

# PROBABILIDADE E ESTATÍSTICA - **Lab. Informática**

## Contrastando hipóteses: da formulação de modelos à atualização do conhecimento

Prof. Fabio Cop (*fcferreira@unifesp.br*)

Instituto do Mar - Unifesp

2026-02-20

### Conteúdo

<b>1</b>	<b>Preparação: Carregando os Dados da Turma</b>	<b>2</b>
<b>2</b>	<b>Reproduzindo a Análise da Aula</b>	<b>2</b>
2.1	Calculando as Frequências . . . . .	3
2.2	Visualizando os Dados com Gráficos . . . . .	3
2.3	Calculando as Verossimilhanças . . . . .	4
2.4	Comparando Modelos com Razão de Verossimilhanças . . . . .	4
2.5	Teste a Independência de H dos Dados . . . . .	5
<b>3</b>	<b>Simular e Calibrar</b>	<b>6</b>
3.1	Simulando sob H . . . . .	6
3.2	Visualizando as Simulações . . . . .	6
<b>4</b>	<b>Modelos a Priori</b>	<b>7</b>
4.1	A Priori da Turma . . . . .	7
4.2	Calculando a Posteriori com Diferentes Prioris . . . . .	7
4.3	Visualizando a Atualização Bayesiana . . . . .	8
<b>5</b>	<b>Recapitulando as Etapas de Hoje</b>	<b>9</b>

**Curso:** Bacharelado Interdisciplinar em Ciências do Mar  
**Unidade Curricular (UC):** Probabilidade e Estatística  
**Atividade:** Prática no Laboratório de Informática  
**Duração:** 30 minutos

### Instruções

1. **Copie e execute o código**
2. **Observe** os resultados que aparecem
3. **Responda** às perguntas no documento que você vai entregar
4. **Entrega:** Acesse o Moodle e adicione suas respostas às perguntas abaixo
5. **Prazo:** Final da aula de hoje

## 1 Preparação: Carregando os Dados da Turma

Carregue os dados coletados no início da aula (dado mental e dado físico) para começar a análise.

### *Código*

```
# Carregar os dados do arquivo CSV
dados <- read.csv("/caminhodoarquivo/dados_turma_aula1.csv")

# Separar os dados em dois vetores
mental <- dados$dado_mental
fisico <- dados$dado_fisico

# Contar quantas observações temos
n_mental <- length(mental)
n_fisico <- length(fisico)

# Mostrar os primeiros valores
print("Primeiras respostas - Dado Mental:")
head(mental, 10)

print("Primeiras respostas - Dado Físico:")
head(fisico, 10)

print(paste("Total de observações (dado mental):", n_mental))
print(paste("Total de observações (dado físico):", n_fisico))
```

### O que observar

- O R carregou o arquivo CSV com sucesso.
- Você vê os primeiros 10 números do dado mental (valores de 1 a 6).
- Você vê os primeiros 10 números do dado físico (valores de 1 a 6).
- Aparece o total de observações de cada experimento.

## 2 Reproduzindo a Análise da Aula

Refaça os mesmos passos mostrados em sala: calcular frequências, fazer gráficos e calcular as verossimilhanças de  $H_0$  e  $H_1$ .

## 2.1 Calculando as Frequências

### Código

```
# Contar quantas vezes cada face apareceu
freq_mental <- table(factor(mental, levels = 1:6))
freq_fisico <- table(factor(fisico, levels = 1:6))

# Mostrar as contagens
print("Frequências absolutas - Dado Mental:")
freq_mental

print("Frequências absolutas - Dado Físico:")
freq_fisico

# Calcular as proporções (dividir pelo total)
prop_mental <- freq_mental / n_mental
prop_fisico <- freq_fisico / n_fisico

print("Proporções - Dado Mental:")
round(prop_mental, 3)

print("Proporções - Dado Físico:")
round(prop_fisico, 3)
```

### O que observar

- As **frequências absolutas** mostram quantas vezes cada face (1, 2, 3, 4, 5, 6) apareceu.
- As **proporções** mostram a porcentagem de cada face (valores entre 0 e 1).
- Compare mentalmente: alguma face apareceu muito mais que as outras?

