

# Tópico 4: Probabilidade e Inferência Estatística com R

MRS

## 4. Probabilidade e Inferência Estatística com R

A estatística inferencial permite tirar conclusões sobre populações com base em amostras, utilizando a teoria da probabilidade. Neste tópico, exploramos os conceitos básicos de probabilidade, distribuições teóricas e testes estatísticos em R.

### Conceitos Básicos de Probabilidade

- A probabilidade mede a chance de ocorrência de um evento.
- **Evento**: um resultado ou conjunto de resultados.
- **Espaço amostral**: conjunto de todos os resultados possíveis.
- **Probabilidade condicional**: probabilidade de um evento dado que outro já ocorreu.

### Exemplo

```
# Simulando 10 lançamentos de moeda, onde o resultado possível é "cara" ou "coroa"  
sample(c("cara", "coroa"), size = 10, replace = TRUE)
```

```
[1] "coroa" "cara" "coroa" "cara" "coroa" "cara" "cara" "coroa" "coroa"  
[10] "coroa"
```

## Exemplo explicado

- `sample(...)`: função que gera amostras aleatórias de um vetor.
- `c("cara", "coroa")`: cria um vetor com dois elementos – “cara” e “coroa”, representando os dois lados de uma moeda.
- `size = 10`: indica que queremos 10 amostras (10 lançamentos de moeda).
- `replace = TRUE`: significa que o sorteio é com reposição – ou seja, após cada sorteio, o item é “colocado de volta” e pode ser sorteado novamente (como acontece com uma moeda real).

## Distribuições de Probabilidade

- Distribuições de probabilidade descrevem como a probabilidade está distribuída entre os possíveis valores de uma variável aleatória.
- Estas distribuições ajudam a modelar incertezas e realizar inferência estatística.
- Tabela de distribuições de probabilidade:

| Distribuição | Tipo     | Uso Principal                                  | Função no R                                      | Exemplo  |
|--------------|----------|--|--|--|
| Normal       | Contínua | Fenômenos naturais, erros, médias              | <code>pnorm()</code> ,<br><code>qnorm()</code>   | <code>pnorm(1.96)</code> ,<br><code>qnorm(0.975)</code>                      |
| Binomial     | Discreta | Número de sucessos em ensaios independentes    | <code>dbinom()</code> ,<br><code>pbinom()</code> | <code>dbinom(2,</code><br><code>size = 5,</code><br><code>prob = 0.3)</code> |
| Poisson      | Discreta | Contagem de eventos raros em tempo/espaço fixo | <code>dpois()</code> ,<br><code>ppois()</code>   | <code>dpois(4,</code><br><code>lambda = 2)</code>                            |

## Exemplos explicados

```
# Distribuição Normal
pnorm(1.96)          # probabilidade acumulada
```

```
[1] 0.9750021
```

```
qnorm(0.975)      # quantil
```

```
[1] 1.959964
```

```
# Distribuição Binomial  
dbinom(2, size = 5, prob = 0.3)
```

```
[1] 0.3087
```

```
# Distribuição de Poisson  
dpois(4, lambda = 2)
```

```
[1] 0.09022352
```

### Tabela resumo

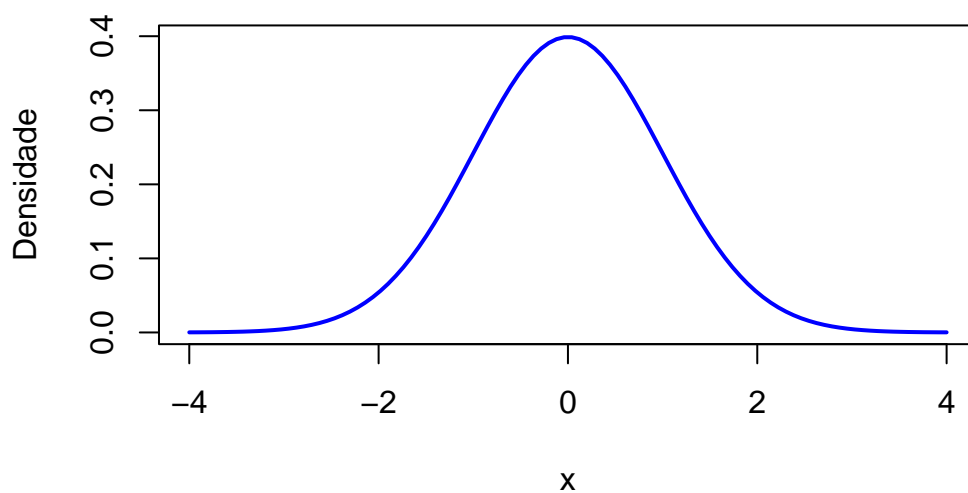
| Distribuição | Código em R                                  | Enunciado  |
|--------------|--|--|
| Normal       | <code>pnorm(1.96)</code>                     | Qual a probabilidade de um valor de uma normal padrão ser menor ou igual a 1.96?                                 |
| Normal       | <code>qnorm(0.975)</code>                    | Qual é o valor x tal que 97,5% dos valores de uma normal padrão estão abaixo dele?                               |
| Binomial     | <code>dbinom(2, size = 5, prob = 0.3)</code> | Qual a probabilidade de acertar exatamente 2 questões em 5, respondendo com 30% de chance de acerto por questão? |
| Poisson      | <code>dpois(4, lambda = 2)</code>            | Qual a probabilidade de receber exatamente 4 chamadas por minuto, sabendo que a média é 2 chamadas por minuto?   |

