

# Métodos Quantitativos e Qualitativos (Análise Estatística) - Parte 2 - Probabilidade, Regressão e Séries

Valderio A. Reisen e Bartolomeu Zamprogno

Departamento de Estatística - UFES

JANEIRO 2018

Valdério Anselmo Reisen, PhD

Departamento de Estatística e  
PPGEA - UFES,  
CentraleSupélec-Paris

Bartolomeu Zamprogno, Dr (PPGEA)

Departamento de Estatística -  
UFES



Curso de Extensão em  
Tecnologias Ambientais



# Noções básicas de probabilidade e Distribuições binomial e normal

Se medirmos a corrente em um fio fino de cobre, estaremos conduzindo um **experimento**.

**Fato:** Se repetirmos o experimento acima diversas vezes, observaremos que os resultados diferem levemente de uma repetição para outra.

Exemplos de variáveis que influenciam no experimento acima:

- variações na temperatura ambiente no momento da realização;
- variações nos equipamentos utilizados para realizar a medição;
- impurezas na composição química do fio, se a medição é realizada em diversas localidades;
- impulsos na fonte da corrente;
- entre uma infinidade de outros fatores.

Não importa quão cuidadosamente tenha sido conduzido o experimento, sempre existem variáveis de perturbação (ou ruído) que não são controladas. Isso provoca aleatoriedade dos resultados obtidos em diferentes realizações do experimento.

## Definição

*Dizemos que um experimento é aleatório se, mesmo quando repetido sob condições idênticas, não é possível prever com absoluta certeza o seu resultado. Frequentemente, experimentos aleatórios são denotados pela letra  $E$ .*

# Espaços amostrais

Em geral, não sabemos o resultado de um experimento aleatório.

Por exemplo:

- ❶  $E_1 =$  “uma peça é fabricada em uma linha de produção e, depois de inspecionada, é classificada como ‘defeituosa’ ( $D$ ) ou ‘não defeituosa’ ( $N$ )”;
- ❷  $E_2 =$  “o número de ligações que chega em determinado dia a um *call center* é observado”;
- ❸  $E_3 =$  “o tempo em minutos necessário para realizar uma reação química é observado”.

Embora não saibamos o resultado que um experimento fornecerá, devemos poder listar todos os seus possíveis resultados.

## Definição

*O CONJUNTO de TODOS os possíveis resultados de um experimento aleatório é denominado **espaço amostral** do experimento.*

*Frequentemente, o espaço amostral é denotado pela letra  $\Omega$ .*

Nos exemplos anteriores os espaços amostrais seriam:

- ①  $\Omega_{E_1} = \{D, N\};$
- ②  $\Omega_{E_2} = \{0, 1, 2 \dots\};$
- ③  $\Omega_{E_3} = \{\omega : \omega > 0\} = (0, \infty).$

Suponha que os exemplos anteriores sejam repetidos 2 vezes. Os espaços amostrais agora passam a ser:

- ①  $\Omega_1^* = \Omega_{E_1} \times \Omega_{E_1} = \{DD, DN, ND, NN\};$
- ②  $\Omega_2^* = \Omega_{E_2} \times \Omega_{E_2} = \{(0, 0), (0, 1), (0, 2), \dots, (1, 0), (1, 1), (1, 2), \dots\};$
- ③  $\Omega_3^* = \Omega_{E_3} \times \Omega_{E_3} = \{(\omega_1, \omega_2) : \omega_1, \omega_2 > 0\}.$

Em geral, quando conduzimos um experimento, não estamos interessados apenas em um resultado em particular, mas sim em uma coleção destes. Isto é, em um **subconjunto** do espaço amostral.

## Definição

*Um **evento** é qualquer subconjunto do espaço amostral de um experimento aleatório. Frequentemente, eventos são denotados por letras iniciais do alfabeto maiúsculas:  $A, B, C, \dots$*

Exemplos de eventos dos espaços amostrais  $\Omega_1^*$ ,  $\Omega_2^*$  e  $\Omega_3^*$ :

- ①  $A_1 = \text{"Pelo menos uma peça defeituosa"} = \{DD, DN, ND\};$
- ②  $A_2 = \text{"Receber 4 ligações no total"} \\ = \{(0, 4), (1, 3), (2, 2), (3, 1), (4, 0)\} = \{(\omega_1, \omega_2) \in \Omega_2^* : \omega_1 + \omega_2 = 4\};$
- ③  $A_3 = \text{"no total as duas reações terminarem em 5 minutos ou mais"} \\ = \{(\omega_1, \omega_2) \in \Omega_3^* : \omega_1 + \omega_2 \geq 5\}$

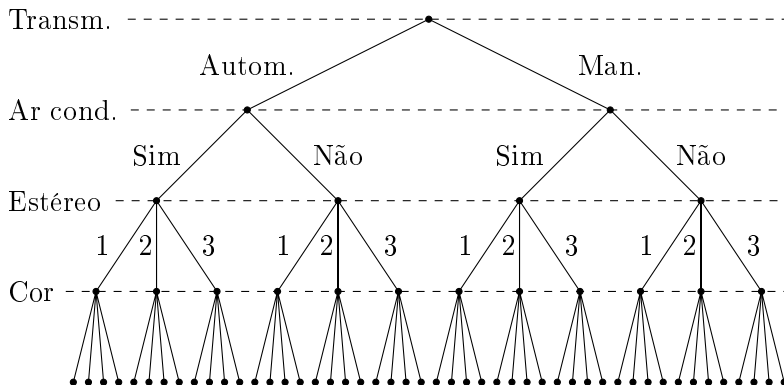
# Exercícios Espaço Amostral

- 1 Se lançarmos uma moeda e observarmos a face para cima, o resultado será? Cara ou Coroa.
- 2 Se lançarmos um dado e observarmos o número obtido na face para cima, o resultado será um dos números do conjunto  $\{1,2,3,4,5,6\}$ .
- 3 Ao fabricarmos parafusos com uma máquina, alguns poderão ter defeito na fabricação. Assim ao escolhermos um parafuso aleatoriamente ele será elemento do conjunto  $\{\text{defeituoso}, \text{não-defeituoso}\}$ .
- 4 De um baralho comum de 52 cartas extraí-se uma carta ao acaso. Descreva o espaço amostral: (a) não levando em conta os naipes; (b) levando-se em conta os naipes.



Um fabricante fornece automóveis que podem ser equipados com os opcionais escolhidos pelo consumidor. Cada veículo pode ser escolhido:

- com ou sem transmissão automática;
- com ou sem ar condicionado;
- com um dos três tipos de sistema estéreo;
- com uma das quatro cores exteriores.



# Interpretação de probabilidade

A princípio consideremos um espaço amostral finito  $\Omega = \{\omega_1, \dots, \omega_n\}$  (ou infinito contável  $\Omega = \{\omega_1, \omega_2, \dots\}$ ).

Existem duas interpretações básicas de probabilidade:

- subjetivista: quando a probabilidade de um evento  $A$ ,  $P(A)$ , representa o “grau de crença” que se tem com respeito a ocorrência de  $A$ . Nesse ponto de vista, indivíduos diferentes podem atribuir probabilidades diferentes para o mesmo evento  $A$ ;
- frequentista: quando a probabilidade atribuída a um evento  $A$ ,  $P(A)$ , é interpretada como o limite da frequência relativa desse evento em  $n$  repetições idênticas do experimento. Neste caso, o limite é avaliado quando  $n \rightarrow \infty$ .

Seja  $A$  = “chover no dia de finados”. Se atribuírmos ao evento  $A$  a probabilidade  $P(A) = 0.8$ . Temos as seguintes interpretações:

- subjetivista: segundo a minha experiência, o grau de crença que eu tenho de que choverá no dia de finados é de 0.8 (ou 80%);
- frequentista: se fosse possível observar indefinidamente o dia de finados ano a ano, a proporção de anos em que o dia de finados é chuvoso seria próxima de 0.8 e ficaria mais próxima conforme aumentássemos o período de observação.

# Axiomas de probabilidade

## Definição

*Uma medida de probabilidade é qualquer função de eventos  $P(\cdot)$  que satisfaça as seguintes propriedades:*

- ❶  $P(\Omega) = 1$ ;
- ❷  $P(A) \geq 0$ , se  $A \subset \Omega$  é evento;
- ❸  $P(A \cup B) = P(A) + P(B)$ , se  $A, B \subset \Omega$  são eventos mutuamente excludentes, isto é, se  $A \cap B = \emptyset$ .

## Comentário

*Os axiomas de probabilidade não determinam probabilidades. As probabilidades devem ser atribuídas com base no nosso conhecimento do fenómeno em estudo e devem sempre ser estabelecidas de forma que obedeçam os axiomas acima.*

# Propriedades

Sejam  $\Omega$  espaço amostral e  $A, B \subset \Omega$  eventos. Uma função de probabilidade definida de acordo com os axiomas de probabilidade satisfaz as seguintes propriedades:

- ①  $P(A^c) = 1 - P(A)$ ;
- ②  $P(\emptyset) = 0$ ;
- ③  $P(A) \leq P(B)$ , se  $A \subset B$ ;
- ④  $P(B - A) = P(B) - P(A \cap B)$ , onde  $B - A = B \cap A^c$ ;
- ⑤  $P(B - A) = P(B) - P(A)$ , se  $A \subset B$ ;
- ⑥  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

# Exemplo

Discos de policarbonato plástico são analisados com relação a resistência a arranhões e choque. Os resultados de 100 discos são resumidos a seguir:

Res. a arranhão	Res. a choque		Total
	Alta	Baixa	
Alta	80	9	89
Baixa	6	5	11
Total	86	14	100

Suponha que um disco será selecionado ao acaso. Sejam  $A$  = “resist. a arranhões alta” e  $B$  = “resist. a choques alta”. Qual a probabilidade que você atribuiria para os eventos:

- $P(A \cap B)$ ;
- $P(A \cup B)$ ;
- $P(B - A)$ ;

## Exemplo

Discos de policarbonato plástico são analisados com relação a resistência a arranhões e choque. Os resultados de 100 discos são resumidos a seguir:

Res. a arranhão	Res. a choque		Total
	Alta	Baixa	
Alta	80	9	89
Baixa	6	5	11
Total	86	14	100

Suponha que um disco será selecionado ao acaso. Sejam  $A$  = “resist. a arranhões alta” e  $B$  = “resist. a choques alta”. Qual a probabilidade que você atribuiria para os eventos:

- $P(A \cap B) = 0.8$ ;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.89 + 0.86 - 0.80 = 0.95$ ;
- $P(B - A) = P(B) - P(A \cap B) = 0.86 - 0.8 = 0.06$ .