## 2.2 Visualizando os Dados com Gráficos

### Código

```
# Configurar para mostrar dois gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico do dado físico
barplot(prop_fisico,
        col = "seagreen",
        main = "Dado Físico",
        xlab = "Face",
        ylab = "Proporção",
        ylim = c(0, 0.35))
abline(h = 1/6, lty = 2, col = "red", lwd = 2)

# Gráfico do dado mental
barplot(prop_mental,
        col = "steelblue",
        main = "Dado Mental",
        xlab = "Face",
        ylab = "Proporção",
        ylim = c(0, 0.35))
abline(h = 1/6, lty = 2, col = "red", lwd = 2)
```

### O que observar

- Você deve ver **dois gráficos de barras** lado a lado.
- A **linha vermelha tracejada** marca  $1/6 = 0.167$  (o esperado se o dado fosse uniforme).
- Compare os dois gráficos: qual deles parece mais próximo da linha vermelha?

### PERGUNTA 1

Olhando para os dois gráficos:

- a) O dado físico parece seguir uma distribuição uniforme (todas as faces com proporção próxima de  $1/6$ )? Justifique.
- b) O dado mental parece seguir uma distribuição uniforme? Quais faces aparecem mais ou menos que o esperado?

## 2.3 Calculando as Verossimilhanças

### Código

```
# Definir as probabilidades de cada modelo
p_h1 <- rep(1/6, 6) # H: todas as faces com prob = 1/6
p_h2 <- c(1, 2, 4, 4, 2, 1) / 14 # H: viés central

# Ver as probabilidades de H
print("Probabilidades do modelo H:")
round(p_h2, 3)

# Calcular log-verossimilhança de H para o dado físico
loglik_H1_fisico <- sum(freq_fisico * log(p_h1))

# Calcular log-verossimilhança de H para o dado físico
loglik_H2_fisico <- sum(freq_fisico * log(p_h2))

# Calcular log-verossimilhança de H para o dado mental
loglik_H1_mental <- sum(freq_mental * log(p_h1))

# Calcular log-verossimilhança de H para o dado mental
loglik_H2_mental <- sum(freq_mental * log(p_h2))

# Mostrar os resultados
print("=== DADO FÍSICO ===")
print(paste("Log-verossimilhança H:", round(loglik_H1_fisico, 2)))
print(paste("Log-verossimilhança H:", round(loglik_H2_fisico, 2)))

print("=== DADO MENTAL ===")
print(paste("Log-verossimilhança H:", round(loglik_H1_mental, 2)))
print(paste("Log-verossimilhança H:", round(loglik_H2_mental, 2)))
```

### O que observar

- Você vê os valores de log-verossimilhança (números negativos).
- Valores **menos negativos** = maior verossimilhança = modelo mais compatível com os dados.
- Compare H vs H para cada experimento.

## 2.4 Comparando Modelos com Razão de Verossimilhanças

### Código

```
# Calcular a razão H/H para cada experimento
razao_fisico <- exp(loglik_H2_fisico - loglik_H1_fisico)
razao_mental <- exp(loglik_H2_mental - loglik_H1_mental)

print("=== RAZÃO DE VEROSSIMILHANÇAS (H/H) ===")
print(paste("Dado Físico:", round(razao_fisico, 2)))
print(paste("Dado Mental:", round(razao_mental, 2)))
```

#### O que observar

- A **razão de verossimilhanças** compara os dois modelos.
- Se a razão  $> 1$ :  $H_1$  é mais verossímil que  $H_0$ .
- Se a razão  $< 1$ :  $H_0$  é mais verossímil que  $H_1$ .
- Quanto maior o número, mais forte é a evidência.

#### PERGUNTA 2

- Para o dado físico, qual hipótese ( $H_0$  ou  $H_1$ ) é mais compatível com os dados? Por quê?
- Para o dado mental, qual hipótese é mais compatível? A diferença é grande?

## 2.5 Teste a Independência de $H_0$ dos Dados

### Código

```
# Vamos criar dados fictícios diferentes, mas com o MESMO total
dados_equilibrados <- c(rep(1, 8), rep(2, 9), rep(3, 8),
                        rep(4, 9), rep(5, 8), rep(6, 8))

dados_extremos <- c(rep(1, 25), rep(2, 0), rep(3, 0),
                    rep(4, 0), rep(5, 0), rep(6, 25))

# Ambos têm n = 50 observações
n_teste <- 50

# Calcular verossimilhança de  $H_0$  para os dois casos
freq_eq <- table(factor(dados_equilibrados, levels = 1:6))
freq_ex <- table(factor(dados_extremos, levels = 1:6))

loglik_H1_equilibrado <- sum(freq_eq * log(p_h1))
loglik_H1_extremo <- sum(freq_ex * log(p_h1))

print("=== TESTE:  $H_0$  É CEGO AO PADRÃO? ===")
print(paste("H para dados equilibrados:", round(loglik_H1_equilibrado, 2)))
print(paste("H para dados extremos:", round(loglik_H1_extremo, 2)))
print(paste("São iguais?",
            round(loglik_H1_equilibrado, 2) == round(loglik_H1_extremo, 2)))
```

#### O que observar

- Os dois conjuntos de dados são **completamente diferentes**.
- Mas a verossimilhança de  $H_0$  é **exatamente a mesma**.
- Isso confirma que  $H_0$  só depende do total de observações, não do padrão.

