j)}$  dado que  $X_{(i)} = x_i$  calcula-se dividindo a densidade conjunta de  $X_{(i)}$  e  $X_{(j)}$ , dada em (2.7), pela densidade marginal de  $X_{(i)}$ , esta obtida no Teorema 2.4. Temos então que, quando

$i < j \leq n$  e  $x_i \leq x_j < \infty$ ,

$$\begin{aligned} f(x_j|X_{(i)} = x_i) &= \frac{f_{ij}(x_i, x_j)}{f_i(x_i)} \\ &= \frac{(n-i)!}{(j-i-1)!(n-j)!} \left[ \frac{F(x_j) - F(x_i)}{1 - F(x_i)} \right]^{j-i-1} \times \left[ \frac{1 - F(x_j)}{1 - F(x_i)} \right]^{n-j} \frac{f(x_j)}{1 - F(x_i)}. \end{aligned}$$

O resultado segue observando que  $\frac{F(x_j) - F(x_i)}{1 - F(x_i)}$  e  $\frac{f(x_j)}{1 - F(x_i)}$  são, respectivamente, as funções de distribuição e de densidade truncando à esquerda em  $x_i$  a distribuição  $F$ . ■

Na demonstração do teorema anterior utiliza-se o conceito de distribuição truncada, o que é isso? define-se a seguir este conceito e incluem-se exemplos explicativos.

### Definição 2.5 (*Distribuição truncada*)

Seja  $X$  uma variável aleatória no espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  e seja  $A \in \mathcal{B}(\mathbb{R})$ , a  $\sigma$ -álgebra de Borel, tal que  $0 < P(X \in A) < 1$ . A distribuição condicional  $P(X \leq x|X \in A)$ , definida para todo  $x \in \mathbb{R}$ , é chamada de distribuição truncada de  $X$ .

Caso a variável aleatória  $X$  seja discreta com função de probabilidade  $P$ , a distribuição truncada de  $X$  é dada por

$$P(X = x|X \in A) = \frac{P(X = x, X \in A)}{P(X \in A)} = \begin{cases} \frac{P(X = x)}{\sum_{a \in A} P(X = a)}, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}.$$

Na situação  $X$  do tipo contínua, com função de densidade  $f$ , temos que

$$P(X \leq x|X \in A) = \frac{P(X \leq x, X \in A)}{P(X \in A)} = \frac{\int_{(-\infty, x] \cap A} f(y) dy}{\int_A f(y) dy}. \quad (2.8)$$

Concluindo então que, a função de densidade da distribuição truncada é dada por

$$h(x) = \begin{cases} \frac{f(x)}{\int_A f(y) dy}, & \text{caso } x \in A, \\ 0, & \text{caso } x \notin A \end{cases}. \quad (2.9)$$

### Exemplo 2.11

Suponhamos  $X$  uma variável aleatória com distribuição normal padrão e  $A = (-\infty, 0]$ . Então,  $P(X \in A) = 1/2$ , dado que  $X$  é simétrica e contínua. Para a densidade truncada temos que

$$h(x) = \begin{cases} 2f(x), & \text{caso } -\infty < x \leq 0, \\ 0, & \text{caso } x > 0 \end{cases}.$$

O truncamento é especialmente importante nos casos em que a distribuição  $F$  em questão não tem média finita. Se  $X$  é uma variável aleatória, truncamos  $X$  em algum  $c > 0$ , onde  $c$  é finito, substituindo  $X$  por  $X^c = X$  caso  $|X| \leq c$  e zero caso  $|X| > c$ . Então  $X^c$  é  $X$  truncada em  $c$  e todos os momentos de  $X^c$  existem e são finitos. Na verdade, sempre podemos selecionar  $c$  suficientemente grande para que

$$P(X \neq X^c) = P(|X| > c),$$

seja arbitrariamente pequena.

A distribuição de  $X^c$  é então dada por

$$P(X^c \leq x) = P(X \leq x | |X| \leq c) = \frac{\int_{(-\infty, x] \cap [-c, +c]} f(y) dy}{P(|X| \leq c)},$$

no caso contínuo com função de densidade  $f$  e é dada por

$$P(X^c = x) = \begin{cases} \frac{P(X = x)}{\sum_{a \in [-c, +c]} P(X = a)}, & \text{se } x \in [-c, +c] \\ 0, & \text{se } x \notin [-c, +c] \end{cases},$$

no caso discreto. Observemos que, para algum  $\alpha > 0$ ,

$$E(|X^c|)^\alpha \leq c^\alpha.$$

### Exemplo 2.12

Caso  $X \sim \text{Cauchy}(0, 1)$ , sabemos que  $E(X)$  não existe. Seja  $c > 0$  um número finito, truncando  $X$  em  $c$  definimos

$$X^c = \begin{cases} X, & \text{caso } |X| \leq c, \\ 0, & \text{caso } |X| > c \end{cases}.$$

Então

$$P(|X| \leq c) = \frac{1}{\pi} \int_{-c}^{+c} \frac{1}{1+x^2} dx = \frac{2}{\pi} \tan^{-1}(c).$$

Sendo que a função de densidade truncada é dada por

$$h(x) = \begin{cases} \frac{1}{2} \frac{1}{1+x^2} \frac{1}{\tan^{-1}(c)}, & \text{caso } x \in [-c, +c], \\ 0, & \text{caso } x \notin [-c, +c] \end{cases}.$$

Desta expressão obtemos que

$$E(X^c) = \frac{1}{2 \tan^{-1}(c)} \int_{-c}^{+c} \frac{x}{1+x^2} dx = 0,$$

e também que

$$E(X^c)^2 = \frac{1}{2 \tan^{-1}(c)} \int_{-c}^{+c} \frac{x^2}{1+x^2} dx = \frac{c}{\tan^{-1}(c)} - 1.$$

Por último, temos o seguinte resultado estabelecendo novamente relação entre estatísticas de ordem e distribuições truncadas.

**Teorema 2.8**

Seja  $X_1, \dots, X_n$  uma amostra aleatória da função de distribuição absolutamente contínua  $F$  e  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  as estatísticas de ordem correspondentes. A distribuição condicional de  $X_{(i)}|X_{(j)} = x_j$ , para  $i < j$ , coincide com a distribuição da  $i$ -ésima estatística de ordem obtida de uma amostra de tamanho  $j - 1$  de uma população cuja função de distribuição é  $F$  truncada à direita em  $x_j$ .

**Demonstração:** A densidade condicional de  $X_{(i)}$  dado que  $X_{(j)} = x_j$  calcula-se dividindo a densidade conjunta de  $X_{(i)}$  e  $X_{(j)}$ , dada em (2.7), pela densidade marginal de  $X_{(j)}$ , esta obtida no Teorema 2.4. Temos então que, quando  $i < j \leq n$  e  $x_i \leq x_j < \infty$ ,

$$f(x_i|X_{(j)} = x_j) = \frac{(j-i)!}{(i-1)!(j-i-1)!} \left[ \frac{F(x_i)}{F(x_j)} \right]^{i-1} \times \left[ \frac{F(x_j) - F(x_i)}{F(x_j)} \right]^{j-i-1} \frac{f(x_i)}{F(x_j)}.$$

O resultado segue observando que  $F(x_i)/F(x_j)$  e  $f(x_i)/F(x_j)$  são, respectivamente, as funções de distribuição e de densidade truncando à direita em  $x_j$  a distribuição  $F$ . ■

**2.2.2 Quantis**

Lembremos que a função de distribuição  $F$  é contínua à direita e que o número de descontinuidades é, no máximo, enumerável. Estas são propriedades importantes que farão toda diferença na definição dos quantis amostrais, por isso, demonstraremos as propriedades mencionadas da função de distribuição.

A prova de que  $F$  é contínua à direita advém do seguinte fato

$$F(x + h_n) - F(x) = P(x < X \leq x + h_n),$$

onde  $\{h_n\}$  é uma sequência de números reais estritamente positivos tais que  $\lim_{n \rightarrow \infty} h_n = 0$ . Segue, da propriedade de continuidade da função de probabilidade,<sup>1</sup> que

$$\lim_{n \rightarrow \infty} [F(x + h_n) - F(x)] = 0,$$

e, portanto,  $F$  é contínua à direita.

Definamos por  $D$  o conjunto dos pontos de descontinuidade de  $F$  e seja

$$D_n = \left\{ x \in D : P(X = x) \geq \frac{1}{n} \right\},$$

onde  $n$  é um inteiro positivo. Dado que  $F(\infty) - F(-\infty) = 1$ , o número de elementos em  $D_n$  não pode exceder  $n$ . Logicamente

$$D = \bigcup_{n=1}^{\infty} D_n$$

e, então, o conjunto  $D$  é enumerável. Demonstrando-se assim a segunda propriedade importante mencionada da função de distribuição. Definimos a seguir o conceito de quantil teórico e depois mostramos a forma de cálculo.

<sup>1</sup>A função de distribuição é contínua, devido a que

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n),$$

se o limite  $\lim_{n \rightarrow \infty} A_n$  existir.

**Definição 2.6 (Quantil)**

Seja  $X$  uma variável aleatória com distribuição  $F$ . Para  $0 < p < 1$  definimos o  $p$ -ésimo quantil de  $F$  como

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}.$$

A função  $F^{-1}(t)$ ,  $0 < t < 1$  foi definida em (1.29) e é chamada de função inversa de  $F$ . O seguinte teorema fornece-nos propriedades úteis. Fica claro que as propriedades apresentadas no seguinte teorema nos permitirão o cálculo dos quantis e é por isso que dedicamos atenção a este conceito.

**Teorema 2.9**

Seja  $F$  uma função de distribuição. A função  $F^{-1}(t)$ ,  $0 < t < 1$ , é não-decrescente, contínua a esquerda e satisfaz:

- i)  $F^{-1}[F(x)] \leq x$ ,  $-\infty < x < \infty$ ,
- ii)  $F[F^{-1}(t)] \geq t$ ,  $0 < t < 1$ ,
- iii)  $F(x) \geq t$  se, e somente se,  $x \geq F^{-1}(t)$ .

**Demonstração:** Exercício. ■

**Exemplo 2.13**

Seja  $X \sim \text{Exponencial}(\theta)$ . Sabemos que a função de distribuição neste caso é  $F(x) = 1 - e^{-x/\theta}$ . Resulta que a expressão de qualquer um dos quantis é possível de ser encontrada de maneira exata via

$$\begin{aligned} F(\xi_p) &= p \\ 1 - e^{-\xi_p/\theta} &= p \\ 1 - p &= e^{-\xi_p/\theta}, \end{aligned}$$

obtendo-se que

$$\xi_p = -\theta \ln(1 - p)$$

é a expressão teórica do  $p$ -ésimo quantil. Devemos mencionar que a expressão dos quantis está bem definida, no sentido de que o resultado é sempre positivo. Isto é importante porque devemos lembrar que a distribuição exponencial está definida somente para valores positivos, então o quantil teórico deve ser positivo, já que é um dos possíveis valores da variável.

Observemos que caso  $F$  seja contínua e estritamente crescente,  $F^{-1}$  é definida como

$$F^{-1}(y) = x \quad \text{quando} \quad y = F(x).$$

Ainda podemos observar que, se  $x_0$  é um ponto de descontinuidade de  $F$  e supondo que

$$F(x_0^-) < y < F(x_0) = F(x_0^+)$$

vemos que, embora não exista  $x$  tal que  $y = F(x)$ ,  $F^{-1}(y)$  é definido como igual a  $x_0$ . A situação na qual  $F$  não é estritamente crescente, por exemplo, caso da variável aleatória ser discreta, podemos escrever

$$F(x) = \begin{cases} < y, & \text{caso } x < a \\ = y, & \text{caso } a \leq x \leq b \\ > y, & \text{caso } x > b \end{cases} .$$

Então, qualquer valor  $a \leq x \leq b$  poderia ser escolhido como  $x = F^{-1}(y)$ . A convenção é que, neste caso, definimos  $F^{-1}(y) = a$ . Em particular

$$\xi_{1/2} = F^{-1}(1/2), \quad (2.10)$$

é chamada de mediana de  $F$ . Observemos que  $\xi_p$  satisfaz a desigualdade

$$F(\xi_{p-}) \leq p \leq F(\xi_p).$$

### Exemplo 2.14 (*Continuação do Exemplo 2.13*)

Caso  $p = 1/2$ , a mediana amostral será  $\xi_{1/2} = -\theta \ln(1/2) = 0.6931472\theta$ .

## 2.3 Momentos amostrais

Nesta seção vamos estudar algumas estatísticas amostrais comumente utilizadas e suas distribuições.

### Definição 2.7

Define-se a função  $\hat{F}_n(x)$ , chamada de função de distribuição amostral ou empírica, como

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - X_i),$$

onde  $\delta$  é a conhecida função delta, definida em (1.2).

Observemos que  $n\hat{F}_n(x)$  é o número de  $X_k$  ( $1 \leq k \leq n$ ) menores ou iguais a  $x$ . Se  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  são as estatísticas de ordem de  $X_1, X_2, \dots, X_n$  então claramente

$$\hat{F}_n(x) = \begin{cases} 0, & \text{se } x < X_{(1)} \\ \frac{k}{n}, & \text{se } X_{(k)} \leq x < X_{(k+1)}, \quad (k = 1, 2, \dots, n-1) \\ 1, & \text{se } x \geq X_{(n)} \end{cases} . \quad (2.11)$$

Para  $x \in \mathbb{R}$  fixo, mas arbitrário,  $\hat{F}_n(x)$  é uma variável aleatória discreta e, como toda variável aleatória discreta, têm função de probabilidade, média e variância. Temos então o seguinte resultado.

### Teorema 2.10

A variável aleatória  $\hat{F}_n(x)$  tem por função de probabilidade

$$P \left[ \hat{F}_n(x) = \frac{i}{n} \right] = \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i}, \quad i = 0, 1, \dots, n, \quad (2.12)$$

com esperança

$$E[\widehat{F}_n(x)] = F(x) \quad (2.13)$$

e variância

$$\text{Var}[\widehat{F}_n(x)] = \frac{F(x)[1 - F(x)]}{n}. \quad (2.14)$$

**Demonstração:** Dado que  $\delta(x - X_i)$ ,  $i = 1, 2, \dots, n$  são variáveis aleatórias independentes igualmente distribuídas cada uma com função de probabilidade

$$P[\delta(x - X_i) = 1] = P(x - X_i \geq 0) = F(x)$$

e

$$P[\delta(x - X_i) = 0] = 1 - F(x),$$

sua soma  $nF_n^*(x)$  é uma variável aleatória com distribuição *Binomial*( $n, p$ ), onde  $p = F(x)$ . As relações (2.12), (2.13) e (2.14) seguem-se imediatamente. ■

### Corolário 2.11

A variável aleatória  $\widehat{F}_n(x)$  converge em probabilidade a  $F(x)$ , isto é

$$\widehat{F}_n(x) \xrightarrow{P} F(x) \quad \text{quando } n \rightarrow \infty.$$

**Demonstração:** É uma consequência da Lei dos Grandes Números. ■

### Corolário 2.12

$$\frac{\sqrt{n}[\widehat{F}_n(x) - F(x)]}{\sqrt{F(x)[1 - F(x)]}} \xrightarrow{D} Z \quad \text{quando } n \rightarrow \infty,$$

onde  $Z \sim N(0, 1)$ .

**Demonstração:** É consequência do Teorema do Limite Central. ■

### Exemplo 2.15

Vamos apresentar o conceito de função distribuição empírica no caso de termos uma amostra aleatória da distribuição  $N(0, 1)$ . A lista de comandos na linguagem de programação R está disponível abaixo. O primeiro comando destina-se a fixar o gerador de amostras e, assim, em qualquer momento podemos obter a mesma amostra aleatória.

Na Figura 2.2 mostramos a forma da distribuição empírica, de três formas diferentes, para uma amostra de tamanho 12. A representação da função de distribuição empírica é realizada permitindo escolher qual utilizar segundo o agrado.