# Introdução variável aleatória

Consideremos os experimentos:

- ①  $E_1 =$  “duas peças de uma linha de produção são inspecionadas e classificadas como ‘defeituosa’ ( $D$ ) ou ‘não defeituosa’ ( $N$ )”;
- ②  $E_2 =$  “uma linha de produção é observada até a produção da primeira peça defeituosa”;
- ③  $E_3 =$  “um ponto em um círculo de raio unitário é escolhido ao acaso”.



# Introdução variável aleatória

Consideremos os experimentos:

- ①  $E_1 =$  “duas peças de uma linha de produção são inspecionadas e classificadas como ‘defeituosa’ ( $D$ ) ou ‘não defeituosa’ ( $N$ )”;
- ②  $E_2 =$  “uma linha de produção é observada até a produção da primeira peça defeituosa”;
- ③  $E_3 =$  “um ponto em um círculo de raio unitário é escolhido ao acaso”.

Defina:

- ① em  $E_1$ ,  $X_1 =$  “número de peças defeituosas observadas”;
- ② em  $E_2$ ,  $X_2 =$  “número de peças produzidas até a interrupção”;
- ③ em  $E_3$ ,  $X_3 =$  “distância do ponto sorteado ao centro do círculo”.

# Introdução variável aleatória

Consideremos os experimentos:

- 1  $E_1$  = “duas peças de uma linha de produção são inspecionadas e classificadas como ‘defeituosa’ ( $D$ ) ou ‘não defeituosa’ ( $N$ )”;
- 2  $E_2$  = “uma linha de produção é observada até a produção da primeira peça defeituosa”;
- 3  $E_3$  = “um ponto em um círculo de raio unitário é escolhido ao acaso”.

Defina:

- 1 em  $E_1$ ,  $X_1$  = “número de peças defeituosas observadas”;
- 2 em  $E_2$ ,  $X_2$  = “número de peças produzidas até a interrupção”;
- 3 em  $E_3$ ,  $X_3$  = “distância do ponto sorteado ao centro do círculo”.

Observe que  $X_1$ ,  $X_2$  e  $X_3$  são funções (numéricas) dos resultados (não necessariamente numéricos) do experimento aleatório ao qual estão associadas. Neste caso, dizemos que  $X_1$ ,  $X_2$  e  $X_3$  são **variáveis aleatórias**.

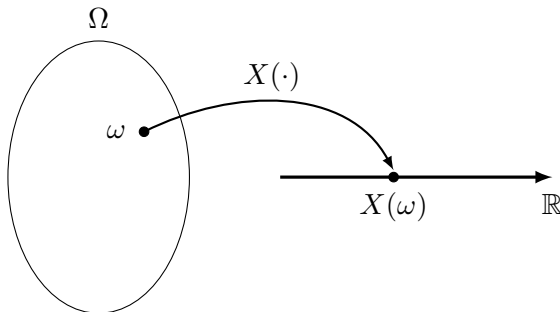
# Definição de variável aleatória

## Definição

Sejam  $\Omega$  espaço amostral de um experimento aleatório  $E$  e  $\omega \in \Omega$  um de seus resultados. Se

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

dizemos que  $X$  é variável aleatória (v.a.).



## Definição

*O conjunto de todos os possíveis valores que uma variável aleatória  $X$  pode assumir é denominado **imagem** de  $X$  e denotado por  $Im(X)$ .*

Temos que

- ❶  $Im(X_1) = \{0, 1, 2\};$
- ❷  $Im(X_2) = \{1, 2, 3, \dots\};$
- ❸  $Im(X_3) = \{x : 0 \leq x \leq 1\} = [0, 1].$

# Tipos de variáveis aleatórias

## Definição

*O conjunto de todos os possíveis valores que uma variável aleatória  $X$  pode assumir é denominado **imagem** de  $X$  e denotado por  $Im(X)$ .*

Temos que

- ①  $Im(X_1) = \{0, 1, 2\}$ ;
- ②  $Im(X_2) = \{1, 2, 3, \dots\}$ ;
- ③  $Im(X_3) = \{x : 0 \leq x \leq 1\} = [0, 1]$ .

Há diferenças entre  $Im(X_1)$ ,  $Im(X_2)$  e  $Im(X_3)$ :

- ①  $Im(X_1)$  é um conjunto finito;
- ②  $Im(X_2)$  é um conjunto infinito contável (ou enumerável);
- ③  $Im(X_3)$  é um conjunto infinito não contável (ou não enumerável).

## Definição

Dizemos que uma variável aleatória  $X$  é **discreta** se  $Im(X)$  é um conjunto finito ou infinito enumerável.

## Definição

Dizemos que uma variável aleatória  $X$  é **contínua** se  $Im(X)$  é um conjunto não-enumerável.

Portanto,  $X_1$  e  $X_2$  são v.a. discretas, enquanto  $X_3$  é v.a. contínua.

- A concentração de um poluente  $\rightarrow$  **contínua**;
- Incomodo à algum tipo de poluição  $\rightarrow$  **discreta**;
- Número de crianças internadas decorrentes de poluição  $\rightarrow$  **discreta**;
- Nível de precipitação  $\rightarrow$  **contínua**;
- O número de moléculas em uma amostra de gás  $\rightarrow$  **discreta**;
- Volume de água perdido por dia, num sistema de abastecimento  $\rightarrow$  **contínua**;
- O volume de gasolina que é perdido por evaporação, durante o enchimento de um tanque de gasolina  $\rightarrow$  **contínua**;
- O tempo que um projétil gasta para retornar à Terra  $\rightarrow$  **contínua**;
- O número de vezes que um transistor em uma memória de computador muda de estado em uma operação  $\rightarrow$  **discreta**;
- A concentração de saída de um reator  $\rightarrow$  **contínua**;
- A corrente em um circuito elétrico  $\rightarrow$  **contínua**.

- 1 O resultado  $\omega$  é desconhecido *a priori*;
- 2 O valor resultante da v.a.  $X(\omega)$  também o é;
- 3 Devemos tratar probabilisticamente  $X$ ;
- 4 Abordagens diferentes nos casos discreto e contínuo.



# Variável aleatória discreta: o Modelo binomial

## Definição

*Considere um experimento  $E$  o qual pode ocasionar apenas dois resultados:  $s$  = “sucesso”; e  $f$  = “fracasso”. O espaço amostral desse experimento é:*

$$\Omega_E = \{s, f\}.$$

*Denominamos  $E$  de **experimento de Bernoulli**.*

**Exemplos:** Observe alguns experimentos de Bernoulli a seguir:

- $E$  = “observar uma amostra de ar e verificar se ela possui alguma molécula rara”;
- $E$  = “observar um bit transmitido através de um canal digital e verificar se ele foi recebido com erro”;
- $E$  = “em uma questão de múltipla escolha um candidato tenta adivinhar a resposta correta”.

Seja  $E$  o experimento que consiste em  $n$  repetições **independentes** de um experimento de Bernoulli nos quais a probabilidade de sucesso  $P(\{s\}) := p$  permanece **inalterada**. Defina  $X =$  “o número de sucessos observados nas  $n$  réplicas”.

É evidente que  $Im(X) = \{0, 1, \dots, n\}$ . Vamos encontrar a f.p. de  $X$ , ou seja, desejamos obter  $P_X(x) = P(X = x)$ ,  $x = 0, 1, \dots, n$ .

- Fixe algum  $x \in Im(X)$ ;
- Em  $n$  repetições,  $x$  sucessos (e  $n - x$  fracassos) ocorrem com probabilidade  $p^x(1 - p)^{n-x}$ ;
- Não importa a ordem em que os  $x$  sucessos e os  $n - x$  fracassos ocorrem, a probabilidade acima permanece inalterada. Por exemplo:

$$P(\{\overbrace{s \dots ss}^x \overbrace{ff \dots ff}^{n-x}\}) = P(\{\overbrace{s \dots s}^{x-1} fs \overbrace{ff \dots ff}^{n-x-1}\}) = p^x(1 - p)^{n-x};$$

A pergunta que surge é: de quantas maneiras diferentes podemos ter  $x$  sucessos em  $n$  repetições?

Posição escolhida	1 <sup>a</sup>	2 <sup>a</sup>	...	$x$ -ésima
Posições disponíveis	$n$	$n - 1$	...	$n - x + 1$

A princípio poderíamos imaginar **erroneamente** que temos

$$n(n-1)(n-2)\dots(n-x+1) = \frac{n!}{(n-x)!}$$

maneiras de ter  $x$  sucessos em  $n$  repetições. Na equação acima  $a! = a(a-1)\dots 2 \cdot 1$ . Define-se  $0! = 1$ . Entretanto, existem redundâncias nestas combinações. Por exemplo, tome  $n = 4$  e  $x = 2$ . Adotando o raciocínio acima teríamos as seguintes combinações de posições

$$\begin{array}{cccc} (1,2) & (2,1) & (3,1) & (4,1) \\ (1,3) & (2,3) & (3,2) & (4,2) \\ (1,4) & (2,4) & (3,4) & (4,3) \end{array} \Rightarrow \frac{n!}{(n-x)!} = \frac{4!}{2!} = 12.$$

A pergunta que surge é: de quantas maneiras diferentes podemos ter  $x$  sucessos em  $n$  repetições?