#### PERGUNTA 3

Com base neste teste, explique por que precisamos de pelo menos **duas hipóteses** para avaliar modelos. O que aconteceria se só tivéssemos  $H_0$ ?

## 3 Simular e Calibrar

Simule 1000 “turmas virtuais” sob  $H$  e verifique onde nossos dados reais se encaixam nessa distribuição de simulações.

### 3.1 Simulando sob $H$

#### *Código*

```
# Número de simulações
n_sim <- 1000

# Vetor para guardar as razões de verossimilhança
razoes_simuladas <- numeric(n_sim)

# Loop: para cada simulação...
for (i in 1:n_sim) {
  # Simular uma turma sob H (dado justo)
  amostra_sim <- sample(1:6, n_mental, replace = TRUE, prob = p_h1)

  # Calcular frequências
  freq_sim <- table(factor(amostra_sim, levels = 1:6))

  # Calcular log-verossimilhanças
  loglik_H1_sim <- sum(freq_sim * log(p_h1))
  loglik_H2_sim <- sum(freq_sim * log(p_h2))

  # Guardar a razão
  razoes_simuladas[i] <- exp(loglik_H2_sim - loglik_H1_sim)
}

print("Simulação concluída!")
print(paste("Média das razões simuladas:", round(mean(razoes_simuladas), 2)))
print(paste("Desvio padrão:", round(sd(razoes_simuladas), 2)))
```

#### O que observar

- O R simulou 1000 turmas onde **todos** os dados vêm de um dado justo ( $H$ ).
- Para cada turma simulada, calculamos a razão  $H/H$ .
- A média nos diz qual razão é “típica” quando  $H$  é verdadeiro.

### 3.2 Visualizando as Simulações

#### *Código*

```
# Fazer o histograma das razões simuladas
hist(razoes_simuladas,
     breaks = 30,
     col = "lightblue",
     main = "Distribuição de H/H sob simulações de H",
     xlab = "Razão de Verossimilhança (H/H)",
     ylab = "Frequência")

# Marcar onde caiu o dado físico
abline(v = razao_fisico, col = "green", lwd = 3, lty = 2)
text(razao_fisico, 100, "Dado Físico", pos = 4, col = "green")

# Marcar onde caiu o dado mental
abline(v = razao_mental, col = "red", lwd = 3, lty = 2)
```

```
text(razao_mental, 120, "Dado Mental", pos = 4, col = "red")

# Marcar a linha de 1 (onde H e H têm mesma verossimilhança)
abline(v = 1, col = "black", lwd = 2)
```

#### O que observar

- O **histograma azul** mostra as razões quando geramos dados de um dado justo.
- A **linha verde** marca onde caiu o dado físico da nossa turma.
- A **linha vermelha** marca onde caiu o dado mental da nossa turma.
- A **linha preta** marca razão = 1 (empate entre H e H).

#### PERGUNTA 4

- a) O valor do dado físico está dentro da região típica das simulações sob H ? O que isso significa?
- b) O valor do dado mental está dentro ou fora da região típica? O que isso indica sobre a necessidade de H ?
- c) Baseado neste histograma, você diria que os dados do dado mental foram gerados por um processo uniforme (H) ?

## 4 Modelos a Priori

Explore como diferentes crenças iniciais (priors) afetam a conclusão final (posteriori).

### 4.1 A Priori da Turma

#### Código

```
# Inserir os valores REAIS da votação da turma
# (substituir pelos números do quadro)
n_turma <- 50
votos_H1 <- 32 # Quantos votaram em H
votos_H2 <- 18 # Quantos votaram em H

# Calcular a priori
prior_turma <- c(H1 = votos_H1 / n_turma, H2 = votos_H2 / n_turma)

print("Priori da Turma:")
round(prior_turma, 2)
```