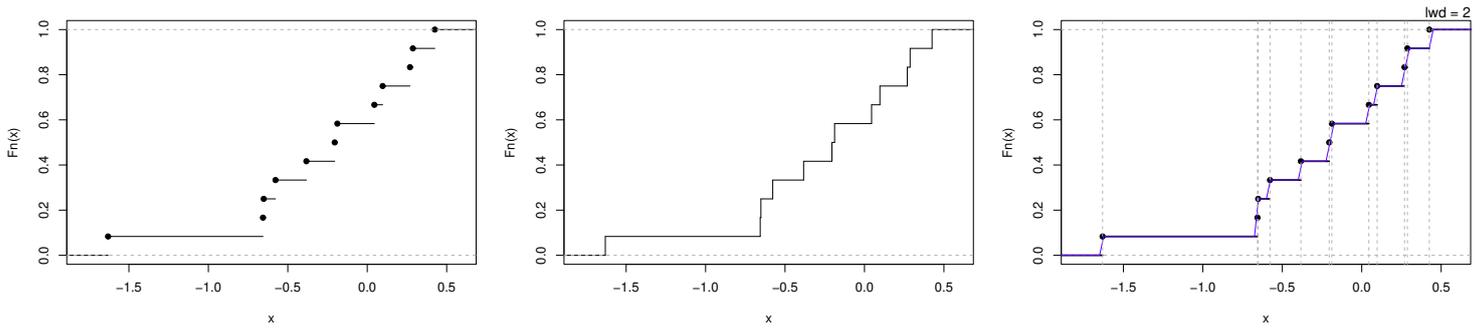


Figura 2.2: Representação da função de distribuição amostral ou empírica, de três formas diferentes, para uma amostra normal padrão de tamanho 12.

A linhas de comando a seguir permitiram-nos gerar os gráficos na Figura 2.2: construímos :

```
set.seed(5739); x=rnorm(12); Fn=ecdf(x)
par(mar=c(5,4,3,1), cex=0.9)
plot(Fn, main="")
plot(Fn, verticals = TRUE, do.points = FALSE, main="")
plot(Fn, lwd = 2, main=""); mtext("lwd = 2", adj = 1)
xx=unique(sort(c(seq(-3, 2, length = 201), knots(Fn12))))
lines(xx, Fn(xx), col = "blue")
abline(v = knots(Fn), lty = 2, col = "gray70")
```

Observemos que a convergência da distribuição empírica, segundo o Teorema 2.10, é para cada valor de  $x$ . É possível fazer uma demonstração da convergência em probabilidade simultaneamente para todos os  $x$ , ou seja, da convergência uniforme.

### Teorema 2.13 (Glivenko-Cantelli)

Seja  $X_1, \dots, X_n$  uma amostra aleatória com distribuição  $F$ . A variável aleatória  $\widehat{F}_n$  converge uniformemente a  $F$ , isto é, para todo  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{-\infty < x < +\infty} |\widehat{F}_n(x) - F(x)| > \epsilon \right\} = 0.$$

**Demonstração:** Seja  $\epsilon > 0$ . Escolhemos um inteiro  $k > 1/\epsilon$  e números

$$-\infty = x_0 < x_1 \leq x_2 \leq \dots \leq x_{k-1} < x_k = \infty,$$

tais que  $F(x_j^-) \leq j/k \leq F(x_j)$ , para  $j = 1, \dots, k-1$ . Observe que se  $x_{j-1} < x_j$ , então

$$F(x_j^-) - F(x_{j-1}) \leq \epsilon.$$

Pela Lei dos Grandes Números

$$\widehat{F}_n(x_j) \xrightarrow{q.c.} F(x_j)$$

e

$$\widehat{F}_n(x_j^-) \xrightarrow{q.c.} F(x_j^-),$$

para  $j = 1, \dots, k-1$ . Consequentemente,

$$\Delta_n = \max\{|\widehat{F}_n(x_j) - F(x_j)|, |F_n^*(x_j^-) - F(x_j^-)|, j = 1, \dots, k-1\} \xrightarrow{q.c.} 0.$$

Seja  $x$  arbitrário e encontremos  $j$  tal que  $x_{j-1} < x \leq x_j$ . Então,

$$\widehat{F}_n(x) - F(x) \leq \widehat{F}_n(x_j^-) - F(x_{j-1}) \leq \widehat{F}_n(x_j^-) - F(x_j^-) + \epsilon,$$

e

$$\widehat{F}_n(x) - F(x) \geq \widehat{F}_n(x_{j-1}) - F(x_j^-) \geq \widehat{F}_n(x_{j-1}) - F(x_{j-1}) - \epsilon.$$

Isto implica que

$$\sup_x |\widehat{F}_n(x) - F(x)| \leq \Delta_n + \epsilon \xrightarrow{q.c.} \epsilon.$$

Como isso vale para todo  $\epsilon > 0$ , o teorema segue. ■

Agora, dado que  $F_n^*(x)$  tem pontos de salto em  $X_i$ ,  $i = 1, 2, \dots, n$  é claro que existem todos os momentos de  $F_n^*(x)$ . Vamos considerar alguns valores típicos da função de distribuição  $F$ , chamados de estatísticas amostrais. Escrevamos

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad (2.15)$$

para os momentos de ordem  $k$  ao redor do 0 (zero). Aqui  $a_k$ , serão chamados de momentos amostrais de ordem  $k$ . Com esta notação

$$a_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

O momento amostral central é definido por

$$b_k = \frac{1}{n} \sum_{i=1}^n (X_i - a_1)^k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (2.16)$$

Logicamente,

$$b_1 = 0 \quad \text{e} \quad b_2 = \left( \frac{n-1}{n} \right) S_n^2.$$

Como mencionado anteriormente, não chamamos  $b_2$  a variância amostral.  $S_n^2$  será chamada como a variância amostral por razões que se tornarão claras posteriormente. Temos que

$$b_2 = a_2 - a_1^2.$$

Para a função geradora de momentos de  $\widehat{F}_n$  podemos afirmar que

$$M_{\widehat{F}_n}(t) = \frac{1}{n} \sum_{i=1}^n e^{tX_i}.$$

Definições similares são realizadas para momentos amostrais de distribuições multivariadas. Por exemplo, se  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  é uma amostra de uma distribuição bivariada, podemos escrever

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

para as duas médias amostrais e para os momentos de segunda ordem centrais escrevemos

$$b_{20} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad b_{02} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

$$b_{11} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Mais uma vez, escrevemos

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \quad (2.17)$$

para as duas variâncias amostrais e para a covariância amostral utilizamos

$$S_{11} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}). \quad (2.18)$$

Em particular, o coeficiente de correlação amostral é definido por

$$R = \frac{b_{11}}{\sqrt{b_{20}b_{02}}} = \frac{S_{11}}{S_1 S_2}.$$

Pode ser demonstrado que  $|R| \leq 1$  e que os valores extremos  $\pm 1$  ocorrem somente quando todos os pontos amostrais  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  estão alinhados.

Correspondendo a uma amostra  $X_1, X_2, \dots, X_n$  de observações em  $F$ ,  $p$ -ésimo quantil amostral é definido como o  $p$ -ésimo quantil da função de distribuição amostral, ou seja, como  $\hat{F}_n^{-1}$ .

Os quantis amostrais são definidos de maneira similar. Então, se  $0 < p < 1$ , o quantil amostral de ordem  $p$ , denotado como  $\hat{\xi}_p$ , é a estatística de ordem  $X_{(r)}$ , onde

$$r = \begin{cases} [np] & \text{se } n \text{ é um número par} \\ [np] + 1 & \text{se } n \text{ é um número ímpar} \end{cases}. \quad (2.19)$$

Como usual,  $[x]$  denota o maior inteiro  $\leq x$ . Observe que, se  $[n]$  for par, podemos escolher qualquer valor entre  $X_{([np])}$  e  $X_{([np]+1)}$  como o  $p$ -ésimo quantil amostral. Então, se  $p = \frac{1}{2}$  e  $n$  par podemos escolher qualquer valor entre  $X_{(n/2)}$  e  $X_{(n/2+1)}$ , os dois valores do meio, como a mediana amostral. Habitualmente é escolhido o ponto médio. Assim, a mediana amostral é definida como

$$\hat{\xi}_{1/2} = \begin{cases} X_{((n+1)/2)} & \text{se } n \text{ é ímpar} \\ \frac{X_{(n/2)} + X_{((n/2)+1)}}{2} & \text{se } n \text{ é par} \end{cases}. \quad (2.20)$$

Observe que

$$\left[ \frac{n}{2} + 1 \right] = \left( \frac{n+1}{2} \right)$$

se  $n$  é ímpar.

Consideraremos agora os momentos de características amostrais. Nos seguintes desenvolvimentos denotaremos  $E(X^k) = m_k$  e  $E[(X - \mu)^k] = \mu_k$  como os momentos populacionais e os momentos populacionais centrais de  $k$ -ésima ordem, respectivamente. Nas situações onde utilizamos  $m_k$  ou  $\mu_k$  assumiremos que estes existem. Também,  $\sigma^2$  representará a variância populacional.

#### Teorema 2.14

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de uma população com função de distribuição  $F$ . Então

$$E(\bar{X}) = \mu, \quad (2.21)$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad (2.22)$$

$$E(\bar{X}^3) = \frac{m_3 + 3(n-1)m_2\mu + (n-1)(n-2)\mu^3}{n^2} \quad (2.23)$$

$$E(\bar{X}^4) = \frac{1}{n^3} [m_4 + 4(n-1)m_3\mu + 6(n-1)(n-2)m_2\mu^2 + 3(n-1)m_2^2 + (n-1)(n-2)(n-3)\mu^4]. \quad (2.24)$$

**Demonstração:** Para provar (2.23) observemos que

$$\left(\sum_{i=1}^n X_i\right)^3 = \sum_{i=1}^n X_i^3 + 3 \sum_{\substack{i=1 \\ j \neq k}}^n \sum_{j=1}^n X_i^2 X_j + \sum_{\substack{i=1 \\ j \neq k \\ j \neq l}}^n \sum_{j=1}^n \sum_{k=1}^n X_i X_j X_k,$$

desta expressão obtemos o resultado em (2.23). Similarmente

$$\begin{aligned} \left(\sum_{i=1}^n X_i\right)^4 &= \left(\sum_{i=1}^n X_i\right) \left( \sum_{i=1}^n X_i^3 + 3 \sum_{\substack{i=1 \\ j \neq k}}^n \sum_{j=1}^n X_i^2 X_j + \sum_{\substack{i=1 \\ j \neq k \\ j \neq l}}^n \sum_{j=1}^n \sum_{k=1}^n X_i X_j X_k \right) \\ &= \sum_{i=1}^n X_i^4 + 4 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n X_i X_j^3 + 3 \sum_{i=1}^n \sum_{j=1}^n X_i^2 X_j^2 \\ &= +6 \sum_{\substack{i=1 \\ j \neq k \\ j \neq l}}^n \sum_{j=1}^n \sum_{k=1}^n X_i^2 X_j X_k + \sum_{\substack{i=1 \\ j \neq k, i \neq l \\ j \neq k, j \neq l \\ k \neq l}}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n X_i X_j X_k X_l. \quad \blacksquare \end{aligned}$$

Um detalhe importante é que os momentos centrais podem ser calculados a partir dos momentos, por exemplo,

$$\begin{aligned} \mu_2 &= E[(X - \mu)^2] = m_2 - \mu^2, \\ \mu_3 &= E[(X - \mu)^3] = m_3 - 3\mu m_2 + 2\mu^3 \end{aligned}$$

e assim por diante. Sabemos agora como calcular os momentos, até quarta ordem, de  $\bar{X}$ . Vejamos a seguir como calcular os momentos centrais.

### Teorema 2.15

Os momentos centrais de terceira e quarta ordem de  $\bar{X}$  são

$$\mu_3(\bar{X}) = \frac{\mu_3}{n^2}, \quad (2.25)$$

$$\mu_4(\bar{X}) = \frac{\mu_4}{n^3} + 3 \frac{(n-1)\mu_2^2}{n^3}. \quad (2.26)$$

**Demonstração:** Temos que

$$\mu_3(\bar{X}) = E(\bar{X} - \mu)^3 = \frac{1}{n^3} E\left\{\sum_{i=1}^n (X_i - \mu)^3\right\} = \frac{1}{n^3} \sum_{i=1}^n E(X_i - \mu)^3 = \frac{\mu_3}{n^2}.$$

No caso do quarto momento central

$$\mu_4(\bar{X}) = E(\bar{X} - \mu)^4 = \frac{1}{n^4} E\left\{\sum_{i=1}^n (X_i - \mu)^4\right\},$$

da qual obtemos que

$$\mu_4(\bar{X}) = \frac{1}{n^4} \sum_{i=1}^n E(X_i - \mu)^4 + \binom{4}{2} \frac{1}{n^4} + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n E\{(X_i - \mu)^2(X_j - \mu)^2\}.$$

Desenvolvendo adequadamente chegamos ao resultado em (2.26). ■

### Exemplo 2.16

Seja  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $\text{Gamma}(\alpha, \beta)$ . Sabemos da Seção 1.2 que

$$E(X) = \alpha\beta, \quad \text{Var}(X) = \alpha\beta^2$$

e que

$$m_k = \beta^k (\alpha + k - 1)(\alpha + k - 2) \cdots \alpha, \quad k \geq 1.$$

Então,

$$E(\bar{X}) = \alpha\beta, \quad \text{Var}(\bar{X}) = \frac{\alpha\beta^2}{n}$$

e

$$\mu_3(\bar{X}) = \frac{1}{n^2} \mu_3 = \frac{1}{n} (6\alpha^3\beta^3 + 3\alpha^2\beta^3 + 2\alpha\beta^3).$$

Até o momento estudamos como calcular os momentos da média amostral. Mais complexo é obter expressões para os momentos da variância amostral  $S_n^2$ . O teorema a seguir dedica-se ao objetivo de encontrarmos expressões, até segunda ordem, dos momentos amostrais centrais. Como consequência deste resultado obtemos os momentos da variância amostral.

### Teorema 2.16

Os momentos de  $b_2$  são

$$E(b_2) = \frac{(n-1)\sigma^2}{n}, \tag{2.27}$$

$$\text{Var}(b_2) = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}, \tag{2.28}$$

$$E(b_3) = \frac{(n-1)(n-2)}{n^2} \mu_3 \tag{2.29}$$

e

$$E(b_4) = \frac{(n-1)(n^2 - 3n + 3)}{n^3} \mu_4 + \frac{3(n-1)(2n-3)}{n^3} \mu_2^2. \tag{2.30}$$

**Demonstração:** Temos que

$$\begin{aligned} E(b_2) &= E \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n X_i \right)^2 \right] \\ &= m_2 - \frac{1}{n^2} E \left[ \sum_{i=1}^n X_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n X_i^2 X_j \right] \\ &= m_2 - \frac{1}{n^2} [nm_2 + n(n-1)\mu^2] = \left( \frac{n-1}{n} \right) (m_2 - \mu^2). \end{aligned}$$

Agora

$$n^2 b_2^2 = \left[ \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right]^2.$$

Escrevendo  $Y_i = X_i - \mu$ , vemos que  $E(Y_i) = 0$ ,  $\text{Var}(Y_i) = \sigma^2$  e  $E(Y^4) = \mu_4$ . Temos então que

$$\begin{aligned} n^2 E(b_2^2) &= E \left( \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right)^2 \\ &= E \left[ \sum_{i=1}^n Y_i^4 + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n Y_i^2 Y_j^2 - \frac{2}{n} \left( \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n Y_i^2 Y_j^2 + \sum_{i=1}^n Y_i^4 \right) \right. \\ &\quad \left. + \frac{1}{n} \left( 3 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n Y_i^2 Y_j^2 + \sum_{i=1}^n Y_i^4 \right) \right]. \end{aligned}$$

Segue então que

$$\begin{aligned} n^2 E(b_2^2) &= n\mu_4 + n(n-1)\sigma^2 - \frac{2}{n} [n(n-1)\sigma^4 + n\mu_4] + \frac{1}{n^2} [3n(n-1)\sigma^4 + n\mu_4] \\ &= \left( n - 2 + \frac{1}{n} \right) \mu_4 + \left( n - 2 + \frac{3}{n} \right) (n-1)\mu_2^2. \quad (\mu_2 = \sigma^2) \end{aligned}$$

Portanto

$$\begin{aligned} \text{Var}(b_2) &= E(b_2) - [E(b_2)]^2 \\ &= \left( n - 2 + \frac{1}{n} \right) \frac{\mu_4}{n^2} + (n-1) \left( n - 2 + \frac{3}{n} \right) \frac{\mu_2^2}{n^2} - \left( \frac{n-1}{n} \right)^2 \mu_2^2 \\ &= \left( n - 2 + \frac{1}{n} \right) \frac{\mu_4}{n^2} + (n-1)(3-n) \frac{\mu_2^2}{n^3}, \end{aligned}$$

como afirmado. As relações (2.29) e (2.30) podem ser provadas de forma semelhante. ■

### Corolário 2.17

$$E(S_n^2) = \sigma^2.$$

Este é justamente o motivo pelo qual chamamos  $S_n^2$  e não  $b_2$  de variância amostral.