Posição escolhida	1 <sup>a</sup>	2 <sup>a</sup>	...	$x$ -ésima
Posições disponíveis	$n$	$n - 1$	...	$n - x + 1$

A princípio poderíamos imaginar **erroneamente** que temos

$$n(n-1)(n-2)\dots(n-x+1) = \frac{n!}{(n-x)!}$$

maneiras de ter  $x$  sucessos em  $n$  repetições. Na equação acima  $a! = a(a-1)\dots 2 \cdot 1$ . Define-se  $0! = 1$ . Entretanto, existem redundâncias nestas combinações. Por exemplo, tome  $n = 4$  e  $x = 2$ . Adotando o raciocínio acima teríamos as seguintes combinações de posições

$$\begin{array}{cccc} (1,2)^1 & (2,1)^1 & (3,1)^2 & (4,1)^3 \\ (1,3)^2 & (2,3)^4 & (3,2)^4 & (4,2)^5 \\ (1,4)^3 & (2,4)^5 & (3,4)^6 & (4,3)^6 \end{array} \Rightarrow \frac{n!}{(n-x)!} = \frac{4!}{2!} = 12.$$

Logo, devemos dividir  $\frac{n!}{(n-x)!}$  pela quantidade de permutações que podemos fazer com as  $x$  posições escolhidas, a saber:  $x!$ . Portanto, observamos que obtem-se  $x$  sucessos em  $n$  repetições de

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

maneiras possíveis.

Dessa forma,

$$\begin{aligned} P_X(x) = P(X = x) &= \overbrace{p^x(1-p)^{n-x} + \cdots + p^x(1-p)^{n-x}}^{\binom{n}{x} \text{ vezes}} \\ &= \binom{n}{x} p^x (1-p)^{n-x}, \end{aligned}$$

onde  $x = 0, 1, \dots, n$ .

## Definição

Seja  $X$  v.a. discreta com conjunto imagem  $Im(X) = \{0, 1, \dots, n\}$  e f.p. dada por

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

Dizemos que  $X$  tem distribuição binomial com  $n$  repetições e probabilidade de sucesso  $p$ . Notação:  $X \sim B(n, p)$ .

Se  $X \sim B(n, p)$ , então sua esperança e sua variância são dadas, respectivamente, por

$$\mu = E(X) = np$$

e

$$\sigma^2 = Var(X) = np(1-p).$$

## Exemplo

Um estudante faz um teste de múltipla escolha com 25 questões, cada uma com 4 alternativas, apenas chutando as respostas. Qual a probabilidade de o estudante acertar mais do que 20 questões? Quantas questões ele acerta em média e com qual variância?

Estamos interessados em  $X =$  “número de questões certas das 25 do teste”. Temos que  $X \sim B(25, \frac{1}{4})$ . Logo,

$$P(X > 20) = \sum_{x=21}^{25} P(X = x) = \sum_{x=21}^{25} \binom{25}{x} (0.25)^x (0.75)^{25-x} = 9 \cdot 10^{-10},$$

$$E(X) = np = 25 \cdot 0.25 = 6.25$$

e

$$Var(X) = np(1 - p) = 25 \cdot 0.25 \cdot 0.75 = 4.6875.$$

Estudo de Melo (2015): O quanto o(a) Sr(a) se sente incomodado com a poluição do ar?

```
#Fazer leitura dos dados no formato numérico
```

```
dados<-read.table("clipboard",h=T)
```

```
table(dados)
```

```
questao_A3<-table(dados)
```

```
names(questao_A3)<-
```

```
c("nada","pouco","moderado","muito","extremamente","não respondeu",'
```

```
> round(prop.table(questao_A3),2)
```

```
nada pouco moderado muito extremamente não respondeu não sabe
```

```
0.03    0.09    0.16    0.57    0.14            0.01            0.01
```

Logo o percentual de incomodados são  $0.57 + 0.14 = 0.71$  ou 71%.



Dessa forma, se observamos 50 pessoas na região do estudo, e supondo a independência entre as pessoas da região, qual a probabilidade dentre as 50 pessoas, termos 30 pessoas incomodadas?

```
> dbinom(30, 50, 0.71)
0.02877402
```

Dessa forma, se observamos 50 pessoas na região do estudo, e supondo a independência entre as pessoas da região, qual a probabilidade dentre as 50 pessoas, termos no máximo 30 pessoas incomodadas?

- `dbinom(0, 50, 0.71)+dbinom(1, 50, 0.71)+dbinom(2, 50, 0.71)+dbinom(3, 50, 0.71)+...+dbinom(30, 50, 0.71)`

ou use

- `> pbinom(30, 50, 0.71)`  
0.06257097

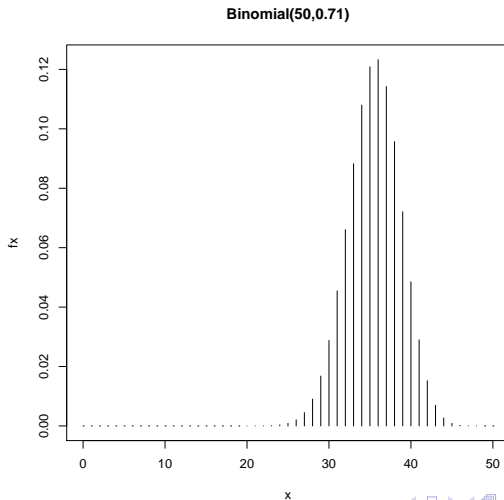
Obs: **pbinom** calcula a probabilidade no ponto e **rbinom** calcula a probabilidade acumulada até o ponto.

Construção do gráfico da binomial do exemplo acima.

```
> x<-0:50
```

```
> fx <- dbinom(x, 50, 0.71)
```

```
> plot(x, fx, type='h',main="Binomial(50,0.71)")
```



# Modelo de Poisson - Uma outra v.a. discreta

Suponha que, para qualquer instante  $t > 0$ , denotamos a quantidade de ocorrências de um determinado evento até  $t$  por  $X_t$ . Se:

- o tempo entre ocorrências do evento tem distribuição **exponencial** com parâmetro  $\lambda$ ; e
- as ocorrências acontecem independentemente.

Então, dizemos que  $X_t$ ,  $t > 0$ , é um processo de Poisson com taxa de ocorrências  $\lambda$ . Isto porque o número de ocorrências em qualquer intervalo de comprimento  $t$  tem distribuição **Poisson** com parâmetro  $\lambda t$ .

## Definição

Suponha  $X$  v.a. discreta com conjunto imagem  $Im(X) = \{0, 1, \dots\}$ . Admita que a f.p. de  $X$  seja

$$P_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, \dots,$$

onde  $\lambda > 0$ . Dizemos que  $X$  tem v.a. Poisson com parâmetro  $\lambda$ . Notação:  $X \sim \text{Poisson}(\lambda)$ .

É possível mostrar que se  $X \sim \text{Poisson}(\lambda)$ , então  $\mu = E(X) = \lambda$  e  $\sigma^2 = \text{Var}(X) = \lambda$ .

## Comentário

### *Comentários:*

- *em um processo de Poisson com taxa de ocorrências  $\lambda$ , o número de ocorrências em qualquer intervalo de comprimento 1 tem distribuição de Poisson com parâmetro  $\lambda$ ;*
- *a distribuição de Poisson é muito utilizada para modelar o número de ocorrências de eventos, não somente no tempo, mas também por unidades de medida, de área, entre outros.*

## Exemplo

Em determinada cidade o número de casos de dengue tem distribuição de Poisson com 100 ocorrências por  $km^2$  em média. Qual a probabilidade de se observar menos que 3 ocorrências em uma região de  $10000m^2$ ?

Primeiramente, temos que

$10000m^2 = 100m \cdot 100m = 0.1km \cdot 0.1km = 0.01km^2$ . Logo, o número de ocorrências  $X \sim Poisson(\lambda t)$ , com  $\lambda = 100$  e  $t = 0.01$ . Assim,

$$\begin{aligned} P(X < 3) &= \sum_{x=0}^2 P_X(x) = \sum_{x=0}^2 \frac{e^{-(\lambda t)} (\lambda t)^x}{x!} = \sum_{x=0}^2 \frac{e^{-1} (1)^x}{x!} \\ &= e^{-1} \left( 1 + 1 + \frac{1^2}{2} \right) = e^{-0.1} (1 + 1 + 0.5) \\ &= 2.5e^{-1} \approx 0.92. \end{aligned}$$

# Exercícios Binomial

- Qual é a probabilidade de 3 caras em 5 lançamentos de uma moeda honesta? R: 31,25%
- Um engenheiro de inspeção extrai uma amostra de 15 itens aleatoriamente de um processo de fabricação, onde é sabido que esse processo produz 85% de itens aceitáveis. Qual a probabilidade de que 10 dos itens extraídos sejam aceitáveis? R: 4,5%
- Um inspetor de qualidade extrai uma amostra de 10 tubos aleatoriamente de carga muito grande de tubos que se sabe que contem 20% de tubos defeituosos. Qual é a probabilidade de que não mais do que 2 tubos extraídos sejam defeituosos? R: 6,78%