## Visualização de Distribuições de Probabilidade

### Distribuição Normal

```
x <- seq(-4, 4, length = 100)  
y <- dnorm(x)  
  
plot(x, y, type = "l", lwd = 2, col = "blue",  
      main = "Distribuição Normal (média = 0, desvio padrão = 1)",  
      xlab = "x", ylab = "Densidade")
```

### Distribuição Normal (média = 0, desvio padrão = 1)

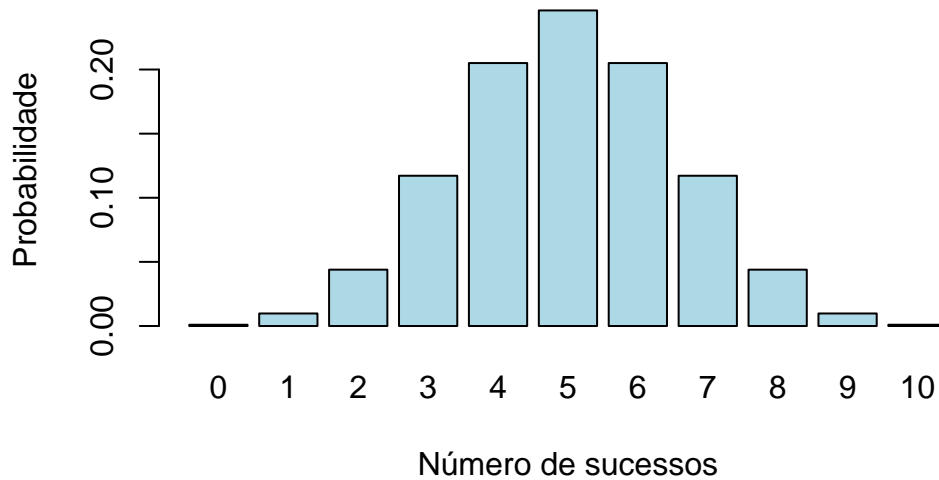


### Distribuição Binomial

```
x <- 0:10
y <- dbinom(x, size = 10, prob = 0.5)

barplot(y, names.arg = x,
        main = "Distribuição Binomial (n = 10, p = 0.5)",
        xlab = "Número de sucessos", ylab = "Probabilidade",
        col = "lightblue")
```

### Distribuição Binomial (n = 10, p = 0.5)



### Distribuição de Poisson

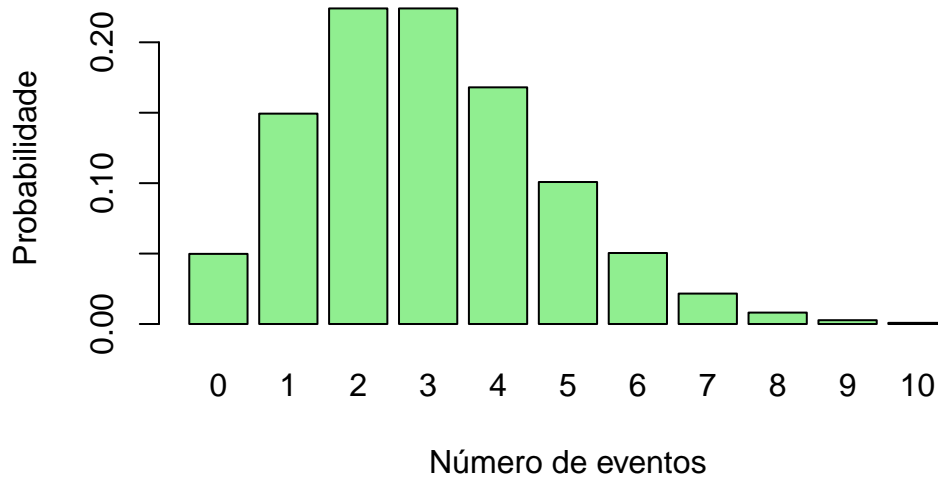
```
x <- 0:10
y <- dpois(x, lambda = 3)

barplot(y, names.arg = x,
        main = "Distribuição de Poisson ( = 3)",
        xlab = "Número de eventos", ylab = "Probabilidade",
        col = "lightgreen")
```

Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
conversion failure on 'Distribuição de Poisson ( = 3)' in 'mbcsToSbcs': dot  
substituted for <ce>

Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
conversion failure on 'Distribuição de Poisson ( = 3)' in 'mbcsToSbcs': dot  
substituted for <bb>

### Distribuição de Poisson ( $\lambda = 3$ )



### Inferência Estatística

- É o processo de tirar conclusões sobre uma população com base em dados de uma amostra.
- Em vez de medir todos os elementos de uma população, usamos dados amostrais para fazer estimativas, testes de hipóteses e previsões.
- Envolve **estimativa** e **testes de hipóteses**.
- Objetivos:
  - Estimar parâmetros populacionais.
    - \* Ex: usar a média de uma amostra para estimar a média da população.
  - Testar hipóteses.
    - \* Ex: verificar se há evidência suficiente de que um novo medicamento é mais eficaz que o atual.
  - Fazer previsões ou decisões com base em dados.
    - \* Ex: prever resultados eleitorais.

### Métodos principais da Inferência Estatística

| Método                        | O que faz   | Exemplo   |
|-------------------------------|---|---|
| <b>Estimativa Pontual</b>     | Calcula um valor único como aproximação de um parâmetro                                     | <code>mean(amostra)</code>                        |
| <b>Intervalo de Confiança</b> | Fornece um intervalo onde o parâmetro populacional deve estar com certa confiança (ex: 95%) | <code>t.test(amostra)\$conf.int</code>            |
| <b>Teste de Hipóteses</b>     | Avalia se uma afirmação sobre a população é estatisticamente plausível                      | <code>t.test(amostra, mu = valor_hipotese)</code> |

## Exemplos

```
# Estimativa Pontual e Intervalar
amostra <- c(7.5, 6.8, 7.2, 7.9, 6.5)
t.test(amostra, conf.level = 0.95)
```

### One Sample t-test

```
data: amostra
t = 28.976, df = 4, p-value = 8.444e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 6.492024 7.867976
sample estimates:
mean of x
 7.18
```

```
# Testes de Hipóteses: teste t para duas amostras
grupo1 <- c(5.1, 4.9, 5.3, 5.0)
grupo2 <- c(5.8, 6.0, 6.1, 5.9)
```

```
# Testes de Hipóteses: teste Qui-quadrado
tabela <- matrix(c(10, 20, 15, 25), nrow = 2)
chisq.test(tabela)
```

### Pearson's Chi-squared test with Yates' continuity correction

```
data:  tabela
X-squared = 0.011667, df = 1, p-value = 0.914
```

## Exercício Prático

1. Gere uma amostra com 30 valores de uma distribuição normal com média 100 e desvio padrão 15.
2. Calcule a média e o intervalo de confiança da média.
3. Compare duas amostras usando o teste t.
4. Realize um teste qui-quadrado com duas variáveis fictícias.

## Solução

```
# 1. Amostra normal
set.seed(123)
amostra1 <- rnorm(30, mean = 100, sd = 15)
amostra1
```

```
[1]  91.59287  96.54734 123.38062 101.05763 101.93932 125.72597 106.91374
[8]  81.02408  89.69721  93.31507 118.36123 105.39721 106.01157 101.66024
[15]  91.66238 126.80370 107.46776  70.50074 110.52034  92.90813  83.98264
[22]  96.73038  84.60993  89.06663  90.62441  74.69960 112.56681 102.30060
[29]  82.92795 118.80722
```

```
# 2. Estatísticas
mean(amostra1)
```

```
[1] 99.29344
```

```
t.test(amostra1)
```

### One Sample t-test

```
data:  amostra1
t = 36.958, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
```



```
93.7986 104.7883
sample estimates:
mean of x
99.29344
```

```
# 3. Comparando duas amostras
amostra2 <- rnorm(30, mean = 105, sd = 15)
t.test(amostra1, amostra2)
```

#### Welch Two Sample t-test

```
data: amostra1 and amostra2
t = -2.3755, df = 56.559, p-value = 0.02094
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-15.448139 -1.315124
sample estimates:
mean of x mean of y
99.29344 107.67508
```

```
# 4. Qui-quadrado
genero <- c("F", "M", "F", "M", "F", "M", "F", "M")
resposta <- c("Sim", "Não", "Sim", "Sim", "Não", "Não", "Sim", "Não")
tabela <- table(genero, resposta)
# chisq.test(tabela) # o teste do qui-quadrado requer que todas as frequências esperadas > 5
fisher.test(tabela)
```

#### Fisher's Exact Test for Count Data

```
data: tabela
p-value = 0.4857
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.001607888 4.722931239
sample estimates:
odds ratio
0.156047
```