**Corolário 2.18**

$$\text{Var}(S_n^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\mu_2^2.$$

**Exemplo 2.17 (Continuação do Exemplo 2.16)**

inteir Nesta situação,  $\sigma^2 = \alpha\beta^2$ ,  $\mu_2 = \sigma^2$  e  $\mu_4 = m_4 - 4m_3\mu + 6m_2\mu^2 - 3\mu^4$ . Obtemos que

$$E(S_n^2) = \alpha\beta^2$$

e

$$\text{Var}(S_n^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\alpha^2\beta^4.$$

O seguinte resultado fornece uma justificativa para a nossa definição de covariância amostral.

**Teorema 2.19**

Seja  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  uma amostra aleatória de uma distribuição bivariada, com variâncias respectivas  $\sigma_1^2$  e  $\sigma_2^2$  e covariância  $\rho\sigma_1\sigma_2$ . Então

$$E(S_1^2) = \sigma_1^2, \quad E(S_2^2) = \sigma_2^2 \quad \text{e} \quad E(S_{11}) = \rho\sigma_1\sigma_2. \quad (2.31)$$

**Demonstração:** Do Corolário 2.17 sabemos que  $E(S_1^2) = \sigma_1^2$  e  $E(S_2^2) = \sigma_2^2$ . Para provar que  $E(S_{11}) = \rho\sigma_1\sigma_2$  observemos que  $X_i$  é independente de  $X_j$ , ( $i \neq j$ ) e de  $Y_j$ , ( $i \neq j$ ). Temos que

$$(n-1)E(S_{11}) = E\left[\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})\right].$$

Agora

$$\begin{aligned} E\{(X_i - \bar{X})(Y_i - \bar{Y})\} &= \\ &= E\left(X_i Y_i - X_i \frac{\sum_{j=1}^n Y_j}{n} - Y_i \frac{\sum_{j=1}^n X_j}{n} + \frac{\sum_{j=1}^n X_j \sum_{j=1}^n Y_j}{n^2}\right) \\ &= E(XY) - \frac{1}{n}[E(XY) + (n-1)E(X)E(Y)] \\ &\quad - \frac{1}{n}[E(XY) + (n-1)E(X)E(Y)] \\ &\quad - \frac{1}{n^2}[nE(XY) + n(n-1)E(X)E(Y)] \\ &= \frac{n-1}{n}[E(XY) - E(X)E(Y)] \end{aligned}$$

e segue que

$$(n-1)E(S_{11}) = n\left(\frac{n-1}{n}\right)[E(XY) - E(X)E(Y)],$$

isto é

$$E(S_{11}) = E(XY) - E(X)E(Y) = \text{Cov}(X, Y) = \rho\sigma_1\sigma_2. \quad \blacksquare$$

A seguir, voltamos nossa atenção para as distribuições das características da amostra. Existem várias possibilidades. Se for necessária a distribuição exata o método de transformação de variáveis pode ser utilizado. As vezes, a técnica da função geradora de momentos pode ser aplicada. Assim, se  $X_1, X_2, \dots, X_n$  é uma amostra aleatória de uma população com distribuição para a qual existe a função geradora de momentos, a função geradora de momentos da média amostral  $\bar{X}$  é dada por

$$M_{\bar{X}}(t) = \prod_{i=1}^n E(e^{tX_i/n}) = [M_X(t/n)]^n, \quad (2.32)$$

onde  $M_X$  é a função geradora de momentos da distribuição populacional. Se  $M_{\bar{X}}(t)$  tiver alguma forma conhecida seria possível escrever a função de probabilidade ou de densidade de  $\bar{X}$ . Embora este método tem a desvantagem óbvia que se aplica apenas à distribuições para as quais existem todos os momentos, veremos sua efetividade na situação importante de amostras da distribuição normal.

### Exemplo 2.18

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição Gama( $\alpha, 1$ ). Nesta situação podemos encontrar a função de densidade de  $\bar{X}$ . Temos que

$$M_{\bar{X}}(t) = [M_X(t/n)]^n = \frac{1}{(1-t/n)^{\alpha n}}, \quad \frac{t}{n} < 1,$$

da qual obtemos que  $\bar{X} \sim \text{Gama}(n\alpha, 1/n)$ .

### Exemplo 2.19

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória da distribuição Uniforme no intervalo  $(0, 1)$ . Considere a média geométrica

$$Y_n = \left( \prod_{i=1}^n X_i \right)^{1/n}.$$

Sabemos que  $\log(Y_n) = (1/n) \sum_{i=1}^n \log(X_i)$  e, desta forma,  $\log(Y_n)$  é a média amostral de  $\log(X_1), \dots, \log(X_n)$ . A função de densidade comum de  $\log(X_1), \dots, \log(X_n)$  é

$$f(x) = \begin{cases} e^x, & \text{se } x < 0 \\ 0, & \text{caso contrário} \end{cases},$$

que é a distribuição exponencial negativa com parâmetro  $\beta = 1$ . Vemos que a função geradora de momentos de  $\log(Y_n)$  é dada por

$$M_{\log(Y_n)}(t) = \prod_{i=1}^n E(e^{t \log(X_i)/n}) = \frac{1}{(1-t/n)^n},$$

e a função de densidade de  $\log(Y_n)$  é dada por

$$f_{\log(Y_n)}(y) = \begin{cases} \frac{n^n}{\Gamma(n)} [-y]^{n-1} e^{ny}, & \text{se } -\infty < y < 0 \\ 0, & \text{caso contrário} \end{cases}.$$

Segue então que  $Y_n$  tem por função de densidade

$$f_{Y_n}(y) = \begin{cases} \frac{n^n}{\Gamma(n)} y^{n-1} [-\log(y)]^{n-1}, & \text{se } 0 < y < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Voltemos ao quantil amostral de ordem  $p$ ,  $\hat{\xi}_p$ , o qual sabemos é ou  $X_{([np])}$  ou  $X_{([np]+1)}$  dependendo se  $[np]$  é um número par ou ímpar, como definido em (2.19). Simplificando, vamos discutir as propriedades de  $X_{([np])}$ , onde  $p \in (0, 1)$  e  $n$  é grande. Isso, por sua vez, nos informará sobre as propriedades de  $\hat{\xi}_p$ .

Primeiro observemos que, se  $U_1, U_2, \dots, U_n$  é uma amostra aleatória da distribuição  $U(0, 1)$  então, pelo Teorema 2.3, temos que

$$U_{([np])} \sim \text{Beta}([np], n - [np] + 1),$$

do qual obtemos que

$$\begin{aligned} \mathbb{E}(U_{([np])}) &= \frac{[np]}{n+1} \xrightarrow{n \rightarrow \infty} p, \\ \text{Cov}(U_{([np_1])}, U_{([np_2])}) &= n \frac{[np_1](n - [np_2] + 1)}{(n+1)^2(n+2)} \xrightarrow{n \rightarrow \infty} p_1(1 - p_2). \end{aligned}$$

Utilizando este resultado e a desigualdade de Chebychev, demonstramos que

$$U_{([np])} \xrightarrow{P} p. \quad (2.33)$$

Isso gera a questão

$$\hat{\xi}_p \xrightarrow{P} \xi_p?$$

qualquer seja a distribuição da amostra aleatória  $X_1, \dots, X_n$ .

Para respondermos a pergunta acima vamos utilizar o Lema de Hoeffding, ou seja, para respondermos se o quantil amostral de ordem  $p$  converge em probabilidade para o quantil teórico correspondente, utilizaremos o seguinte resultado devido a Hoeffding (1963).

### Lema 2.20 (Hoeffding)

Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes satisfazendo que  $P(0 \leq X_i \leq 1) = 1$ , para cada  $i$ . Então, para  $t > 0$ ,

$$P\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \geq nt\right) \leq \exp(-2nt^2). \quad (2.34)$$

**Demonstração:** Dado que as variáveis aleatórias são limitadas ao intervalo  $(0, 1)$ , sabemos que

$$e^{hX} \leq (1 - X) + Xe^h,$$

isto deve-se a que a função exponencial  $e^{hX}$  é convexa e, portanto, seu gráfico é limitado por cima no intervalo  $0 \leq X \leq 1$  pela linha que conecta as ordenadas  $X = 0$  e  $X = 1$ . Então

$$\mathbb{E}(e^{hX}) \leq (1 - \mathbb{E}(X)) + \mathbb{E}(X)e^h. \quad (2.35)$$

Seja  $S_n = \sum_{i=1}^n X_i$ . Sabemos que

$$P(S_n - \mathbb{E}(S_n) \geq nt) = \mathbb{E}(\mathbf{1}_{[S_n - \mathbb{E}(S_n) - nt \geq 0]}),$$

também sabemos que

$$\mathbf{1}_{[S_n - \mathbb{E}(S_n) - nt \geq 0]} \leq \exp\left(h(S_n - \mathbb{E}(S_n) - nt)\right),$$

qualquer seja  $h$  uma constante positiva arbitrária. Então

$$P(S_n - \mathbb{E}(S_n) \geq nt) \leq \mathbb{E}\left(e^{h(S_n - \mathbb{E}(S_n) - nt)}\right) \quad (2.36)$$

e como estamos assumindo que as variáveis são independentes, podemos escrever

$$\mathbb{E}\left(e^{h(S_n - \mathbb{E}(S_n) - nt)}\right) = e^{-hnt} \prod_{i=1}^n \mathbb{E}\left(e^{h(X_i - \mathbb{E}(X_i))}\right). \quad (2.37)$$

Escrevendo  $\mu_i = \mathbb{E}(X_i)$  temos, pela expressão em (2.35) que

$$\mathbb{E}\left(e^{h(X_i - \mu_i)}\right) \leq e^{-h\mu_i} \left((1 - \mu_i) + \mu_i e^h\right) = e^{f(h)}, \quad (2.38)$$

onde  $f(h) = -h\mu_i + \ln(1 - \mu_i + \mu_i e^h)$ . As primeiras duas derivadas são:

$$f'(h) = -\mu_i + \frac{\mu_i}{e^{-h}(1 - \mu_i) + \mu_i} \quad \text{e} \quad f''(h) = \frac{\mu_i e^{-h}(1 - \mu_i)}{[\mu_i + e^{-h}(1 - \mu_i)]^2}.$$

Na segunda derivada, escolhendo  $u = \frac{\mu_i}{\mu_i + e^{-h}(1 - \mu_i)}$  vemos que este quociente é da forma  $u(1 - u)$ , sendo  $0 < u < 1$ . Portanto,  $f''(h) \leq \frac{1}{4}$ . Pela série de Taylor

$$f(h) \leq f(0) + f'(0)h + \frac{1}{8}h^2 = \frac{1}{8}h^2.$$

Então, pela expressão em (2.38)

$$\mathbb{E}\left(e^{h(X_i - \mu_i)}\right) \leq e^{\frac{1}{8}h^2}.$$

Substituindo em (2.36) temos que

$$P(S_n - \mathbb{E}(S_n) \geq nt) \leq e^{-nht + \frac{1}{8}nh^2},$$

e o mínimo no expoente é atingido quando  $h = 4t$ . Então, o mínimo do limite superior da probabilidade é  $\exp(-2nt^2)$ . ■

Devemos lembrar que esta não é a única maneira de termos uma taxa de convergência para Teorema do Limite Central. Por exemplo, se  $Y_1, Y_2, \dots, Y_n$  forem variáveis aleatórias independentes e identicamente distribuídas, utilizando o Teorema de Berry-Esseen<sup>2</sup>, temos que

$$P\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \geq nt\right) \leq \Phi\left(t\sqrt{\frac{\text{Var}(Y_1)}{n}}\right) + \frac{C}{\sqrt{n}} \frac{\mathbb{E}|Y_1 - \mathbb{E}(Y_1)|^3}{\text{Var}^{3/2}(Y_1)}.$$

2

### Teorema 2.21 (Teorema de Berry-Esseen)

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com terceiro momento finito, tais que  $\mathbb{E}(X_i) = 0$  e  $\mathbb{E}(X_i^2) = 1$ ,  $\forall i$ . Então

$$\sup_{\substack{\alpha, \beta \in \mathbb{R} \\ \alpha < \beta}} \left| P\left(\alpha \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \leq \beta\right) - \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2} dt \right| \leq \frac{C\gamma^3}{\sqrt{n}},$$

onde  $\gamma = \left(\sum_{i=1}^n \mathbb{E}|X_i|^3/n\right)^{1/3} \leq \max_{i=1, \dots, n} (\mathbb{E}|X_i|^3)^{1/3}$  e  $C > 0$  uma constante universal.

**Demonstração:** Berry (1941); Esseen (1942). ■

Pode-se consultar o livro de Feller (1971) para uma demonstração moderna.

**Exemplo 2.20**

Caso a amostra aleatória seja *Bernoulli*( $\mu$ ), temos que

$$\sum_{i=1}^n X_k \sim \text{Binomial}(n, \mu).$$

Então, segundo a desigualdade de Hoeffding

$$P(\bar{X} - \mu \geq t) \leq \exp(-2nt^2).$$

Uma vantagem da desigualdade no Lema de Hoeffding é que não assume-se conhecimento da variância e, em geral, o limite da probabilidade é mais acurado do que outras desigualdades. Caso as variáveis aleatórias sejam limitadas como  $a \leq X_i \leq b$ , com  $a < b$ , o limite superior da desigualdade (2.34) seria  $\exp(-2nt^2/(b-a)^2)$ .

**Exemplo 2.21**

Sejam  $X_1, \dots, X_n$  variáveis aleatórias com distribuição  $U(-1, 1)$ . Nesta situação  $E(X) = 0$ ,  $a = -1$  e  $b = 1$ . A desigualdade de Hoeffding assume a forma

$$P(\bar{X} \geq t) \leq \exp(-nt^2/2).$$

**Teorema 2.22**

Suponhamos que  $\xi_p$  seja solução única da desigualdade

$$F(\xi_p^-) \leq p \leq F(\xi_p), \quad 0 < p < 1.$$

Então, para todo  $\epsilon > 0$ ,

$$P(|\hat{\xi}_p - \xi_p| > \epsilon) \leq 2 \exp(-2n\delta_\epsilon^2), \quad \forall n,$$

onde  $\delta_\epsilon = \min\{F(\xi_p + \epsilon) - p, p - F(\xi_p - \epsilon)\}$ .