### 4.2 Calculando a Posteriori com Diferentes Priors

#### Código

```
# Função auxiliar para calcular posteriori
calcular_posteriori <- function(prior, loglik_H1, loglik_H2) {
  lik <- c(H1 = exp(loglik_H1), H2 = exp(loglik_H2))
  post <- prior * lik
  post <- post / sum(post)
  return(post)
}

# Posteriori com priori da turma
post_turma_mental <- calcular_posteriori(prior_turma,
                                         loglik_H1_mental,
                                         loglik_H2_mental)
```

```

# Posteriori com priori 50/50 (sem preferência)
prior_neutro <- c(H1 = 0.5, H2 = 0.5)
post_neutro_mental <- calcular_posteriori(prior_neutro,
                                           loglik_H1_mental,
                                           loglik_H2_mental)

# Posteriori com priori 90/10 (forte preferência por H)
prior_forte_H1 <- c(H1 = 0.9, H2 = 0.1)
post_forte_H1_mental <- calcular_posteriori(prior_forte_H1,
                                             loglik_H1_mental,
                                             loglik_H2_mental)

# Posteriori com priori 10/90 (forte preferência por H)
prior_forte_H2 <- c(H1 = 0.1, H2 = 0.9)
post_forte_H2_mental <- calcular_posteriori(prior_forte_H2,
                                             loglik_H1_mental,
                                             loglik_H2_mental)

# Mostrar resultados
print("=== PRIORI DA TURMA ===")
print(round(prior_turma, 3))
print("Posteriori:")
print(round(post_turma_mental, 3))

print("=== PRIORI NEUTRO (50/50) ===")
print(round(prior_neutro, 3))
print("Posteriori:")
print(round(post_neutro_mental, 3))

print("=== PRIORI FORTE EM H (90/10) ===")
print(round(prior_forte_H1, 3))
print("Posteriori:")
print(round(post_forte_H1_mental, 3))

print("=== PRIORI FORTE EM H (10/90) ===")
print(round(prior_forte_H2, 3))
print("Posteriori:")
print(round(post_forte_H2_mental, 3))

```

### 4.3 Visualizando a Atualização Bayesiana

#### *Código*

```

# Criar uma figura com 4 painéis
par(mfrow = c(2, 2))

# Paine 1: Priori e Posteriori da Turma
barplot(rbind(prior_turma, post_turma_mental),
        beside = TRUE,
        col = c("grey70", "salmon"),
        main = "Priori da Turma",
        ylab = "Probabilidade",
        ylim = c(0, 1),
        legend.text = c("Priori", "Posteriori"))

# Paine 2: Priori e Posteriori Neutro
barplot(rbind(prior_neutro, post_neutro_mental),
        beside = TRUE,

```



```

col = c("grey70", "salmon"),
main = "Priori Neutro (50/50)",
ylab = "Probabilidade",
ylim = c(0, 1),
legend.text = c("Priori", "Posteriori"))

# Paine 3: Priori e Posteriori Forte em H
barplot(rbind(prior_forte_H1, post_forte_H1_mental),
        beside = TRUE,
        col = c("grey70", "salmon"),
        main = "Priori Forte em H (90/10)",
        ylab = "Probabilidade",
        ylim = c(0, 1),
        legend.text = c("Priori", "Posteriori"))

# Paine 4: Priori e Posteriori Forte em H
barplot(rbind(prior_forte_H2, post_forte_H2_mental),
        beside = TRUE,
        col = c("grey70", "salmon"),
        main = "Priori Forte em H (10/90)",
        ylab = "Probabilidade",
        ylim = c(0, 1),
        legend.text = c("Priori", "Posteriori"))

```

#### O que observar

- Quatro gráficos mostrando como diferentes prioris levam a diferentes posteriores.
- Em cada gráfico, as barras cinzas são a **priori** (antes dos dados).
- As barras salmão são a **posteriori** (depois dos dados).
- Observe como a evidência dos dados “puxa” a crença em direção a H .

#### PERGUNTA 5

- a) Em qual dos quatro cenários a priori teve **maior** influência na posteriori? Por quê?
- b) Em qual cenário a priori teve **menor** influência? O que isso nos ensina sobre quando a priori importa mais ou menos?
- c) Mesmo com uma priori muito forte em H (90/10), a posteriori mudou em direção a H ? O que isso indica sobre a força da evidência nos dados do dado mental?

## 5 Recapitulando as Etapas de Hoje

Conceitos fundamentais da inferência estatística/probabilística:

1. **Modelos probabilísticos:** Traduzir hipóteses em equações matemáticas com probabilidades.
2. **Verossimilhança:** Medir quão compatível um modelo é com os dados observados.
3. **Comparação de modelos:** Comparar a verossimilhança entre os modelos.
4. **Simulação:** Gerar dados para entender o que cada modelo pode prever.
5. **Inferência Bayesiana:** Atualizar as crenças combinando conhecimento prévio (priori) com evidências (dados).