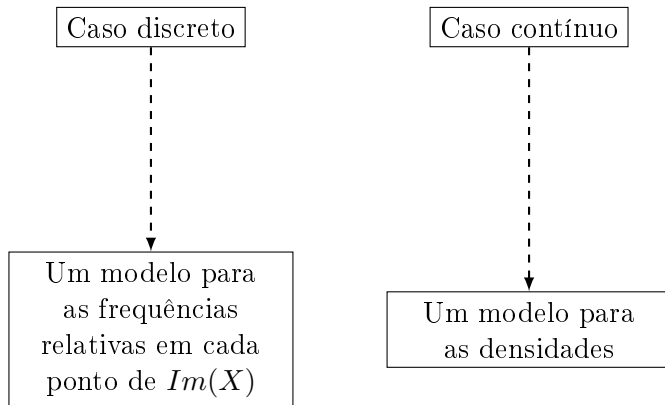
# Exercícios Binomial

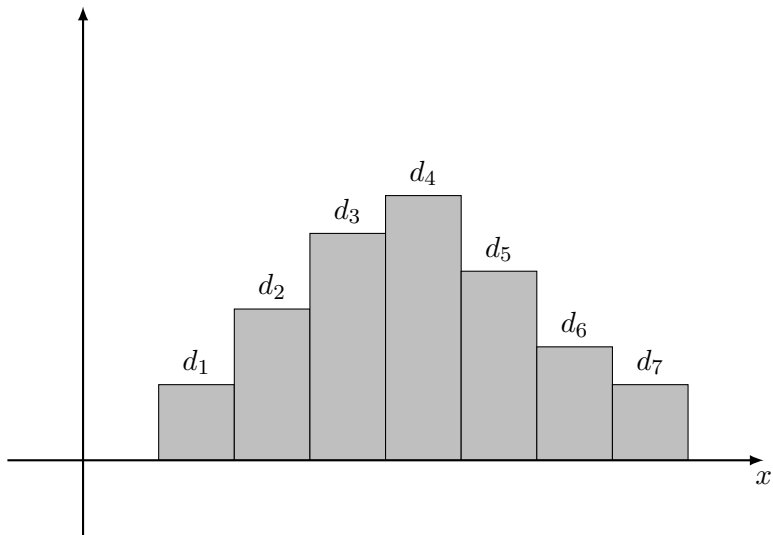
- Acredita-se que 20% dos moradores das proximidades de uma grande siderúrgica têm alergia aos poluentes lançados ao ar. Admitindo que este percentual de alérgicos seja real (correto), calcule a probabilidade de que pelo menos 4 moradores tenham alergia entre 13 selecionados ao acaso. R: 25,26%
- Três em cada quatro alunos de uma universidade fizeram cursinho antes de prestar vestibular. Se 16 alunos são selecionados ao acaso, qual é a probabilidade de que:
  - a) Pelo menos 12 tenham feito cursinho? R: 63,02%
  - b) No máximo 13 tenham feito cursinho? R: 80,29%
  - c) Exatamente 12 tenham feito cursinho? R: 22,51%




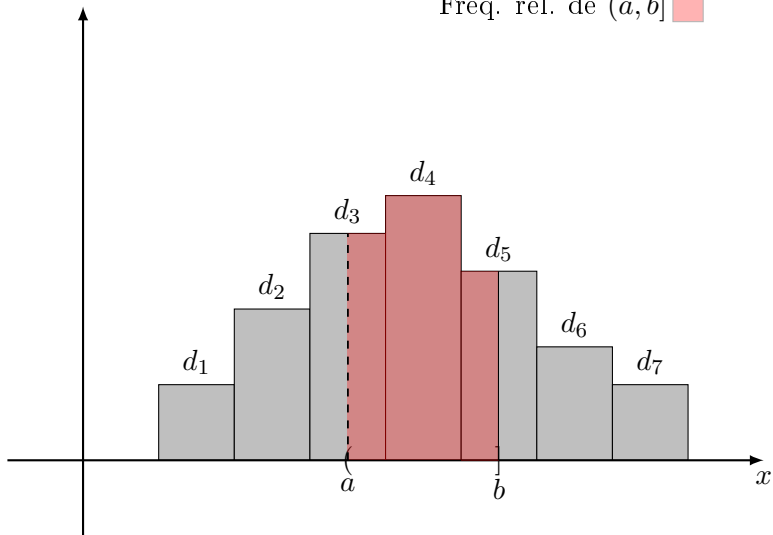
# Variáveis aleatórias contínuas


Suponha que  $X$  é uma v.a. contínua. Devido a natureza não contável de  $Im(X)$  neste caso, não devemos estabelecer probabilidades “ponto-a-ponto” conforme no caso discreto.

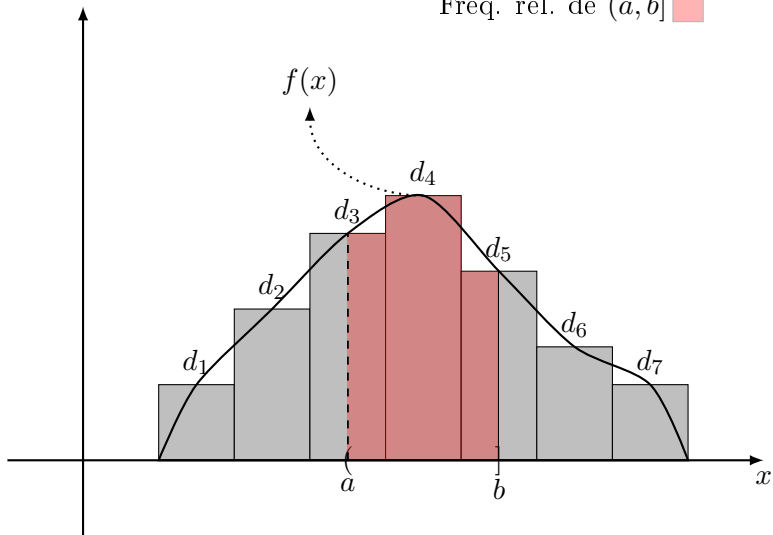


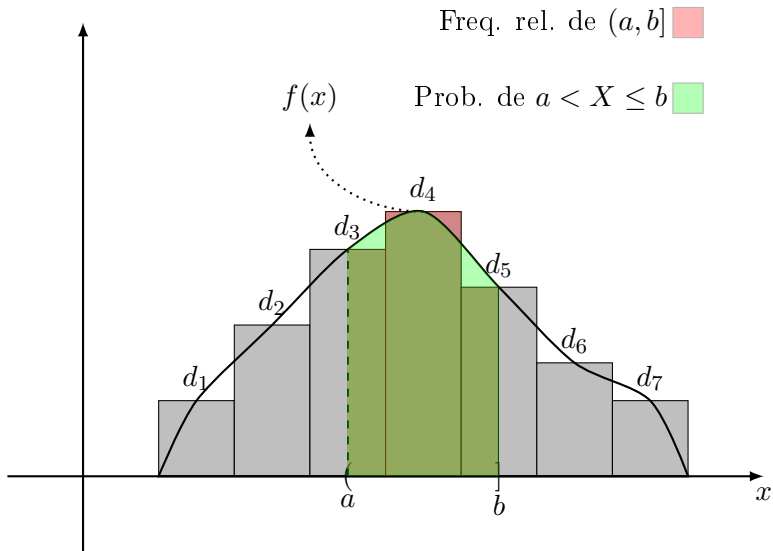


Freq. rel. de  $(a, b]$  



Freq. rel. de  $(a, b]$  





# Função densidade de probabilidade

## Definição

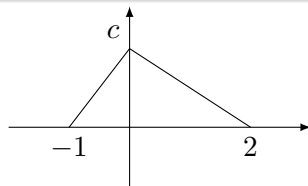
Seja  $f(x)$  uma função, tal que:

- $f(x) \geq 0$ , para todo  $x \in \mathbb{R}$ ;
- a área entre  $f(x)$  e o eixo horizontal é igual a 1.

Dizemos que  $f(x)$  é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.

**Exemplo:** Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1-x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$



Para qual valor de  $c$ ,  $f(x)$  é densidade? Primeiramente,  $c > 0$ . A área abaixo de  $f(x)$  por sua vez é  $\frac{c}{2} + c = 1 \rightarrow c = \frac{2}{3}$ . Logo, devemos ter  $c = \frac{2}{3}$ .

# Como obter as probabilidades

Dada uma v.a. contínua  $X$  com função densidade de probabilidade  $f(x)$ . A probabilidade  $P(a < X \leq b)$  é dada pela área entre a f.d.p.  $f(x)$  e o eixo horizontal compreendida no intervalo  $(a, b]$ .

Por este motivo

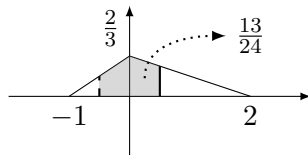
$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b)$$

e, além disso,

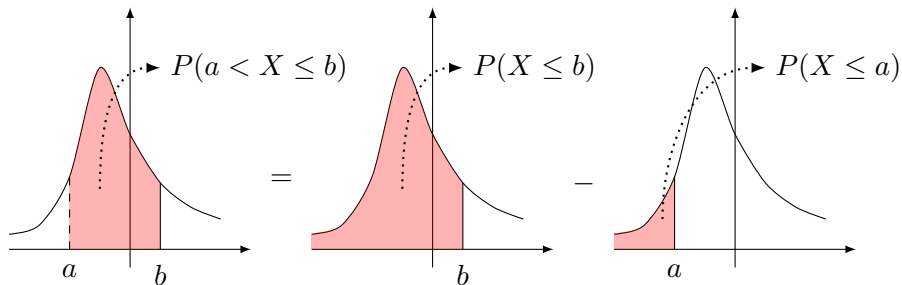
$$P(X = x) = 0, \text{ para todo } x \in \mathbb{R}.$$

**Exercício:** Mostre que, se  $X$  tem a f.d.p. do exemplo anterior, então

$$P(-0.5 < X < 0.5) = \frac{13}{24}.$$



Esquemáticamente temos a seguinte figura:



Logo,

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



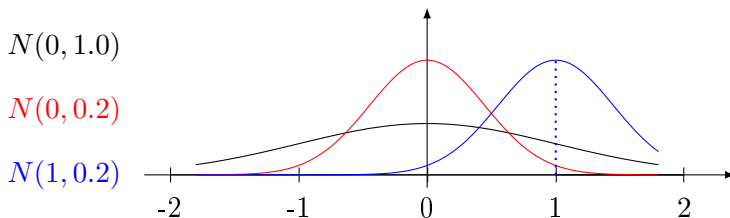
# Modelo normal (ou gaussiano)

## Definição

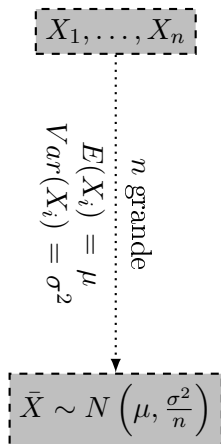
Dizemos que  $X$  tem distribuição normal com média  $\mu$  e variância  $\sigma^2$  se sua f.d.p. é dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Notação:  $X \sim N(\mu, \sigma^2)$ .