**Demonstração:** Para  $\epsilon > 0$  qualquer, podemos escrever

$$P(|\hat{\xi}_p - \xi_p| > \epsilon) = P(\hat{\xi}_p > \xi_p + \epsilon) + P(\hat{\xi}_p < \xi_p - \epsilon).$$

Pelo Teorema 2.9, podemos escrever

$$\begin{aligned} P(\hat{\xi}_p > \xi_p + \epsilon) &= P(p > \hat{F}_n(\xi_p + \epsilon)) \\ &= P\left(\sum_{i=1}^n \mathbf{1}_{[X_i > \xi_p + \epsilon]} > n(1-p)\right) \\ &= P\left(\sum_{i=1}^n V_i - \sum_{i=1}^n E(V_i) > n\delta_1\right), \end{aligned}$$

onde  $V_i = \mathbf{1}_{[X_i > \xi_p + \epsilon]}$  e  $\delta_1 = F(\xi_p + \epsilon) - p$ . Da mesma forma,

$$\begin{aligned} P(\hat{\xi}_p < \xi_p - \epsilon) &= P(p > \hat{F}_n(\xi_p - \epsilon)) \\ &= P\left(\sum_{i=1}^n W_i - \sum_{i=1}^n E(W_i) > n\delta_2\right), \end{aligned}$$

onde  $W_i = \mathbf{1}_{[X_i < \xi_p - \epsilon]}$  e  $\delta_2 = p - F(\xi_p - \epsilon) - p$ . Portanto, utilizando o Lema de Hoeffding (Lema 2.20), temos

$$P(\widehat{\xi}_p > \xi_p + \epsilon) \leq \exp(-2n\delta_1^2)$$

e

$$P(\widehat{\xi}_p < \xi_p - \epsilon) \leq \exp(-2n\delta_2^2).$$

Colocando  $\delta_\epsilon = \min\{\delta_1, \delta_2\}$ , a prova está completa. ■

Demonstramos que

$$\lim_{n \rightarrow \infty} P(|\widehat{\xi}_p - \xi_p| > \epsilon) \leq \lim_{n \rightarrow \infty} 2 \exp(-2n\delta_\epsilon^2) = 0,$$

o qual significa que  $\widehat{\xi}_p \xrightarrow{P} \xi_p$ . Em outras palavras, sempre que  $\xi_p$  seja solução única da desigualdade  $F(\xi_p^-) \leq p \leq F(\xi_p)$ ,  $0 < p < 1$ , o quantil amostral converge em probabilidade para o quantil populacional e isto sempre acontece nas distribuições contínuas. Um detalhe importante é que para demonstrarmos a convergência em probabilidade de  $\widehat{\xi}_p$  utilizamos o Lema de Hoeffding e ele depende da existência da esperança.

O seguinte resultado fornece a distribuição assintótica da  $r$ -ésima estatística de ordem amostral de uma população com uma função de distribuição  $F$ , absolutamente contínua, e função de densidade  $f$ .

### Teorema 2.23

Seja  $0 < p < 1$ . Suponhamos que a amostra aleatória  $X_1, X_2, \dots, X_n$  tenha função de distribuição  $F$  com função de densidade  $f$  na vizinhança de  $\xi_p$ , sendo  $f$  positiva e contínua em  $\xi_p$ . Então

$$\sqrt{n}(\widehat{\xi}_p - \xi_p) \xrightarrow{D} N\left(0, \frac{p(1-p)}{f^2(\xi_p)}\right), \quad (2.39)$$

quando  $n \rightarrow \infty$  e  $\xi_p$  é a solução única da equação  $F(\xi_p) = p$ .

**Demonstração:** Vamos demonstrar somente para o caso  $p = 1/2$ . Observemos que  $\xi_{1/2}$  é mediana única dado que  $f(\xi_{1/2}) > 0$ . Primeiro, consideremos que  $n$  seja ímpar, por exemplo,  $n = 2m - 1$ , logo

$$P\left[\sqrt{n}\left(X_{(m)} - F^{-1}(1/2)\right) \leq t\right] = P\left(X_{(m)} \leq t/\sqrt{n} + F^{-1}(1/2)\right).$$

Seja  $S_n$  o número de  $X$  que excedem  $t/\sqrt{n} + F^{-1}(1/2)$ . Então

$$X_{(m)} \leq \frac{t}{\sqrt{n}} + F^{-1}(1/2) \quad \text{se, e somente se,} \quad S_n \leq m - 1 = \frac{n-1}{2}.$$

Percebemos que

$$S_n \sim \text{Binomial}\left(n, 1 - F\left(F^{-1}(1/2) + t/\sqrt{n}\right)\right).$$

Fazendo  $p_n = 1 - F\left(F^{-1}(1/2) + t/\sqrt{n}\right)$ , temos que

$$\begin{aligned} P\left[\sqrt{n}\left(X_{(m)} - F^{-1}(1/2)\right) \leq t\right] &= P\left(S_n \leq \frac{n-1}{2}\right) \\ &= P\left(\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \leq \frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right). \end{aligned}$$

Utilizando o Teorema de Berry-Esseen, temos que

$$\lim_{n \rightarrow \infty} \left\{ P\left(S_n \leq \frac{n-1}{2}\right) - \Phi\left(\frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right) \right\} = 0.$$

Escrevendo

$$\begin{aligned} \frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}} &\approx \frac{\sqrt{n}(\frac{1}{2} - p_n)}{1/2} \\ &= \frac{\sqrt{n}\left(-\frac{1}{2} + F(t/\sqrt{n} + F^{-1}(1/2))\right)}{1/2} \\ &= 2t \frac{F(t/\sqrt{n} + F^{-1}(1/2)) - F(F^{-1}(1/2))}{t/\sqrt{n}} \rightarrow 2tf(F^{-1}(1/2)). \end{aligned}$$

Então

$$\Phi\left(\frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right) \approx \Phi\left(2tf(F^{-1}(1/2))\right)$$

ou

$$\sqrt{n}\left(X_{(m)} - F^{-1}\left(\frac{1}{2}\right)\right) \xrightarrow{D} N\left(0, \frac{1}{4f^2(F^{-1}(1/2))}\right).$$

Quando  $n$  é par, digamos  $n = 2m$ , ambos  $P\left(\sqrt{n}(X_{(m)} - F^{-1}(1/2)) \leq t\right)$  quanto  $P\left(\sqrt{n}(X_{(m+1)} - F^{-1}(1/2)) \leq t\right)$  convergem a  $\Phi\left(2tf(F^{-1}(1/2))\right)$ . ■

Observe que o quantil amostral de ordem  $p$ ,  $\hat{\xi}_p$ , como consequência do Teorema 2.23, tem por distribuição assintótica

$$N\left(\xi_p, \frac{1}{[f(\xi_p)]^2} \frac{p(1-p)}{n}\right),$$

onde  $\xi_p$  é o correspondente quantil populacional e  $f$  é a função de densidade populacional. Por exemplo, suponha temos uma amostra aleatória da distribuição  $N(\mu, \sigma^2)$  de tamanho  $n$ . Seja  $\hat{\xi}_{1/2}$  a mediana amostral obtida dessa amostra. Então porque  $f(\mu) = 1/(\sqrt{2\pi}\sigma)$ , temos que

$$\sqrt{n}(\hat{\xi}_{1/2} - \mu) \xrightarrow{D} N\left(0, \frac{\pi\sigma^2}{2}\right).$$

Também devemos ter em consideração que para demonstrarmos o Teorema 2.23 utilizamos a Teorema de Berry-Esseen, o qual depende da existência dos primeiros dois momentos da variável aleatória. Com isso, caso  $X \sim \text{Cauchy}(\mu, \sigma)$ , o Teorema 2.23 não se aplica.

## 2.4 Gráficos descritivos

Vejamos alguns conjuntos de dados disponíveis na linguagem de programação R (R Core Team, 2014), especificamente na biblioteca `datasets`, que nos permitiram mostrar a utilidade dos momentos amostrais para resumir as informações contidas nos dados.

Para consultar estes conjuntos de dados basta digitar

```
library(help = "datasets")
```

Alguns dos diversos exemplos disponíveis serão apresentados aqui.

### Exemplo 2.22 (*Puromicina*)

Os dados sobre a velocidade de uma reação enzimática são obtidos por Treloar (1974) e disponíveis no arquivo de dados *Puromycin*. O número de contagens por minuto de produto radioativo a partir da reação foi medida como uma função da concentração do substrato em partes por milhão (ppm) e a partir destas contagens a taxa

inicial (ou velocidade) da reação foi calculada (contagens/min/min). O experimento foi realizado uma vez com a enzima tratada com puromicina e depois com a enzima não tratada.

A estrutura destes dados tem 23 linhas e 3 colunas, cada coluna contendo as informações das variáveis:

*conc*: um vector numérico de concentrações de substrato (ppm);

*rate*: um vector numérico de taxas de reação instantânea (contagens/min/min);

*state*: um fator com níveis *treated* (tratada) ou *untreated* (não tratada).

Para a leitura e observação dos nomes das variáveis utilizamos os comandos a seguir:

```
data(Puromycin)
names(Puromycin)
```

Uma maneira de obtermos estatísticas descritivas é utilizando as linhas de comando a seguir:

```
summary(rate[state=='treated'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  47.0  104.5  145.5  141.6  193.2  207.0
```

e

```
summary(rate[state=='untreated'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  51.0   85.0  115.0  110.7  137.5  160.0
```

para o caso da variável *rate*, as concentrações, obtidas as estatísticas descritivas segundo os níveis do fator *state*, se as concentrações foram ou não tratadas com puromicina.

No caso das estatísticas descritivas acerca das concentrações de substrato, variável *conc*, temos:

```
summary(conc[state=='treated'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.020  0.060  0.165  0.345  0.560  1.100
```

e

```
summary(conc[state=='untreated'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0200 0.0600 0.1100 0.2764 0.3900 1.1000
```

Os valores mínimos e máximos foram registrados sempre com os nomes de *Min.* e *Max.*, respectivamente. O primeiro e terceiro quantis ou quantis de 25% e 75% respectivos são registrados com os nomes *1st Qu.* e *3rd Qu.* e, finalizando, o resumo de informações de estatísticas de posição temos os valores de medianas (*Median*) e médios (*Mean*).

### Exemplo 2.23 (*Rock*)

Medições em 48 amostras de rochas de um reservatório de petróleo estão disponíveis no arquivo de dados *rock*. Este conjunto de dados contém 48 linhas e 4 colunas numéricas, descritas a seguir:

*area* área do espaço de poros, em pixels de 256 por 256;

*peri* perímetro em pixels;

*shape* perímetro/sqrt(*area*)

*perm* permeabilidade em mili-Darcies.

Doze amostras do núcleo de reservatórios de petróleo foram amostrados por 4 seções transversais. Cada amostra foi medida no núcleo para a permeabilidade e cada seção transversal tem uma área total de poros, perímetro total de poros e forma. A fonte destes dados é a BP Research e a análise das imagens foi de Ronit Katz, Oxford University.

Na geologia, a permeabilidade é a medida da capacidade de um material (tipicamente uma rocha) para transmitir fluídos. É de grande importância na determinação das características de fluxo dos hidrocarbonetos em reservatórios de petróleo e gás e da água nos aquíferos. A unidade de permeabilidade é o Darcy ou, mais habitualmente, o mili-Darcy ou mD.

### 2.4.1 Gráfico de Boxplot

Em 1977, John Tukey (Tukey, 1977) publicou uma proposta que posteriormente foi reconhecida como sendo um eficiente método para mostrar cinco números que resumem qualquer conjunto de dados. O gráfico proposto é chamado de boxplot (também conhecido como box and whisker plot) e resume as seguintes medidas de posição estatísticas: mediana, quantis inferior e superior e os valores mínimos e máximos. Os quantis inferior e superior entendem-se serem os quantis de 25% e 75%, respectivamente.

No caso do exemplo 2.22, deixamos a disposição os dados digitando

```
attach(Puromycin)
```

e com isso podemos mudar o nome dos níveis do fator da forma

```
state=factor(state,labels=c('Tratada','Não tratada'))
```

Então, com os comandos a seguir geramos o gráfico de boxplot, tanto para a variável `rate` quanto para a variável `conc`, estas segundo os níveis do fator `state`.

```
par(mar=c(5,4,3,1))
boxplot(rate ~ state, col = grey(c(0.4,1)),
        main='Taxas de reação instantânea')
```

para o caso do `rate`. Observemos que a primeira linha `par(mar=c(5,4,3,1))` serve somente para dimensionar a janela gráfica. Para o caso da variável `conc` utilizamos comandos semelhantes.

```
par(mar=c(5,4,3,1))
boxplot(conc ~ state, col = grey(c(0.4,1)),
        main='Concentrações de substrato')
```

O resultado deste trabalho pode ser observado na Figura 2.3.

Interpretemos o gráfico de boxplot. A caixa (box) propriamente contém a metade 50% dos dados. O limite superior da caixa indica o percentil 75% dos dados e o limite inferior da caixa indica o percentil 25%. A distância entre esses dois quantis é conhecida como inter-quantil. A linha na caixa indica o valor de mediana dos dados. Se a linha mediana dentro da caixa não é equidistante dos extremos, diz-se então que os dados são assimétricos. O boxplot da variável `rate` (esquerda na Figura 2.3) é um exemplo de dados simétricos já a situação da variável `conc` (direita na Figura 2.3) é um caso clássico de assimetria dos dados. Os extremos do gráfico indicam os valores mínimo e máximo, a menos que valores outliers<sup>3</sup> estejam presentes, nesse caso o gráfico se estende ao máximo de 1.5 vezes da distância inter-quantil. Os pontos fora do gráfico são então outliers ou suspeitos de serem outliers.

Mais elegante seria utilizar a biblioteca de funções `ggplot2`, para isso, digitamos:

```
library(ggplot2)
```

Para gerar os gráficos de boxplot respectivos, fazemos:

```
par(mar=c(5,4,3,1))
qplot(state, rate, geom=c("boxplot", "jitter"),
      main="Taxas de reação instantânea", xlab="", ylab=" ")
```

e

```
par(mar=c(5,4,3,1))
qplot(state, conc, geom=c("boxplot", "jitter"),
      main="Concentrações de substrato", xlab="", ylab=" ")
```

<sup>3</sup>Em estatística, outlier, valor aberrante ou valor atípico, é uma observação que apresenta um grande afastamento das demais observações em uma amostra. A existência de outliers implica, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados as amostras.

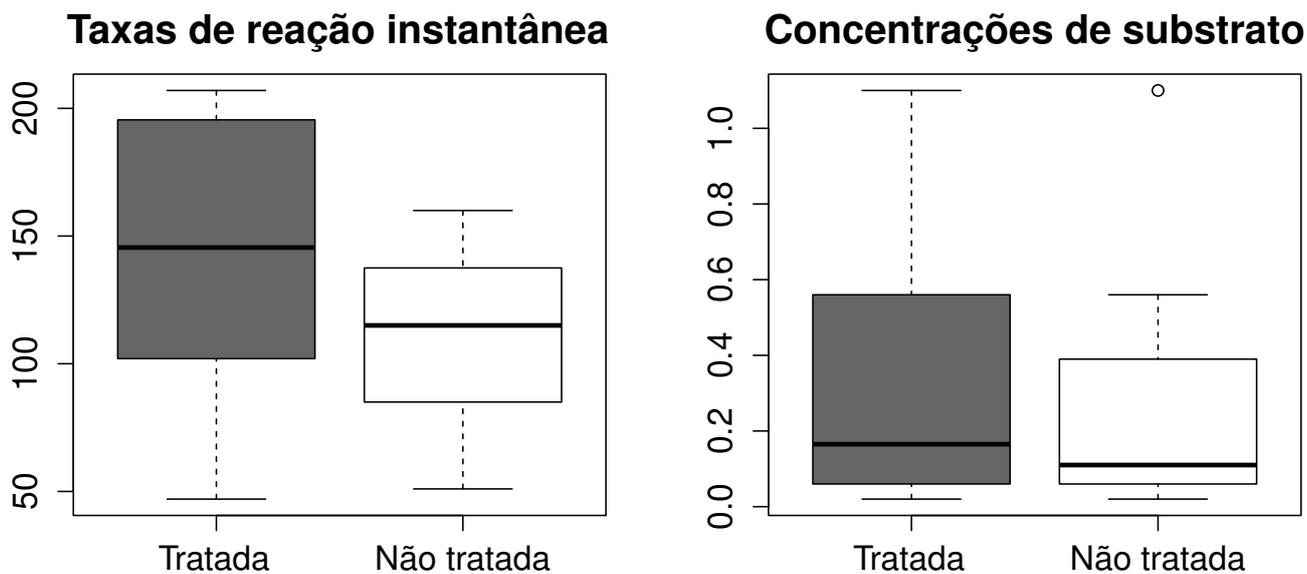


Figura 2.3: Gráfico de boxplot da variável *rate* à esquerda e da variável *conc* à direita, segundo os níveis do fator *state*, se a enzima foi tratada ou não com puromicida. Gráfico gerado utilizando a função R `boxplot`,

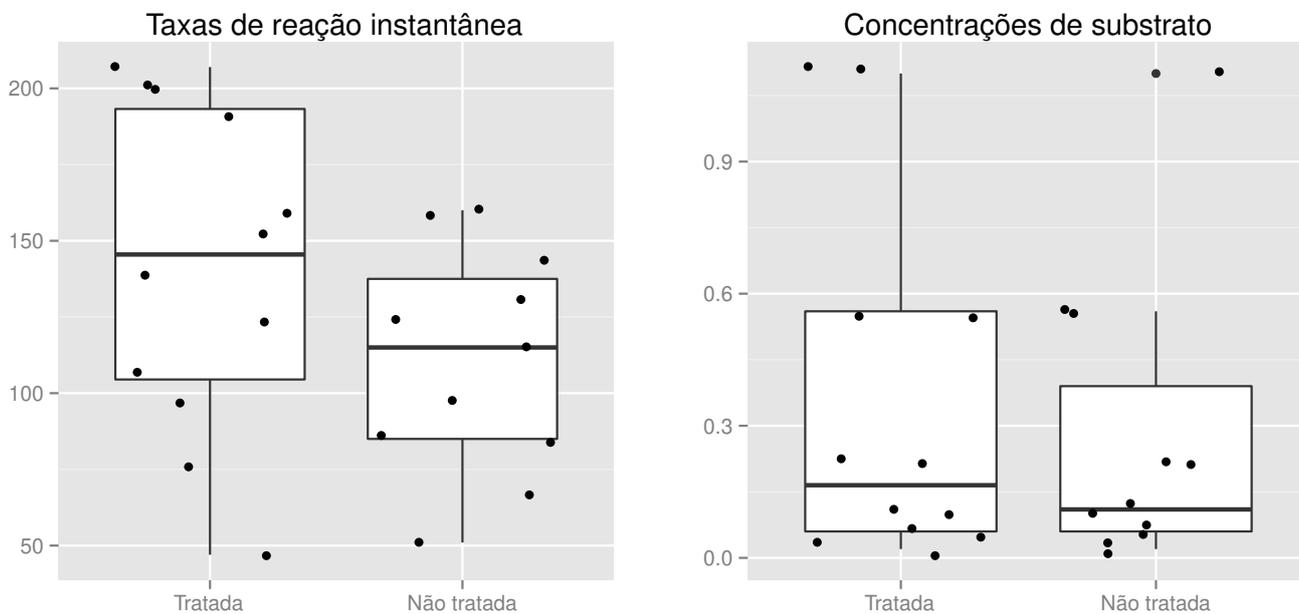


Figura 2.4: Gráfico de boxplot da variável *rate* à esquerda e da variável *conc* à direita, segundo os níveis do fator *state*, se a enzima foi tratada ou não com puromicida. Gráfico gerado utilizando a função R `qplot`, opção `geom=c("boxplot", "jitter")`.

obtendo-se assim os gráficos na Figuras 2.4. Além de melhor qualidade gráfica acrescentamos os pontos observados no boxplot, isso permite termos uma ideia também da dispersão dos dados.

Vejam as vantagens do boxplots. Mostra graficamente a posição central dos dados (mediana) e a tendência. Fornece algum indicativo de simetria ou assimetria dos dados. Ao contrário de muitas outras formas de mostrar os dados, o boxplots mostra os outliers. Utilizando o boxplot para cada variável categórica no mesmo gráfico, pode-se facilmente comparar os dados. Esta é a situação no exemplo na Figura 2.3, podemos observar o comportamento das variáveis *rate* e *conc* segundo os níveis do fator *state*.

Um detalhe do boxplot é que ele tende a enfatizar as caudas da distribuição, que são os pontos ao extremo nos dados. Também fornece detalhes da distribuição dos dados. Mostrar o histograma (Seção 2.4.2) em conjunto com o boxplot ajuda a entender a distribuição dos dados, constituindo estes dos gráficos ferramentas importantes na análise exploratória.

Logicamente, o comportamento dos dados dentro da caixa (box), como podemos perceber nas figuras 2.3 e 2.4, permanece um mistério. Isso porque caso estejam os dados bem espalhados ou não, o gráfico boxplot continua mostrando uma caixa. Somente perceberemos algum comportamento diferente se o valor da mediana estiver mais próximo de um dos extremos desta caixa. Para tentar diminuir essa limitação foi sugerido uma melhoria, obtendo-se o chamada boxplot entalhado (notched boxplot).

Com as linhas de comando a seguir se obtém os gráficos na Figura 2.5.

```
par(mar=c(5,4,3,1))
boxplot(rate ~ state, col = grey(c(0.4,1)), notch=TRUE,
        main='Taxas de reação instantânea')
```

e

```
par(mar=c(5,4,3,1))
boxplot(conc ~ state, col = grey(c(0.4,1)), notch=TRUE,
        main='Concentrações de substrato')
```

Observa-se que a única diferença é a inclusão da opção `notch=TRUE`, permanecendo todas as outras instruções iguais.

Mais elaborado é o chamado *violin plot*, mistura de boxplot com estimação de densidade, tema este tratado na Seção 4.3. Este gráfico, introduzido no artigo Hintze & Nelson (1998), sinergicamente combina o gráfico de boxplot e a estimação da densidade, também chamado de histograma suavizado, em uma única tela que revela a estrutura encontrada nos dados.

Com as linhas de comando a seguir se obtém os gráficos na Figura 2.6.

```
par(mar=c(5,4,3,1))
qplot(state, rate, geom = c("violin", "jitter"), notch=TRUE,
      main="Taxas de reação instantânea", xlab="", ylab=" ")
```

e

```
par(mar=c(5,4,3,1))
qplot(state, conc, geom=c("violin", "jitter"), notch=TRUE,
      main="Concentrações de substrato", xlab="", ylab=" ")
```

Este gráfico é similar ao boxplot excepto que mostra também a densidade de probabilidade dos dados. Pode incluir também um marcador para a média dos dados e uma caixa que indica a distância interquartil, como nos gráficos boxplot. O objetivo do gráfico *violin plot* é o mesmo do que o boxplot original porém, considera de alguma maneira o comportamento dos dados dentro da caixa (box). Assim, percebemos melhor a distribuição dos dados dentro do intervalo interquartil.

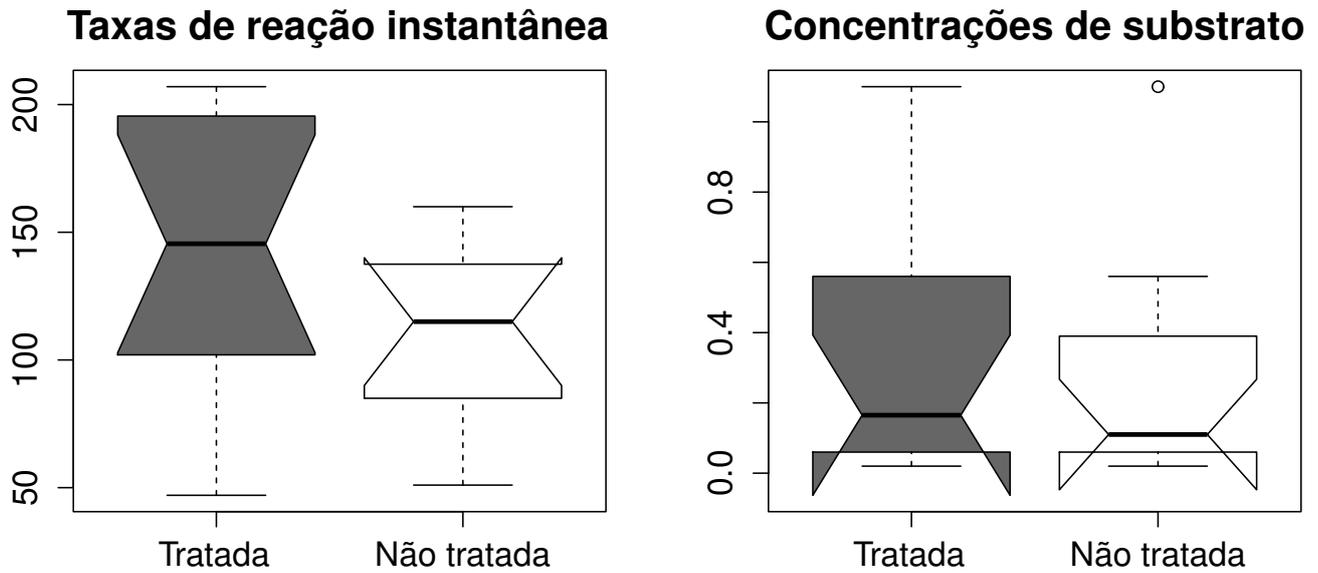


Figura 2.5: Gráfico de boxplot entalhado da variável `rate` à esquerda e da variável `conc` à direita, segundo os níveis do fator `state`, se a enzima foi tratada ou não com puromicida. Gráfico gerado utilizando a função R `qplot`, opção `geom=c("boxplot", "jitter")`, `notch=TRUE`.

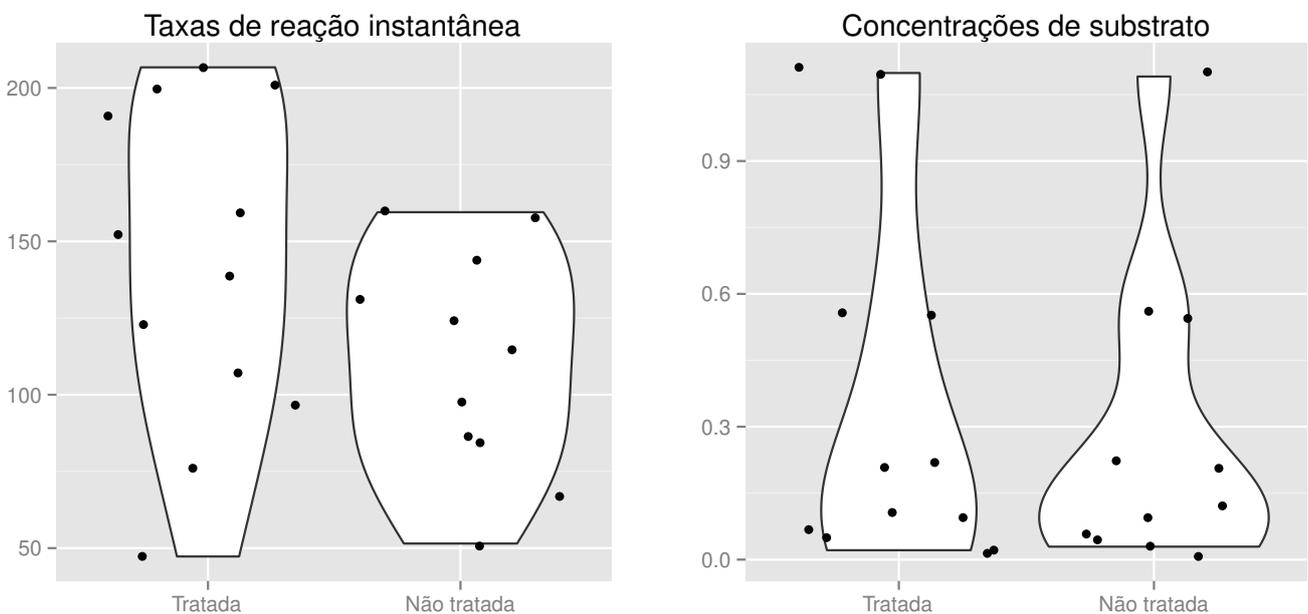


Figura 2.6: Gráfico de violin plot da variável `rate` à esquerda e da variável `conc` à direita, segundo os níveis do fator `state`, se a enzima foi tratada ou não com puromicida. Gráfico gerado utilizando a função R `qplot`, opção `geom=c("vioplot", "jitter")`, `notch=TRUE`.

## 2.4.2 Histograma

Um histograma é uma representação gráfica da função de probabilidades ou da função de densidade de um conjunto de dados independentes e foi introduzido pela primeira vez por Karl Pearson<sup>4</sup>. A representação mais comum do histograma é um gráfico de barras verticais. A palavra histograma é de origem grega, derivada de duas: *histos* que pode significar testemunha no sentido de *aquilo que se vê*, como as barras verticais do histograma, e da também palavra grega *gramma* que significa *desenhar, registrar ou escrever*.

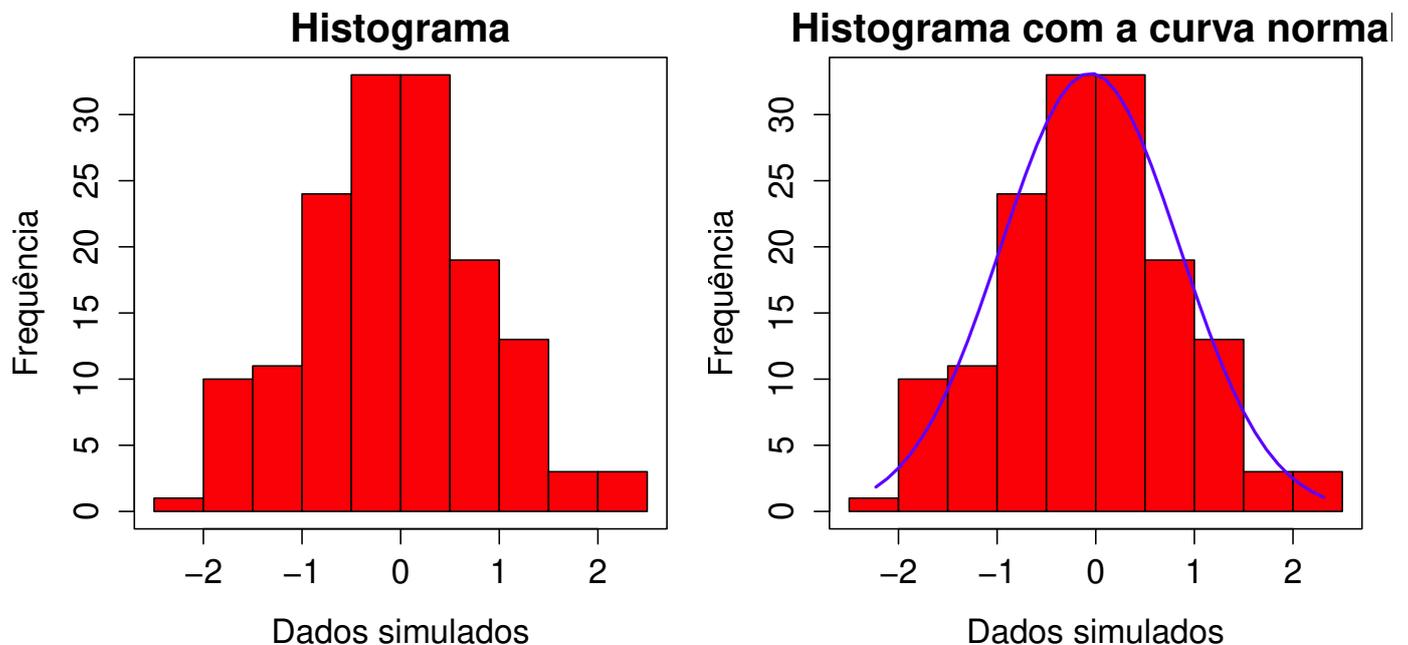


Figura 2.7: Gráfico de histograma para dados simulados.