Principal importância: grande aplicabilidade devido ao Teorema Central do Limite (TCL).



- Se fosse possível observar  $R$  amostras de tamanho  $n$ ;
- Estas gerariam  $R$  médias amostrais;
- Se  $n$  for muito grande o histograma das  $R$  médias amostrais teria o formato aproximado de uma normal;
- Os parâmetros seriam a média teórica,  $\mu$ , e variância teórica dividida pelo tamanho amostral,  $\frac{\sigma^2}{n}$ .

O TCL também nos fornece base teórica para afirmar que, em algumas situações, somas de **muitas** v.a.'s têm distribuição aproximadamente normal.

**Exemplo:** O desvio do comprimento de uma peça usinada do valor em sua especificação pode ser pensado como soma de um grande número de efeitos:

- pulsos na temperatura e na umidade;
- vibrações;
- variações na velocidade rotacional;
- variações nas inúmeras características das matérias-primas.

Se os efeitos forem independentes, então se pode mostrar que o desvio total tem distribuição aproximadamente normal.

# Padronização

É possível mostrar que, se  $X \sim N(\mu, \sigma^2)$ , então a esperança e a variância de  $X$  são, respectivamente,

$$E(X) = \mu \text{ e } Var(X) = \sigma^2.$$

## Definição

*Seja  $Z \sim N(0, 1)$ . Então dizemos que  $Z$  tem distribuição normal padrão.*

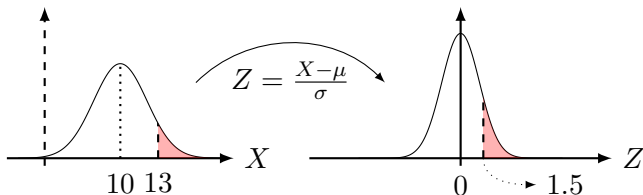
## Teorema

*Se  $X \sim N(\mu, \sigma^2)$ . Então  $Z = \frac{X - \mu}{\sigma}$  tem distribuição normal padrão.*

**Exemplo:** Seja  $X$  a corrente em um pedaço de fio medida em miliampéres. Suponha que  $X \sim N(10, 4)$ . Qual a probabilidade de realizarmos uma medida nesse fio que supere 13 miliampéres?

Queremos encontrar  $P(X > 13)$ .

Observe que  $P(X > 13) = P\left(\frac{X-\mu}{\sigma} > \frac{13-10}{2}\right) = P(Z > 1.5)$ .

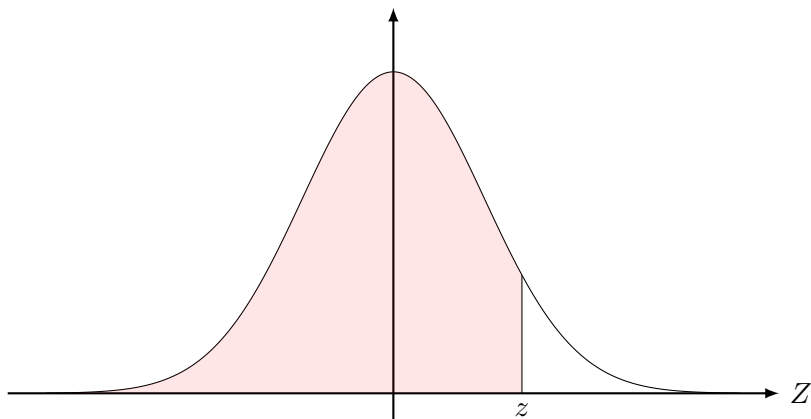


Assim:

- basta conhecer as probabilidades para uma v.a. normal padrão  $Z$ ;
- utilizando a padronização  $Z = \frac{X-\mu}{\sigma}$ , obtemos as probabilidades para normais com quaisquer parâmetros  $\mu$  e  $\sigma^2$ .

De fato, existe uma tabela com as probabilidades da normal padrão.

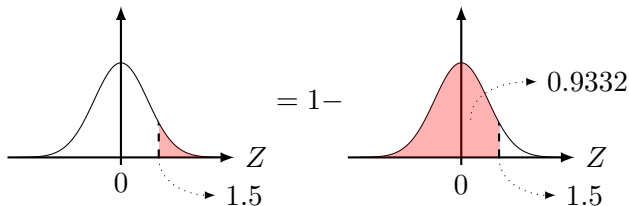
$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



Dessa forma, a probabilidade da corrente ultrapassar 13 miliampéres é

$$P(X > 13) = P(Z > 1.5) = 1 - 0.9332 = 0.0668.$$

Afinal:





# Aproximação da binomial pela normal

**Exemplo:** Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja  $Y$  = “número de peças defeituosas num total de 5000 observadas”. Percebemos que  $Y \sim B(n, p)$ , onde  $n = 5000$  e  $p = 0.05$ . Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

# Aproximação da binomial pela normal

**Exemplo:** Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja  $Y$  = “número de peças defeituosas num total de 5000 observadas”. Percebemos que  $Y \sim B(n, p)$ , onde  $n = 5000$  e  $p = 0.05$ . Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

# Aproximação da binomial pela normal

**Exemplo:** Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja  $Y$  = “número de peças defeituosas num total de 5000 observadas”. Percebemos que  $Y \sim B(n, p)$ , onde  $n = 5000$  e  $p = 0.05$ . Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

Essa tarefa pode ser facilitada percebendo que, para  $n$  suficientemente grande. O gráfico da função de probabilidades de uma v.a.  $B(n, p)$  se assemelha muito com a f.d.p. de uma v.a.  $N(\mu, \sigma^2)$ !

# Aproximação da binomial pela normal

**Exemplo:** Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja  $Y$  = “número de peças defeituosas num total de 5000 observadas”. Percebemos que  $Y \sim B(n, p)$ , onde  $n = 5000$  e  $p = 0.05$ . Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

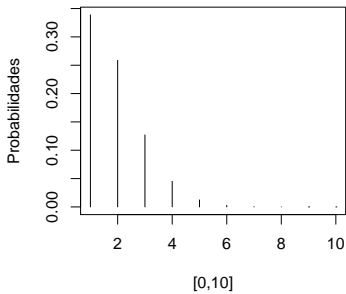
A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

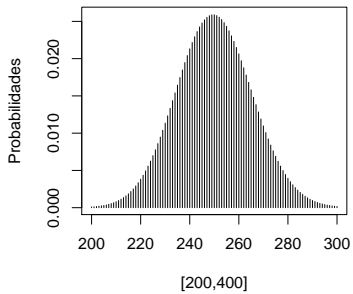
Essa tarefa pode ser facilitada percebendo que, para  $n$  suficientemente grande. O gráfico da função de probabilidades de uma v.a.  $B(n, p)$  se assemelha muito com a f.d.p. de uma v.a.  $N(\mu, \sigma^2)$ !

Mas quais valores utilizar para  $\mu$  e  $\sigma^2$ ?

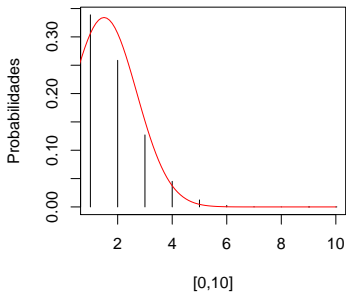
**B(30,0.05)**



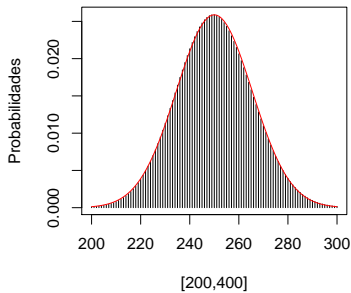
**B(5000,0.05)**



**B(30,0.05)**



**B(5000,0.05)**



$$N(np, np(1 - p)).$$

Portanto, a probabilidade do exemplo anterior pode ser aproximada da seguinte maneira:

$$P(Y < 230) \approx P(X < 230),$$

onde  $X \sim N(\mu, \sigma^2)$ , com

$$\mu = np = 5000 \cdot 0.05 = 250 \text{ e } \sigma^2 = np(1 - p) = 5000 \cdot 0.05 \cdot 0.95 = 237.5.$$

Logo,

$$P(Y < 230) \approx P(X < 230) = P\left(\frac{X - \mu}{\sigma} < \frac{230 - 250}{\sqrt{237.5}}\right) = P(Z < -1.29),$$

onde  $Z$  é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.0985$$

Portanto, a probabilidade do exemplo anterior pode ser aproximada da seguinte maneira:

$$P(Y < 230) \approx P(X < 230),$$

onde  $X \sim N(\mu, \sigma^2)$ , com

$$\mu = np = 5000 \cdot 0.05 = 250 \text{ e } \sigma^2 = np(1 - p) = 5000 \cdot 0.05 \cdot 0.95 = 237.5.$$

Logo,

$$P(Y < 230) \approx P(X < 230) = P\left(\frac{X - \mu}{\sigma} < \frac{230 - 250}{\sqrt{237.5}}\right) = P(Z < -1.29),$$

onde  $Z$  é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.0985$$

O valor verdadeiro é 0.0904.



# Regressão linear

Suponha que temos interesse em investigar o comportamento de uma variável com respeito à um conjunto de outras variáveis.