Para construir um exemplo controlado do gráfico de histograma, simulamos uma amostra de tamanho 150 da distribuição normal padrão, com o comando

```
x=rnorm(150)
```

e, depois, construímos um gráfico colorido com as linhas de comando

```
par(mar=c(5,4,2,1))
hist(x, breaks=12, col="red", xlab="Dados simulados",
      ylab='Frequência', main="Histograma")
box()
```

Posteriormente, acrescentamos a este gráfico uma linha com a densidade normal

```
par(mar=c(5,4,2,1))
h=hist(x, breaks=10, col="red", xlab="Dados simulados",
       ylab='Frequência', main="Histograma com a curva normal")
xfit=seq(min(x),max(x),length=40)
yfit=dnorm(xfit,mean=mean(x),sd=sd(x))
yfit=yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
box()
```

<sup>4</sup>Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 186: 343-414.

Desta forma geramos os gráficos na Figura 2.7. A ideia é mostrar que o histograma assemelha-se ao gráfico da densidade normal, a densidade dos dados.

### Definição 2.8

Matematicamente o histograma é uma função  $m_i$  que conta o número de observações que pertencem a vários intervalos disjuntos, entanto que o gráfico do histograma ou simplesmente histograma é uma mera representação desta função. Assim, se  $n$  representa o total de observações e  $k$  o número de intervalos disjuntos, o histograma satisfaz que

$$n = \sum_{i=1}^k m_i.$$

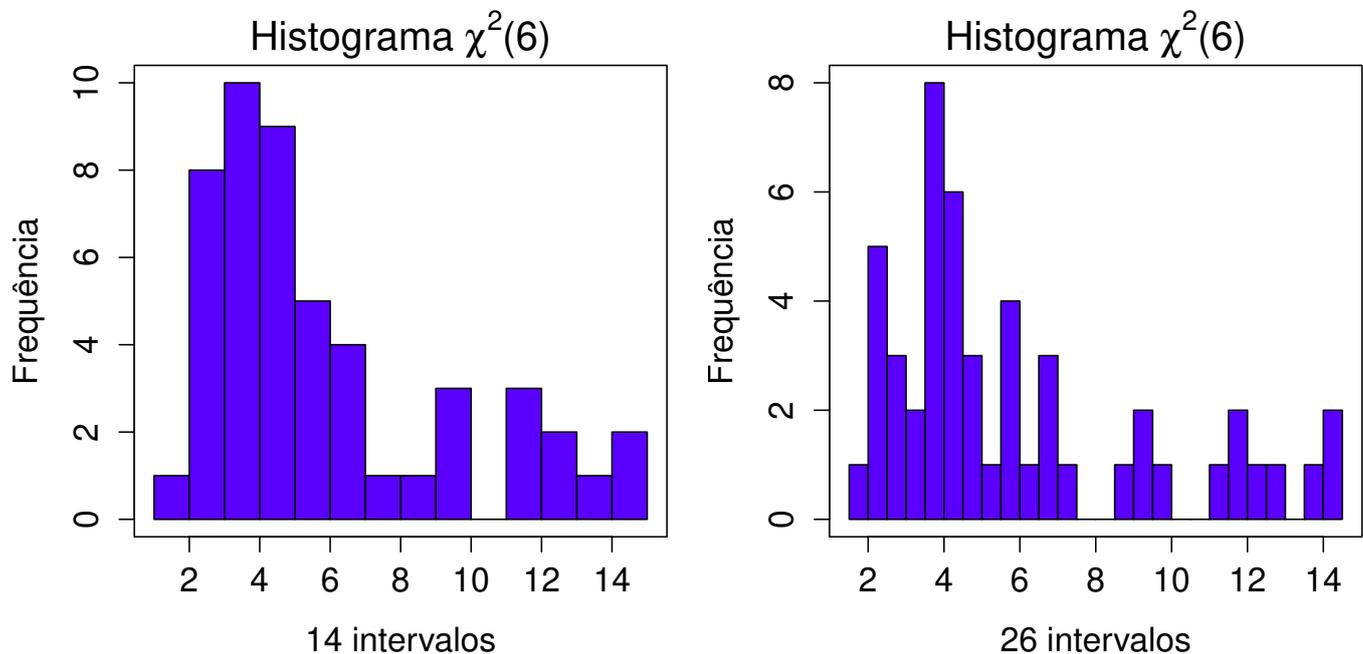


Figura 2.8: Histogramas da distribuição  $\chi^2$  com 6 graus de liberdade. Número de intervalos 14 e 26, respectivamente.

O histograma é um gráfico composto por retângulos justapostos em que a base de cada um deles corresponde ao intervalo de classe e a sua altura à respectiva frequência. A construção de histogramas tem caráter preliminar em qualquer estudo e é um importante indicador da distribuição de dados. Pode indicar se uma distribuição aproxima-se de uma densidade normal como pode indicar mistura de densidades, quando os dados apresentam várias modas.

Os histogramas podem ser um mau método para determinar a forma de uma distribuição porque são fortemente influenciados pelo número de intervalos utilizados. Por exemplo, decidimos gerar 50 amostras da densidade  $\chi^2(6)$ , da forma

```
set.seed(5678)
z=rchisq(50, df=6)
```

Os gráficos de histogramas correspondentes com 14 e 26 intervalos são apresentados na Figura 2.8 e foram gerados com as linhas de comando

```
par(mar=c(5,4,2,1))
hist(z, breaks=14, col="blue", main=expression(paste('Histograma ',
  chi^2,'(6)')), ylab='Frequência', xlab='14 intervalos')
box()
```

e

```
par(mar=c(5,4,2,1))
hist(z, breaks=26, col="blue", main=expression(paste('Histograma ',
  chi^2,'(6)')), ylab='Frequência', xlab='26 intervalos')
box()
```

Na Figura 2.9 podemos observar os gráficos de histograma obtidos das variáveis descritas no Exemplo 2.23. A situação em (a) representa o caso em a distribuição dos dados de assemelha à distribuição normal, já a situação descrita no gráfico em (b) mostra-se uma mistura de densidades, percebemos a existência de duas modas.

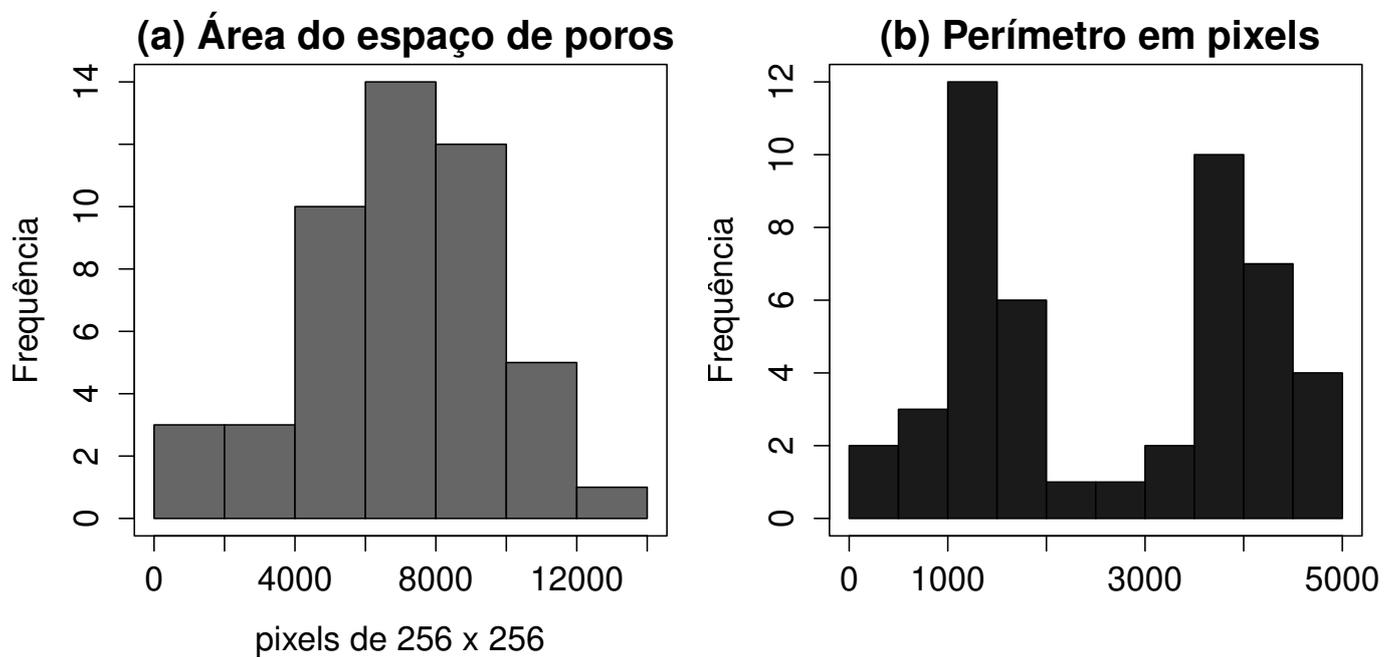


Figura 2.9: Histogramas das variáveis no Exemplo 2.23.

Outras situações no mesmo exemplo, mas diferentes variáveis, são descritas nos gráficos na Figura 2.10. Nessa figura apresentamos dois gráficos, chamados de (c) e (d), nesta figura. Correspondem, como podemos observar, à distribuições assimétricas e descrevem os dados coletados nas variáveis **shape** e **perm** do arquivo de dados Rock, Exemplo 2.23. Os histogramas foram pensados somente para o caso de variáveis contínuas, porém é uma descrição discreta delas. Logicamente, também podemos utiliza-los em situações de variáveis aleatórias discretas, nada impede isso.

Estas figuras foram geradas utilizando a configuração padrão do comando `hist`, isto é, utilizamos uma maneira automática de determinar o número de intervalos, mais adiante dedicamos maior atenção a diferentes formas de calcular este número.

Como pode ter sido observado, além de não ficar claro como determinar o número de intervalos nem como delimitar os intervalos, também não ficou claro o que queremos realmente observar com o gráfico desta função.

Vejam agora uma definição mais clara do histograma, esta definição nos permitirá obter propriedades importantes.

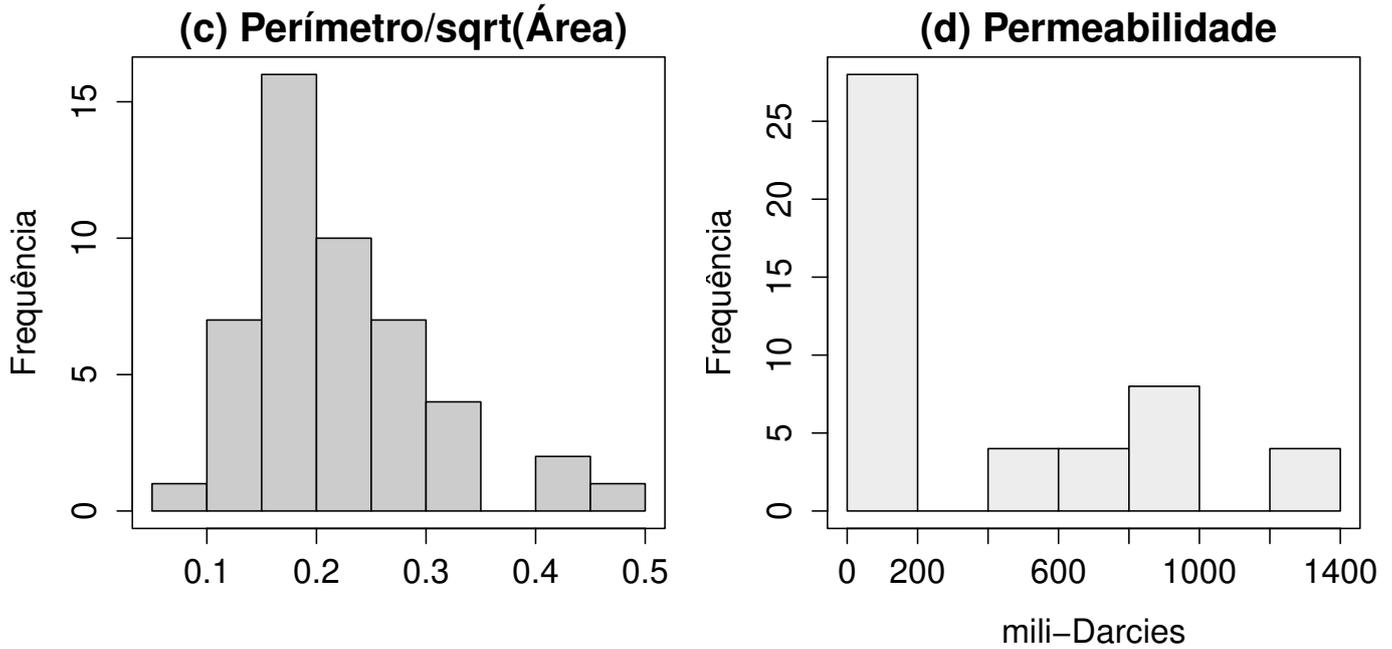


Figura 2.10: Histogramas das variáveis no Exemplo 2.23.

### Definição 2.9

Sejam  $I_1, \dots, I_k$  intervalos disjuntos do suporte da função de probabilidade ou de densidade da variável aleatória  $X$ . O histograma é definido por

$$\hat{f}(x) = \frac{m_i}{n|I_i|}, \quad \forall x \in I_i, i = 1, 2, \dots, k, \quad (2.40)$$

onde  $|I_i|$  representa o comprimento do intervalo  $i$ ,  $m_i$  e  $n$  como na definição 2.8.

Foi provado por Robertson (1967) que, dados os intervalos  $I_1, I_2, \dots, I_k$ , o histograma  $\hat{f}$  é um estimador de máxima verossimilhança<sup>5</sup> dentre os estimadores expressados como funções simples e semicontínuas superiormente, isto se o fecho de cada intervalos contiver duas ou mais observações. Os gráficos apresentados nas figuras 2.7, 2.9 e 2.10 são histogramas também segundo a proposta de Robertson (1967).

Pode-se observar que este estimador tem duas limitações importantes: a dependência do comprimento do intervalo e o fato de o histograma não constituir uma função contínua. A primeira destas limitações foi amplamente estudada por Wegman (1975). Ele provou que os pontos extremos de cada intervalo  $I_k$  devem ser coincidentes com observações e que, se o número mínimo de observações em cada intervalo aumente, conforme aumenta o tamanho da amostra, o estimador  $\hat{f}$  é consistente<sup>6</sup>.

A segunda limitação importante do histograma, isto é, o fato de ele não constituir uma função contínua, incentivou diversos estudos na procura de estimadores contínuos da função de densidade. No Capítulo 3, a Seção 4.3 dedica-se a mostrar estimadores contínuos da função de densidade.

<sup>5</sup>Os estimadores de máxima verossimilhança serão estudados na Seção 4.2

<sup>6</sup>Estimadores consistentes serão estudados na Seção 3.1.1

### Cálculo automático do número de intervalos num histograma

Uma questão importante é determinar de maneira automatizada o número de intervalos disjuntos que serão utilizados para a construção do gráfico.

Uma primeira forma de escolher o número de intervalos foi dada por Sturges (1926) e que constitui a forma padrão no *R*. Conhecida como fórmula de Sturges é dada por

$$k = \lceil \log_2(n) + 1 \rceil, \quad (2.41)$$

isto significa que o número de intervalos é a parte inteira do logaritmo base 2 do número de observações mais 1.

Outras expressões comumente utilizadas são a fórmula de Scott (Scott, 1979)  $h = 3.5s/\sqrt[3]{n}$ , onde  $s$  é o desvio padrão e a fórmula de Freedman Diaconis (Freedman & Diaconis, 1981)  $h = 2IQR(x)/\sqrt[3]{n}$ , onde  $IQR$  é a diferença entre o terceiro e o primeiro quantil.

#### Exemplo 2.24

Na biblioteca de funções *R* `robustbase` temos disponíveis dados do teor de cálcio e do pH em amostras de solo coletadas em diferentes comunidades da região de Condroz, na Bélgica.

Podemos ler estes dados digitando as linhas de comando abaixo, primeiro para escolher a biblioteca de funções e depois para selecionar os dados.

```
library(robustbase)
data(condroz)
```

Temos registadas duas variáveis: *Ca* que registra o teor de cálcio na amostra de solo e o *pH*, o pH correspondente. Construímos histogramas da variável *Ca* segundo a três formas de escolha do número de intervalos e os apresentamos na Figura 2.11.

Os dados deste exemplo foram publicados em: Hubert, M. and Vandervieren, E. (2006). An Adjusted Boxplot for Skewed Distributions, Technical Report TR-06-11, KULeuven, Section of Statistics, Leuven.

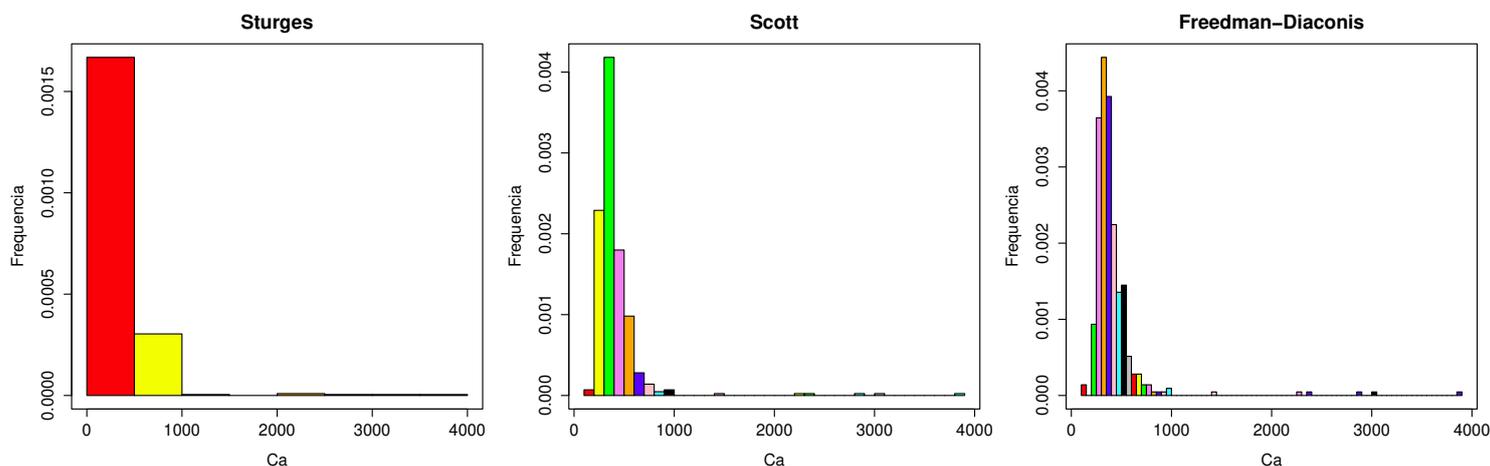


Figura 2.11: Diferentes histogramas da variável *Ca* no Exemplo 2.24.

### 2.4.3 Gráficos para verificar normalidade

Um primeiro gráfico chamado de qq-norm permite a comparação de duas distribuições de probabilidades traçando seus quantis uns contra os outros. Depois exploramos um gráfico mais recente, conhecido como worm plot (gráfico de minhoca), consistindo numa determinada coleção de de qq-norm.

### QQ-norm

O gráfico quantil-quantil ou qq-plot, proposto por Wilk & Gnanadesikan (1968), é um dispositivo gráfico exploratório utilizado para verificar a validade de um pressuposto de distribuição para um conjunto de dados. Em geral, a ideia básica é a de calcular o valor teoricamente esperado para cada ponto de dados com base na distribuição em questão. Se os dados de fato seguirem a distribuição assumida os pontos deste gráfico formarão aproximadamente uma linha reta.

Percebemos que podemos verificar com este gráfico qualquer densidade contínua, eventualmente pode ser utilizado também para funções de probabilidade. O qq-plot vai apresentar-se como uma linha reta se a densidade assumida estiver correta. Vejamos o caso particular de verificarmos se a densidade é normal, nesta situação o gráfico qq-plot será chamado de qq-norm. Primeiro consideraremos a situação da densidade normal padrão.

Seja  $z_1, z_2, \dots, z_n$  uma amostra aleatória de uma distribuição normal com média  $\mu = 0$  e desvio padrão  $\sigma = 1$ . As estatísticas de ordem amostrais são

$$z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}.$$

Estes valores desempenharão o papel dos quantis da amostra. Agora, quais devemos tomar como os quantis teóricas correspondentes? Se a função de distribuição cumulada da densidade normal padrão fosse denotada por  $\Phi$ , usando a notação quantil, se  $\xi_q$  é o  $q$ -ésimo quantil de uma distribuição normal, então

$$\Phi(\xi_q) = q,$$

ou seja, a probabilidade de uma amostra normal ser inferior a  $\xi_q$  é, de fato, apenas  $q$ .

Considere o primeiro valor ordenado  $z_{(1)}$ . O que podemos esperar que o valor  $\Phi(z_{(1)})$  seja? Intuitivamente, esperamos que essa seja a probabilidade de assumir um valor no intervalo  $(0, 1/n)$ . Do mesmo modo, espera-se que  $\Phi(z_{(2)})$  seja a probabilidade de assumir um valor no intervalo  $(1/n, 2/n)$ . Continuando, esperamos que  $\Phi(z_{(n)})$  seja a probabilidade de assumir um valor no intervalo  $(n-1)/n, 1)$ . Assim, o quantil teórico desejamos seja definido pelo inverso da função de distribuição acumulada normal padrão. Em particular, o quantil teórico correspondente ao quantil empírico  $z_{(i)}$  deve ser

$$\xi_q = q \approx \frac{i - 0,5}{n},$$

para  $i = 1, 2, \dots, n$ .

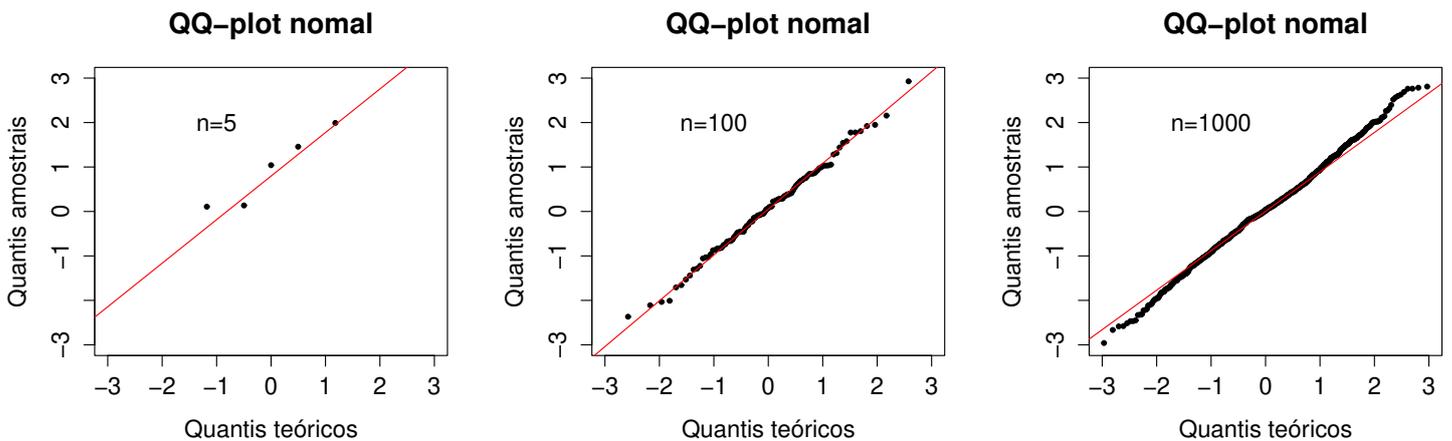


Figura 2.12: Diferentes qqplot para dados normais.

Na Figura 2.12, a esquerda acima exibimos o qq-norm de uma pequena amostra normal de tamanho 5. Os restantes quadros na Figura 2.12 exibem as plotagens de qq-norm para amostras normais de tamanhos  $n = 100$  e

$n = 1000$ , respectivamente. Como o tamanho da amostra aumenta, os pontos encontram-se mais perto da linha  $y = x$ .

Estes gráficos (Figura 2.12) foram gerados utilizando as linhas de comando:

```
set.seed(1278)
x=rnorm(5)
qqnorm(x, xlim=c(-3,3), ylim=c(-3,3), cex=0.6, pch=19,
       ylab='Quantis amostrais', xlab='Quantis teóricos',
       main='QQ-plot nomal')
qqline(x,col="red")
text(-1,2,'n=5')
```

para a situação de amostra de tamanho 5. A primeira linha de comando serve para fixar o gerador de números aleatórios e, dessa forma, podemos simular sempre a mesma amostra e reproduzir o gráfico idêntico. Nas outras situações somente muda-se o tamanho da amostra que se quer gerar.

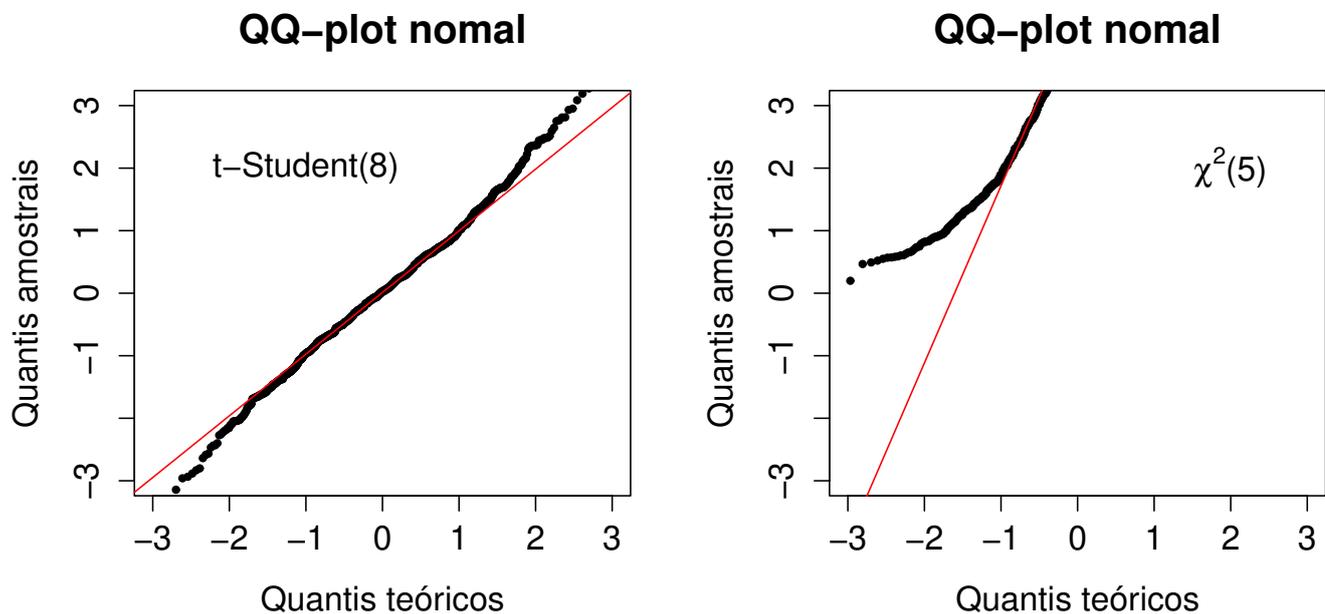


Figura 2.13: Diferentes qqplot para dados não normais.

Assim, os comandos para gerar o segundo e terceiro gráficos são:

```
x=rnorm(100)
qqnorm(x, xlim=c(-3,3), ylim=c(-3,3), cex=0.6, pch=19,
       ylab='Quantis amostrais', xlab='Quantis teóricos',
       main='QQ-plot nomal')
qqline(x,col="red")
text(-1,2,'n=100')

x=rnorm(1000)
qqnorm(x, xlim=c(-3,3), ylim=c(-3,3), cex=0.6, pch=19,
       ylab='Quantis amostrais', xlab='Quantis teóricos',
       main='QQ-plot nomal')
qqline(x,col="red")
text(-1,2,'n=1000')
```

Caso os dados não forem padronizados basta aplicar a transformação  $(X - \hat{\mu})/\hat{\sigma}$ , onde  $X$  representa os dados originais e  $\hat{\mu}$  e  $\hat{\sigma}$  representam os estimadores dos parâmetros  $\mu$  e  $\sigma$ , respectivamente.

Estes gráficos podem indicar afastamentos da normalidade por isso apresentamos duas situações de dados não simétricos e com caudas pesadas. Na Figura 2.13, mostramos o que acontece se os dados forem da distribuição t-Student(8) e da distribuição  $\chi^2(5)$ , sempre de tamanho  $n = 1000$ . Observe, em particular, que os dados a partir da distribuição t-Student seguem a curva normal bem de perto até os últimos pontos em cada extremo. Na outra situação o afastamento da distribuição normal é evidente.

Foi mencionado que o qq-norm é uma situação particular do qq-plot devido a este último permitir comparar os quantis amostrais com os quantis distribucionais. Com isto queremos dizer que o qq-plot serve para verificar se os dados forem t-Student ou  $\chi^2(5)$ , por exemplo. Na Figura 2.14 apresentamos a aparência dos gráficos qq-plot caso queira-se verificar se as amostras seguem distribuição t-Student(8) ou  $\chi^2(5)$ , respectivamente.

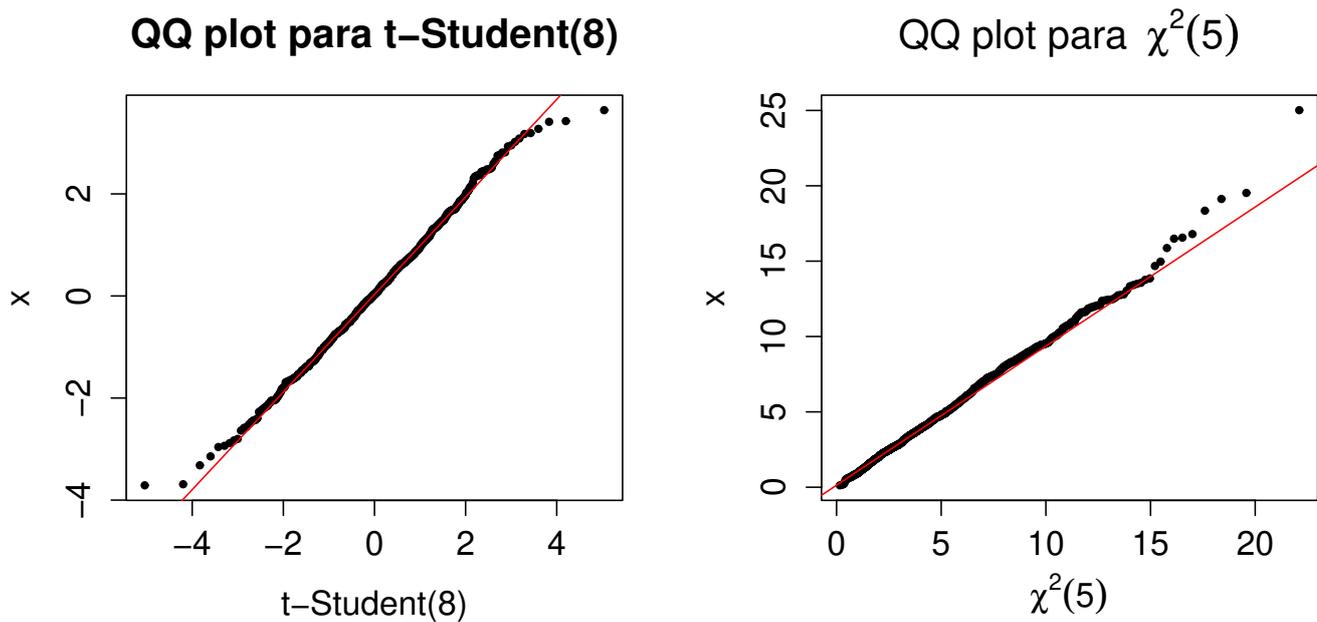


Figura 2.14: Diferentes qqplot para dados não normais.