Considere os seguintes problemas:

- avaliar se as alturas dos filhos estão relacionadas com as alturas dos seus pais.
- verificar se o faturamento de uma empresa é afetado pelo número de funcionários.
- avaliar se a produção de uma máquina depende do nível de treinamento do operador.
- avaliar a relação entre o número de internações e a concentração de determinados poluentes.

Para **predizer** o valor de uma variável a partir de uma outra variável e para **descrever** a relação entre duas ou mais variáveis utiliza-se métodos de regressão.

- A análise de regressão é uma das mais importantes técnicas estatísticas e é utilizada quando há interesse em investigar o comportamento (relação não determinística) de uma variável com respeito à um conjunto de outras variáveis.
- As aplicações de regressão são numerosas e ocorrem em quase todos os campos de estudo, tais como engenharia, ciências físicas e químicas, economia, ciência biológicas, ciências sociais, etc.
- Um problema de regressão consiste em estabelecer e determinar uma função que descreva a relação entre uma variável, chamada de variável resposta ou dependente e denotada por  $Y$ , e um conjunto de variáveis observáveis, chamadas de variáveis explicativas ou covariáveis e denotadas por  $X_1, X_2, \dots, X_p$ .

- Uma vez estabelecida e determinada a relação funcional entre a variável resposta (variável dependente),  $Y$ , e as covariáveis (variáveis explicativas ou independentes),  $X_1, X_2, \dots, X_p$ , a análise de regressão pode explorar esta relação para obter informações sobre  $Y$  a partir do conhecimento de  $X_1, X_2, \dots, X_p$ .
- É importante salientar que as relações estatísticas (correlação entre variáveis) não necessariamente implicam em relações causais, mas a presença de qualquer relação estatística fornece um ponto inicial para outras pesquisas.

Os modelos de regressão podem ser usados para predição, estimação, testes de hipótese e para modelar relações casuais.

Em geral, os principais objetivos em análise de regressão são:

- Verificar se as variáveis estão associadas: verificar se os valores de uma variável tendem a crescer (ou decrescer) à medida que os valores da outra variável crescem (ou decrescem).
- Predizer o valor de uma variável a partir de um valor conhecido da outra.
- Descrever a relação entre as variáveis: por exemplo, verificar qual é o crescimento médio esperado para uma variável dado um aumento específico em outra variável.

# Modelo de regressão linear simples

- A relação funcional mais utilizada para descrever o tipo de associação entre duas variáveis é a função linear.
- O modelo que é aplicado na estrutura de regressão mais simples é denominado por modelo de regressão linear simples.
- O termo “simples” implica em uma única variável explicativa,  $X$ , e o termo “linear” implica linearidade nos parâmetros.
- O modelo de regressão linear simples é descrito por:

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

onde  $Y$  é a variável dependente;  $X$  a variável independente ou explicativa;  $\beta_0$  é o intercepto;  $\beta_1$  é o parâmetro de inclinação; e  $\epsilon$  é o erro aleatório do modelo.

- Este modelo pode ser visto como um modelo de regressão linear simples populacional.

Considere  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , pares de valores observados de  $(X, Y)$ . Suponha que desejamos estimar os parâmetros  $\beta_0$  e  $\beta_1$ . Então, o modelo escrito como

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, n,$$

pode ser visto como um modelo de regressão linear simples amostral. No modelo de regressão linear simples usual, os  $\epsilon_i$ 's são variáveis aleatórias sujeitas às seguintes condições:

### Condição 1

$$E(\epsilon_i) = 0.$$

### Condição 2

$$Var(\epsilon_i) = \sigma^2.$$

### Condição 3

$$Cov(\epsilon_i, \epsilon_j) = 0, \quad i \neq j, \quad j = 1, 2, \dots, n.$$

Escrevendo essas  $n$  equações na forma matricial, temos:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

ou, equivalentemente,

$$Y = X\beta + \epsilon.$$

Vale ressaltar que os  $x_i$ 's são apenas números, ou seja, valores de uma variável  $X$  que não é aleatória. Porém, cada  $Y_i$  depende da quantidade aleatória  $\epsilon_i$  e, portanto, cada  $Y_i$  é uma variável aleatória.

Assim, o modelo de regressão linear simples descreve uma situação em que a média da distribuição de  $Y_i$ , para um valor específico de  $X_i = x_i$ , é dada por:

$$E(Y_i) = E(\beta_0 + \beta_1 x_i + \epsilon_i) = \beta_0 + \beta_1 x_i,$$

e a variância é

$$Var(Y_i) = Var(\beta_0 + \beta_1 x_i + \epsilon_i) = \sigma^2.$$

Assim, a média de  $Y_i$  é uma função linear de  $x_i$  e a variância de  $Y_i$  não depende do valor de  $x_i$ .

Como consequência, temos que quaisquer duas variáveis  $Y_i$  e  $Y_j$ ,  $i \neq j$ ,  $j = 1, 2, \dots, n$ , também são não-correlacionadas.



Note que, analogamente, podemos escrever o modelo de regressão populacional como

$$E(Y|X = x) = \beta_0 + \beta_1 x.$$

Os parâmetros  $\beta_0$  e  $\beta_1$  são chamados de coeficientes da regressão. Estes coeficientes têm uma interpretação simples e útil:

- O parâmetro  $\beta_1$  é a mudança na média da distribuição de  $Y_i$  por cada aumento unitário de  $x_i$ .
- O parâmetro  $\beta_0$  é o intercepto da linha de regressão com o eixo da ordenada. Se a amplitude dos dados dos  $x_i$ 's inclui o zero,  $\beta_0$  fornece a média da distribuição de  $Y_i$  quando  $x_i = 0$ . Quando a amplitude dos dados dos  $x_i$ 's não inclui o zero,  $\beta_0$  não tem uma interpretação prática.

Os parâmetros  $\beta_0$  e  $\beta_1$  são desconhecidos e devem ser estimados utilizando os dados amostrais observados.

O método de mínimos quadrados (MMQ) é mais utilizado do que qualquer outro procedimento de estimação em modelos de regressão. O MMQ fornece os estimadores de  $\beta_0$  e  $\beta_1$  tal que a soma de quadrados das diferenças entre as observações  $y_i$ 's e a linha reta ajustada seja mínima.

Assim, de todos os possíveis valores de  $\beta_0$  e  $\beta_1$ , os estimadores de mínimos quadrados (EMQ) serão aqueles que minimizam a soma de quadrados dos erros que é dada por

$$S(\beta_0, \beta_1) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2.$$

Os estimadores de mínimos quadrados de  $\beta_0$  e  $\beta_1$ , denotados  $\hat{\beta}_0$  e  $\hat{\beta}_1$ , devem satisfazer

$$\left. \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

e

$$\left. \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0.$$

Simplificando essas duas equações, teremos:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n Y_i$$

e

$$n\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n Y_i x_i,$$

que são chamadas de equações de mínimos quadrados normais.

As soluções das equações normais são:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

e

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

onde  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  e  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  são as médias dos  $Y_i$ 's e  $x_i$ 's, respectivamente.

Portanto,  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são os EMQ para o intercepto e inclinação, respectivamente.

Substituindo as observações  $y_i$ ,  $i = 1, 2, \dots, n$ , em  $\hat{\beta}_0$  e  $\hat{\beta}_1$  obtemos as estimativas

$$b_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

e

$$b_1 = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}},$$

onde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  e  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  são as médias amostrais dos  $y_i$ 's e  $x_i$ 's, respectivamente.

O modelo de regressão linear simples ajustado é

$$\hat{y}_i = b_0 + b_1 x_i,$$

que é a estimativa pontual da média de  $Y_i$  para um particular  $x_i$ .

A diferença entre o valor observado  $y_i$  e o valor ajustado correspondente  $\hat{y}_i$  é um resíduo. Matematicamente, o  $i$ -ésimo resíduo é

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i), \quad i = 1, \dots, n.$$

Os resíduos desempenham um papel importante para a investigação da adequação do modelo e para detectar violação de algumas suposições básicas.

Os resíduos podem ser vistos como realizações dos erros do modelo. Assim, para checar as suposições para os erros de variância constante e não correlacionados, devemos verificar se os resíduos parecem com uma amostra aleatória de uma distribuição com essas propriedades.

# Propriedades dos estimadores de mínimos quadrados

Os estimadores de mínimos quadrados  $\hat{\beta}_0$  e  $\hat{\beta}_1$  tem várias propriedades importantes.

Os estimadores  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são não-viciados para os parâmetros do modelo  $\beta_0$  e  $\beta_1$ . Isto é,

$$E(\hat{\beta}_1) = \beta_1$$

e

$$E(\hat{\beta}_0) = \beta_0.$$

A variância de  $\hat{\beta}_1$  é dada por

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{S_{xx}}.$$

A variância de  $\hat{\beta}_0$  é

$$Var(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right).$$

Existem várias outras propriedades importantes do ajuste por mínimos quadrados:

- ❶ A soma dos resíduos em qualquer modelo de regressão que contém o intercepto  $\beta_0$  é sempre igual a zero:  $\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0$ .
- ❷ A soma dos valores observados  $y_i$ 's é igual a soma dos valores ajustados  $\hat{y}_i$ 's:  $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$ .
- ❸ A linha de regressão de mínimos quadrados sempre passa através do ponto  $(\bar{y}, \bar{x})$  dos dados.
- ❹ A soma dos resíduos ponderados pelos correspondentes valores da variável regressora é sempre igual a zero:  $\sum_{i=1}^n x_i e_i = 0$ .
- ❺ A soma dos resíduos ponderados pelos correspondentes valores ajustados é sempre igual a zero:  $\sum_{i=1}^n \hat{y}_i e_i = 0$ .