Os gráficos na Figura 2.14 foram gerados pelas linhas de comandos

```
qqplot(qt(ppoints(1000), df = 8), x, cex=0.6, pch=19, main = "QQ plot para t-Student(8)",
      xlab="t-Student(8)")
qqline(x, distribution = function(p) qt(p, df = 8), prob = c(0.1, 0.6), col = 2)
```

no caso t-Student(8) e

```
qqplot(qchisq(ppoints(1000), df = 5), x, cex=0.6, pch=19,
      main = expression("QQ plot para" ~ {chi^2}(5)), xlab=expression({chi^2}(5)))
qqline(x, distribution = function(p) qchisq(p, df = 5), prob = c(0.1, 0.6), col = 2)
```

para o caso  $\chi^2(5)$ .

### Worn plot

O worm-plot é uma série de parcelas de gráficos qq-plot retificados. Constitui uma ferramenta de diagnóstico para visualização de quão bem um modelo estatístico se ajusta aos dados, para encontrar locais em que o ajuste pode ser melhorado e para comparar o ajuste de diferentes modelos.

Na Figura 2.15 mostramos este gráfico para duas situações: a esquerda os dados são normais e a direita os dados são t-Student com 8 graus de liberdade. Nesta situação aparece bem a qualidade da observação com esta

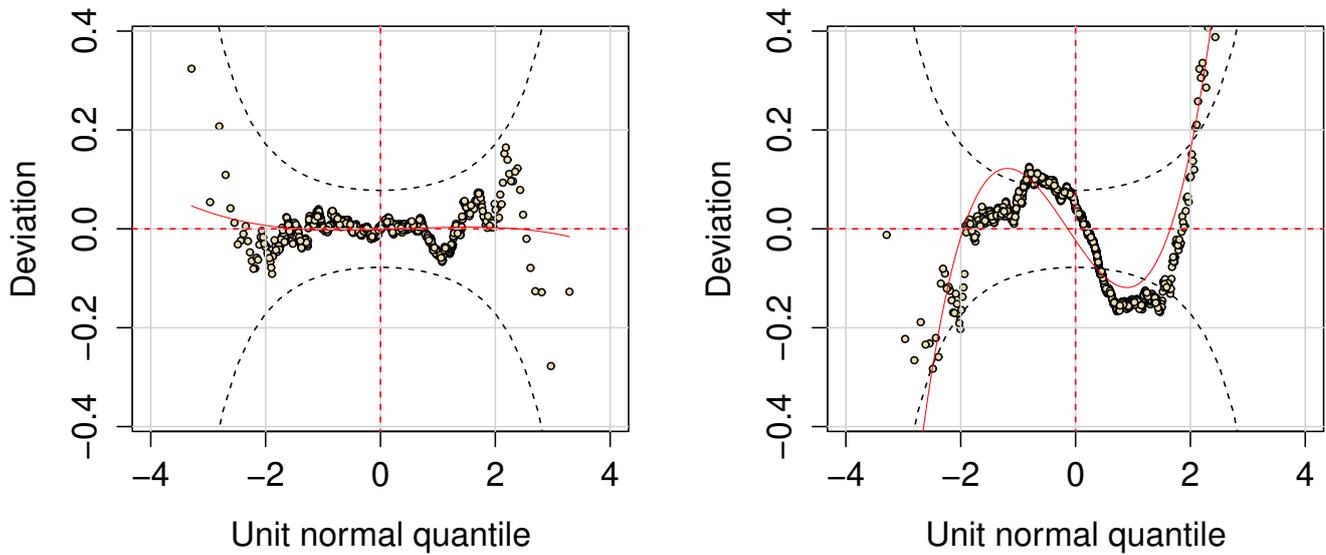


Figura 2.15: Diferentes worm-plot para dados normais.

figura. Se os dados forem normais a curva worm-plot ou gráfico de minhoca deve aparentar um verme achatado, os pontos próximos a curva vermelha e com poucas oscilações. Quando aplicamos este gráfico ao caso t-Student percebemos uma oscilação grande no verme e com pontos fugindo da banda de confiança. Isso comprova que os dados não seguem como referência a distribuição normal.

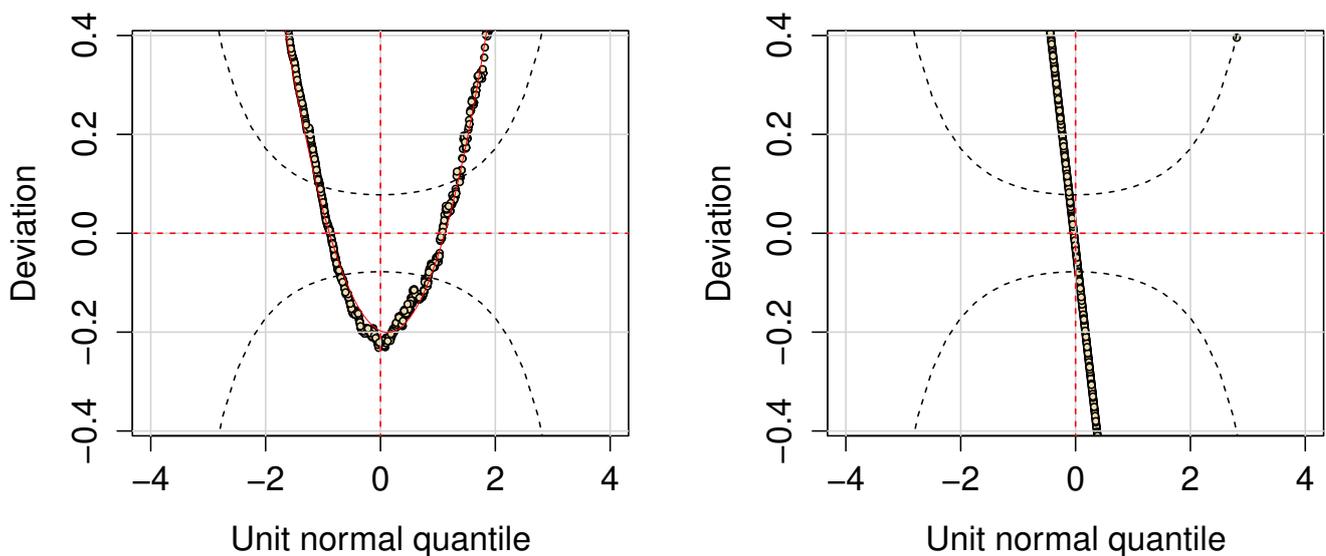


Figura 2.16: Diferentes worm-plot para dados não normais.

As linhas a seguir mostram os comandos necessários para gerar os gráficos na Figura 2.15. Utilizamos a biblioteca de comandos R `gamlss` (Rigby & Stasinopoulos, 2005).

```
library(gamlss)
x=rnorm(1000)
wp(gamlss(x~1), cex=0.6)
x=rt(1000, df=8)
wp(gamlss(x~1), cex=0.6)
```

Na Figura 2.16, a esquerda temos o caso de dados com distribuição  $\chi^2(5)$  e a direita dados com distribuição Cauchy padrão. Nestas situações fica claro que os dados não são normais. Os gráficos na figura foram gerados pelas linhas de comando a seguir.

```
x=rchisq(1000, df=5)
wp(gamlss(x~1), cex=0.6)
x=rcauchy(1000)
wp(gamlss(x~1), cex=0.6)
```

## 2.5 Exercícios

### Exercícios da Seção 2.1

1. Seja  $X \sim \text{Bernoulli}(\frac{1}{2})$  e considere todas as possíveis amostras aleatórias de tamanho  $n = 3$ . Calcule  $\bar{X}_n$  e  $S_n^2$  cada uma das oito amostras. Encontre a função de probabilidade de  $\bar{X}_n$  e  $S_n^2$ .
2. Um dado é lançado. Seja  $X$  o valor da face superior que aparece e  $X_1, X_2$  duas observações independentes de  $X$ . Encontre a função de probabilidade de  $\bar{X}_n$ .
3. Seja  $X_1, \dots, X_n$  uma amostra aleatória de alguma população. Mostre que

$$\max_{1 \leq i \leq n} |X_i - \bar{X}_n| < \frac{(n-1)S_n}{\sqrt{n}},$$

onde  $S_n$  é a raiz quadrada positiva da variância amostral  $S_n^2$ .

### Exercícios da Seção 2.2

1. Seja  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  o conjunto das estatísticas de ordem de  $n$  variáveis aleatórias independentes  $X_1, X_2, \dots, X_n$  com função de densidade comum

$$f(x) = \begin{cases} \beta e^{-x\beta}, & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases}.$$

- a) Mostre que  $X_{(s)}$  e  $X_{(r)} - X_{(s)}$  são independentes para quaisquer  $r > s$ .
  - b) Encontre a função de densidade de  $X_{(r+1)} - X_{(r)}$ .
  - c) Seja  $Z_1 = nX_{(1)}$ ,  $Z_2 = (n-1)(X_{(2)} - X_{(1)})$ ,  $Z_3 = (n-2)(X_{(3)} - X_{(2)})$ , ...,  $Z_n = ((X_{(n)} - X_{(n-1)}))$ . Prove que  $(Z_1, Z_2, \dots, Z_n)$  e  $(X_1, X_2, \dots, X_n)$  são identicamente distribuídas.
2. Provar o Teorema 2.1
  3. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias com distribuição geométrica de parâmetros  $p_1, p_2, \dots, p_n$ , respectivamente. Prove que  $N_n = \min(X_1, X_2, \dots, X_n)$  têm também distribuição geométrica de parâmetro

$$p = 1 - \prod_{i=1}^n (1 - p_i).$$

4. As  $X_1, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas tem por função de probabilidade  $BN(1; p)$  se, e somente se,  $N_n = \min(X_1, \dots, X_n)$  tem distribuição geométrica de parâmetro  $1 - (1 - p)^n$ .
5. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e igualmente distribuídas com função de densidade comum

$$f(x) = \begin{cases} \frac{1}{\sigma} e^{-[(x-\theta)/\sigma]}, & \text{se } x > \theta \\ 0, & \text{se } x \leq \theta \end{cases}$$

Mostre que  $X_{(1)}, X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)}, \dots, X_{(n)} - X_{(n-1)}$  são independentes.

6. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e igualmente distribuídas com função de distribuição acumulada comum

$$F(t) = \begin{cases} 0, & \text{se } t \leq 0 \\ t^\alpha, & \text{se } 0 < t < 1 \\ 1, & \text{se } t \geq 1 \end{cases}$$

para  $\alpha > 0$ . Mostre que  $X_{(i)}/X_{(n)}$ ,  $i = 1, 2, \dots, n-1$  e  $X_{(n)}$  são independentes.

7. Sejam  $X_1$  e  $X_2$  duas variáveis aleatórias discretas independentes com função de probabilidade comum

$$P(X = x) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, \dots; \quad 0 < \theta < 1.$$

Mostre que  $X_{(1)}$  e  $X_{(2)} - X_{(1)}$  são independentes.

8. Sejam  $X_1, \dots, X_n$  duas variáveis aleatórias independentes com função de densidade comum  $f$ . Encontre a função de densidade de  $X_{(1)}$  e de  $X_{(n)}$ .

9. Sejam  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  as estatísticas de ordem de  $n$  variáveis aleatórias independentes e igualmente distribuídas  $X_1, X_2, \dots, X_n$  com função de densidade comum

$$f(x) = \begin{cases} 1 & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Prove que  $Y_1 = X_{(1)}/X_{(2)}$ ,  $Y_2 = X_{(2)}/X_{(3)}$ ,  $\dots$ ,  $Y_{n-1} = X_{(n-1)}/X_{(n)}$  e  $Y_n = X_{(n)}$  são independentes. Encontre a função de densidade conjunta de  $Y_1, Y_2, \dots, Y_n$ .

10. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes identicamente distribuídas não negativas contínuas. Prove que se  $E|X| < \infty$ , então  $E|X_{(r)}| < \infty$ . Definamos  $M_n = X_{(n)} = \max(X_1, X_2, \dots, X_n)$ . Mostre que

$$E(M_n) = E(M_{n-1}) + \int_0^\infty F^{n-1}(x)[1 - F(x)] dx, \quad n = 2, 3, \dots$$

Encontre  $E(M_n)$  em cada uma das seguintes situações:

- $X_k$  tem como função de distribuição comum  $F(x) = 1 - e^{-x^\beta}$ , se  $x \geq 0$ .
  - $X_k$  tem como função de distribuição comum  $F(x) = x$ , se  $0 < x < 1$ .
11. Provar que, qualquer seja a amostra aleatória  $X_1, X_2, \dots, X_n$  sempre cumpre-se que  $X_{(1)} \leq \bar{X} \leq X_{(n)}$ .
12. Demonstrar o Teorema 2.5.
13. Demonstrar o Teorema 2.9.

### Exercícios da Seção 2.3

- Demonstre o Corolário 2.17.
- Demonstre o Corolário 2.18.
- Seja  $X_1, \dots, X_n$  uma amostra aleatória *Poisson*( $\theta$ ). Encontre  $\text{Var}(S^2)$  e compare-a com  $\text{Var}(\bar{X})$ . Observe que  $E(\bar{X}) = \theta = E(S^2)$ .
- Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória da função de distribuição  $F$  e seja  $F_n^*(x)$  a função de distribuição amostral. Encontre  $\text{Cov}[F_n^*(x), F_n^*(y)]$  para números reais fixos  $x, y$ .
- Seja  $F_n^*$  a função de distribuição empírica de uma amostra aleatória com função de distribuição teórica  $F$ . Prove que

$$P \left\{ |F_n^*(x) - F(x)| \geq \frac{\epsilon}{2\sqrt{n}} \right\} \leq \frac{1}{\epsilon^2} \quad \forall \epsilon > 0.$$

- Sejam  $X_1, X_2, \dots, X_n$   $n$  observações independentes da variável aleatória  $X$ . Encontre a distribuição amostral de  $\bar{X}$ , a média amostral, se:
  - $X \sim P(\theta)$ ;
  - $X \sim \text{Cauchy}(1, 0)$ ;
  - $X \sim \chi^2(m)$ .
- Seja  $X_1, \dots, X_n$  uma amostra aleatória *Poisson*( $\theta$ ). Encontre  $\text{Var}(S^2)$  e compare-a com  $\text{Var}(\bar{X})$ . Observe que  $E(\bar{X}) = \theta = E(S^2)$ .
- Demonstre o Teorema 2.23. [**Dica:** para quaisquer reais  $\mu$  e  $\sigma > 0$ , encontre a função de densidade de  $(U_{(r)} - \mu)/\sigma$  e mostre que as variáveis padronizadas de  $U_{(r)}$ ,  $(U_{(r)} - \mu)/\sigma$ , são assintoticamente  $N(0, 1)$  sob as condições do teorema.]
- Provar que o momentos amostral central  $b_1$  é sempre zero.