Dentre a classe de estimadores lineares não viciados de  $\beta_0$  e  $\beta_1$ , os EMQs  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são os únicos estimadores com variância mínima. Dizemos que  $\hat{\beta}_i$  é o estimador BLUE (*best linear unbiased estimate* - melhor estimador não viciado) de  $\beta_i$ .

# Estimação de $\sigma^2$

Além das estimativas dos parâmetros  $\beta_0$  e  $\beta_1$ , a estimativa de  $\sigma^2$  é necessária para testar hipóteses e construir estimativas intervalares associados ao modelo de regressão.

A estimativa de  $\sigma^2$  é obtida dos resíduos ou soma dos quadrados dos erros,

$$\begin{aligned}SQE &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\&= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - b_1 \sum_{i=1}^n y_i(x_i - \bar{x}) \\&= S_{yy} - b_1 S_{xy} = SQT - b_1 S_{xy},\end{aligned}$$

onde  $SQT = S_{yy} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n (y_i - \bar{y})^2$  é a soma de quadrados total das observações das respostas.

A soma dos quadrados residual tem  $n - 2$  graus de liberdade, uma vez que 2 graus de liberdades estão associados com a estimativas de  $\beta_0$  e  $\beta_1$  envolvidas para obter  $\hat{y}_i$ .

Quando consideramos no lugar de  $y_i$  a variável aleatória  $Y_i$ , o valor esperado de  $SQE$  é  $E(SQE) = (n - 2)\sigma^2$ .

Então, um estimador não viciado de  $\sigma^2$  é

$$\hat{\sigma}^2 = \frac{SQE}{n - 2} = QME.$$

A quantidade  $QME$  é chamado quadrado médio do resíduo.

A raiz quadrada de  $\hat{\sigma}^2$  é muitas vezes chamada de erro padrão da regressão.

# Testes de hipóteses

Frequentemente estamos interessados em testar hipóteses sobre os parâmetros do modelo.

Esse procedimento requer uma suposição adicional: que os erros  $\epsilon_i$  do modelo sejam normalmente distribuídos.

Assim, as suposições completas são que os erros sejam independentes e normalmente distribuídos com média zero e variância  $\sigma^2$ .

Uma vez que os erros  $\epsilon_i$  tem distribuição  $N(0, \sigma^2)$ , segue que:

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2).$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right).$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right).$$

$$(n-2)QME/\sigma^2 \sim \chi_{n-2}^2$$

Além disso,  $QME$  e  $\hat{\beta}_1$  são independentes.

Suponha que desejamos testar a hipótese de que o parâmetro de inclinação é igual a uma constante  $\beta_{1_0}$ , isto é,

$$H_0 : \beta_1 = \beta_{1_0} \quad \textit{versus} \quad H_1 : \beta_1 \neq \beta_{1_0}.$$

Se  $\sigma^2$  é conhecido, podemos utilizar a estatística

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{1_0}}{\sqrt{\sigma^2/S_{xx}}} \sim N(0, 1).$$

Se  $\sigma^2$  é desconhecido, utilizamos a estatística

$$T_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{QME/S_{xx}}},$$

que tem distribuição t-Student com  $n - 2$  graus de liberdade se a hipótese nula  $H_0 : \beta_1 = \beta_{10}$  é verdadeira.

O denominador da estatística  $T_0$  é frequentemente chamado de erro padrão estimado ou, mais simplesmente, erro padrão da inclinação.

O procedimento do teste consiste em substituir  $\hat{\beta}_1$  por  $b_1$  em  $T_0$ , obtendo  $t_0$ , e comparar esse valor com o quantil  $\alpha/2$  da distribuição t-Student,  $t_{\alpha/2, n-2}$ .

Esse procedimento rejeita a hipótese nula se

$$|t_0| > t_{\alpha/2, n-2}.$$

Um procedimento similar pode ser usado para testes de hipóteses sobre o intercepto. Para testar

$$H_0 : \beta_0 = \beta_{0_0} \quad \textit{versus} \quad H_1 : \beta_0 \neq \beta_{0_0}$$

podemos utilizar a estatística de teste

$$T_0 = \frac{\hat{\beta}_0 - \beta_{0_0}}{\sqrt{QME(1/n + \bar{x}^2/S_{xx})}}$$

O denominador da estatística  $T_0$  é denominado de erro padrão do intercepto.

O procedimento do teste consiste em substituir  $\hat{\beta}_0$  por  $b_0$  em  $T_0$ , obtendo  $t_0$ , e comparar esse valor com o quantil  $\alpha/2$  da distribuição t-Student,  $t_{\alpha/2, n-2}$ .

Rejeitamos a hipótese nula  $H_0 : \beta_0 = \beta_{0_0}$  se  $|t_0| > t_{\alpha/2, n-2}$ .

Um caso especial e muito importante em teste de hipóteses é testar

$$H_0 : \beta_1 = 0 \quad \textit{versus} \quad H_1 : \beta_1 \neq 0.$$

Estas hipóteses relata a significância da regressão.

Podemos utilizar a análise de variância (ANOVA) para testar a significância da regressão.

A ANOVA é baseada no particionamento da variabilidade total da variável resposta  $Y$ .



Temos que

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i).$$

Como  $2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$ , segue que

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Já sabemos que

$$\sum_{i=1}^n (y_i - \bar{y})^2 = SQT$$

é a soma de quadrado das observações e que

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = SQE$$

é a soma de quadrado dos erros.

É conveniente chamar

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SQR$$

a soma de quadrado do modelo ou da regressão.

Assim, simbolicamente podemos escrever a equação

$$SQT = SQR + SQE.$$

A separação dos graus de liberdade é determinado como segue.

A soma de quadrado total tem  $n - 1$  graus de liberdade.

A soma de quadrado do modelo tem 1 grau de liberdade.

A soma de quadrado dos erros tem  $n - 2$  graus de liberdade.

Note que os graus de liberdade tem a propriedade de aditividade:

$$\begin{aligned} gl_T &= gl_R + gl_E \\ n - 1 &= 1 + (n - 2). \end{aligned}$$

Podemos utilizar o teste  $F$  da ANOVA para testar a hipótese  $H_0 : \beta_1 = 0$ . Temos que:

- ①  $SQE = (n - 2)QME/\sigma^2$  tem distribuição  $\chi^2_{n-2}$ ;
- ② Se a hipótese nula  $H_0 : \beta_1 = 0$  é verdadeira, então  $SQR/\sigma^2$  tem distribuição  $\chi^2_1$ ;
- ③  $SQE$  e  $SQR$  são independentes.

Por definição de uma estatística  $F$ , temos que

$$F_0 = \frac{SQR/gl_R}{SQE/gl_e} = \frac{SQR/1}{SQE/(n-2)} = \frac{QMR}{QME}$$

tem distribuição  $F$  com  $(1, n-2)$  graus de liberdade,  $F_{1, n-2}$ .

Portanto, para testar a hipótese  $H_0 : \beta_1 = 0$ , CALCULA-SE a estatística  $F_0$  e rejeita-SE  $H_0$  se

$$F_0 > F_{\alpha, 1, n-2}.$$

Em resumo, temos a tabela da ANOVA para testar a significância da regressão:

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	$F_0$
Modelo	$SQR$	1	$QME$	$\frac{QMR}{QME}$
Resíduo	$SQE$	$n - 2$	$QME$	
Total	$SQT$	$n - 1$		

# Intervalos de confiança

I.C.  $(1 - \alpha)100\%$  para  $\beta_0$ :

$$b_0 \pm t_{\alpha/2, n-2} \sqrt{QME \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}.$$

I.C.  $(1 - \alpha)100\%$  para  $\beta_1$ :

$$b_1 \pm t_{\alpha/2, n-2} \sqrt{\frac{QME}{S_{xx}}}.$$

**I.C.  $(1 - \alpha)100\%$  para a resposta média  $\mu_{Y|x_0}$  em um ponto  $x = x_0$ :**

Temos que,

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

O estimador de  $\mu_{Y|x_0}$  é dado por

$$\widehat{\mu_{Y|x_0}} = \hat{\beta}_0 + \hat{\beta}_1 x_0,$$

sendo que  $E(\widehat{\mu_{Y|x_0}}) = \beta_0 + \beta_1 x_0$  e  $Var(\widehat{\mu_{Y|x_0}}) = \sigma^2 \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$ .

Então,

$$\widehat{\mu_{Y|x_0}} \pm t_{\alpha/2, n-2} \sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}.$$

**Intervalo de predição  $(1 - \alpha)100\%$  de uma observação futura  $Y_0$  em um ponto  $x = x_0$ :**

Um modelo de regressão pode ser usado para prever a variável resposta correspondente a valores da variável explicativa não considerada no experimento.

Em predição, obtemos um valor de  $Y$  para um  $x_h$  que não pertence aos dados mas que pertence ao intervalo de variação de  $X$ .

Seja  $x_h$  um dado valor da variável explicativa  $x$  que não pertence a amostra. Então,

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h,$$

é um estimador não viciado de  $Y_h = E(Y \mid x_h) = \beta_0 + \beta_1 x_h$ .



Denote por  $e_p$  o erro na previsão a diferença. Temos que

$$e_p = (Y_h - \hat{Y}_h),$$

sendo que  $E(e_p) = 0$  e

$$Var(e_p) = Var(Y_h - \hat{Y}_h) = \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{S_{xx}} \right).$$

Temos que

$$T = \frac{Y_h - \hat{Y}_h}{\sqrt{QME \left( 1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{S_{xx}} \right)}} \sim t_{n-2}.$$

Assim, o intervalo de predição para  $Y_h$  é dado por

$$\hat{Y}_h \pm t_{\alpha/2; n-2} \sqrt{QME \left( 1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{S_{xx}} \right)}.$$

# Coeficiente de determinação

A quantidade

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT},$$

é chamada de coeficiente de determinação.

Uma vez que  $SQT$  é uma medida de variabilidade em  $y$  sem considerar o efeito da variável regressora  $x$  e  $SQE$  é uma medida da variabilidade remanescente em  $y$  após  $x$  ter sido considerada,  $R^2$  é frequentemente dito ser a proporção de variação explicada pelo regressor  $x$ .

Como  $0 \leq SQE \leq SQT$ , segue que  $0 \leq R^2 \leq 1$ .

Valores de  $R^2$  perto de 1, significa que a maior parte da variação em  $y$  é explicada pelo modelo de regressão.

As principais suposições feitas até agora no estudo de análise de regressão são as seguintes:

- 1 A relação entre a resposta  $Y$  e as variáveis regressoras é linear, pelo menos aproximadamente.
- 2 O termo de erro  $\epsilon$  tem média zero.
- 3 O termo de erro  $\epsilon$  tem variância constante  $\sigma^2$ .
- 4 Os erros são não correlacionados.
- 5 Os erros são normalmente distribuídos.

Devemos sempre duvidar da validade destas suposições e realizar análises para examinar a adequação do modelo.

## Identificação de observações influentes

Observações influentes são aquelas que, de acordo com vários critérios aparecem exercendo forte impacto no modelo ajustado.

As principais estatísticas para a identificação de observações influentes são os resíduos padronizados e os resíduos estudentizados.

### Resíduos Padronizados

Uma vez que a variância média aproximada de um resíduo é estimado por  $QME$ , um dimensionamento lógico para os resíduos seriam os resíduos padronizados

$$d_i = \frac{e_i}{\sqrt{QME}}, \quad i = 1, \dots, n.$$

Os resíduos padronizados tem média zero e variância aproximadamente unitária. Consequentemente, um grande resíduo padronizado potencialmente indica um outlier.

## Resíduos Estudentizados

Utilizar  $QME$  como a variância do resíduo  $e_i$  é somente uma aproximação.

Podemos melhorar o escalonamento do resíduo dividindo  $e_i$  por seu desvio-padrão exato.

Considere a equação de vetor de resíduos escrita como

$$e = (I - H)Y,$$

em que  $H = X(X^\top X)^{-1}X^\top$ .

A matriz  $H$  é simétrica ( $H^\top = H$ ) e idempotente ( $HH = H$ ).

Similarmente, a matriz  $I - H$  também é simétrica e idempotente.

Substituindo  $Y = X\beta + \epsilon$  na equação dos resíduos temos a matriz de covariância dos resíduos dada por:

$$Var(e) = Var((I - H)\epsilon) = (I - H)Var(\epsilon)(I - H)^\top = \sigma^2(I - H).$$

A variância do resíduo  $e_i$  é

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii}),$$

em que  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $H$ .

As violações das suposições do modelo são mais prováveis em pontos distantes e estas violações podem ser difíceis de detectar a partir de inspeção dos resíduos ordinários  $e_i$ 's (ou dos resíduos padronizados  $d_i$ 's) porque são geralmente menor.

Um procedimento lógico, então, é examinar os resíduos estudentizados

$$r_i = \frac{e_i}{\sqrt{QME(1 - h_{ii})}}, \quad i = 1, \dots, n.$$

Os resíduos padronizados e estudentizados frequentemente transmitem informações equivalentes. Porém, o resíduo estudentizado é mais preciso.

## Gráficos de Resíduos

A análise gráfica dos resíduos é uma maneira muito eficaz para investigar a adequação do ajuste de um modelo de regressão e de verificar os pressupostos básicos. Alguns gráficos utilizados para diagnóstico dos modelos de regressão são.

**Resíduos  $\times$  valores ajustados:** O gráfico dos resíduos  $e_i$ 's (ou dos  $d_i$ ,  $r_i$ ) *versus* os correspondentes valores ajustados  $\hat{y}_i$ 's é útil para detectar vários tipos comuns de inadequação do modelo.

**Resíduos  $\times$  variáveis explicativas:** O gráfico dos resíduos *versus* os correspondentes valores de cada regressor pode ser útil. É desejável que uma faixa horizontal em torno de zero contenha os resíduos.

**Q-Q norm:** Utilizamos o Q-Q norm para checar a normalidade dos resíduos. Este gráfico apresenta a probabilidade acumulada da normal padrão, representada por uma linha reta, e os pontos referentes aos resíduos ordenados *versus* suas respectivas probabilidades acumuladas. A normalidade dos resíduos é verificada visualmente quando este pontos se aproximam da linha reta.

# Exemplo

Um estudo sobre duração de certas operações está investigando o tempo requerido (em segundos) para acondicionar objetos e o volume (em  $\text{dm}^3$ ) que eles ocupam. Uma amostra foi observada e obtiveram-se os seguintes resultados:

Tempo (Y) 10.8 14.4 19.6 18.0 8.4 15.2 11.0 13.3 23.1

Volume (X) 20.39 24.92 34.84 31.72 13.59 30.87 17.84 23.22 39.65



# Análise descritiva:

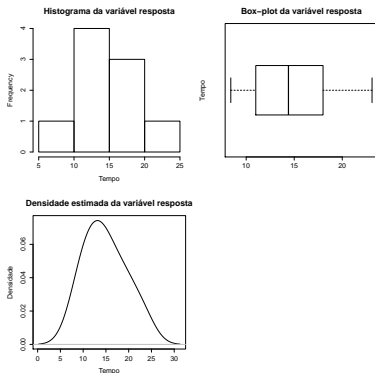


Figura: Análise descritiva da variável resposta Tempo.

Valor mínimo = 8.40; Primeiro quantil = 11.00; Média = 14.40;  
Mediana = 14.87 Terceiro quantil = 18.00; Valor máximo = 23.10

## Estimação dos parâmetros:

Temos que:

$$\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = 26.33778$$

$$\bar{y} = \frac{1}{9} \sum_{i=1}^9 y_i = 14.86667$$

$$S_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 580.8372$$

$$S_{yy} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = 176.1$$

$$S_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = 313.2503$$

Logo, as estimativas dos parâmetros  $\beta_1$  e  $\beta_0$  são dadas por

$$b_1 = \frac{S_{xy}}{S_{xx}} = 0.5393084$$

Temos que

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

e

$$Var(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right).$$

Portanto,

$$\widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}} = 0.001761481$$

e

$$\widehat{Var}(\hat{\beta}_0) = \hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) = 1.335504.$$

## Análise de significância do modelo:

Valores preditos  $\hat{y}_i$  de  $y_i$  são dados por  $y_i = b_0 + b_1x_i$ :  $\hat{y}_i$ : 11.658980  
14.102047 19.451986 17.769344 7.991684 17.310932 10.283744 13.185223  
22.046059

$$SQT = \sum_{i=1}^n (y_i - \bar{y})^2 = 176.1$$

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 7.161477 \text{ e } QME = \frac{SQE}{n-2} = 1.023068$$

$$SQR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 168.9385 \text{ e } QMR = \frac{SQR}{1} = 168.9385$$

$$F_0 = \frac{QMR}{QME} = 165.1293$$

## Coefficiente de determinação:

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT} = 0.9593329$$

Seja  $\mathbf{X}_t = (X_{1t}, \dots, X_{kt})'$  um processo linear vetorial da forma

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi(j) \boldsymbol{\xi}_{t-j}, \quad (1)$$

onde  $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_k)$  é o vetor de médias,  $\Psi(j)$  são matrizes  $K \times K$  absolutamente somáveis, isto é,  $(\sum_{j=0}^{\infty} |\Psi(j)| < \infty)$  e  $\boldsymbol{\xi}'_j = (\xi_{1j}, \dots, \xi_{kj})$  são processos vetoriais ruído branco com média zero e matriz de covariância

$$\Gamma_{\boldsymbol{\xi}}(h) = E(\boldsymbol{\xi}_j \boldsymbol{\xi}'_{j+h}) = \begin{cases} \Sigma, & h = 0 \\ 0, & h \neq 0 \end{cases} \quad (2)$$

onde  $\Sigma$  é uma matriz não negativa definida.

Uma classe paramétrica de modelos de séries temporais pertencente ao processo linear definido em 1 é o vetor autorregressivo média móveis (VARMA) que é a solução do sistema

$$\Phi(B) \Delta^d(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \Theta(B)\boldsymbol{\varepsilon}_t, \quad (3)$$

onde  $B$  é o operador atraso,  $\boldsymbol{\mu}$  é o vetor de médias e  $\boldsymbol{\varepsilon}_t$  é o ruído branco com  $E(\boldsymbol{\varepsilon}_t) = 0$  e  $Var(\boldsymbol{\varepsilon}_t) = \Sigma$ . Os operadores  $\Phi(B) = I - \sum_{i=1}^p \Phi_i B^i$  e  $\Theta(B) = I + \sum_{i=1}^q \Theta_i B^i$  são matrizes polinomiais com ordem  $p, q$  respectivamente,  $d \in \mathbb{N}$  e  $I$  é a matriz identidade de dimensão  $k \times k$  e  $\Phi_i$  e  $\Theta_i$  são matrizes  $k \times k$  de constantes.

- ❶ MORETTIN, P. A. e BUSSAB, W. O. (2010). Estatística Básica. 6a ed. São Paulo: Saraiva.
- ❷ MARTINS, Gilberto de Andrade e DOMINGUES, Osmar. (2011) Estatística Geral e Aplicada. 4a ed. São Paulo: Atlas.
- ❸ Reinsen, V. A. e Silva, A. N. (2011). O uso da linguagem R para cálculos de Estatística Básica. Edufes.
- ❹ R: linguagem de programação usada para análise estatística e gráficos.
- ❺ Download: <https://www.r-project.org/>
- ❻ Editores para R:
  - ▶ R-Studio: <https://www.rstudio.com/>
  - ▶ Tinn-R: <https://sourceforge.net/projects/tinn-r/>
- ❼ Melo, M. M. (2015). Tese: "Correlação entre percepção do incômodo e exposição ao material particulado presente na atmosfera e sedimentado". PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AMBIENTAL (UFES).