Teste de Ajustamento de Rayleigh - Exercício 10

Uma equipa de engenheiros está a analisar a distribuição da velocidade (X, em m/s) do vento dominante em determinado local onde se pretende construir uma aerogare, tendo obtido a seguinte amostra com 200 observações:
2.3, 2.7, 5.2, 0.7, 2.9, 0.6, 2.6, 2.2, 3.8, 0.5, 4.9, 5.4, 3.7, 0.4, 4, 3.6, 2, 0.8, 2.5, 2.8, 1.7, 3.3, 1.5, 0.4, 6.4, 1.5, 6, 2.1, 0.4, 4.6, 3.1, 4.4, 4, 2.1, 5, 3.3, 4.7, 3.4, 4.3, 4.5, 2.3, 0.5, 4.9, 3.5, 1.8, 1.9, 2.6, 4.3, 4.6, 5.2, 1.6, 2.8, 2.4, 2.8, 1.8, 3.6, 0.8, 5.1, 1.4, 3.2, 1, 6.3, 3.6, 3.6, 1.8, 0.9, 4.6, 2.5, 5.8, 0.6, 3.3, 3.2, 6.6, 2.6, 2.5, 1.5, 4.1, 1.7, 2.1, 1.5, 0.4, 4.8, 0.4, 1.5, 4.2, 3.3, 1.2, 8.1, 2.4, 2.8, 2.1, 6.3, 4.2, 1.3, 6, 1.3, 3.7, 2.5, 6.6, 2.7, 1.4, 2, 0.7, 4.3, 3.4, 4.3, 4, 4, 0.8, 2.3, 2.5, 5.4, 4.3, 0.5, 3.9, 2.2, 3.4, 1.3, 2.4, 4.7, 2, 1.3, 4.4, 2.9, 2.1, 2.5, 1.6, 2.3, 4.4, 1.9, 1.9, 1.7, 2, 4.2, 3.4, 3.9, 4.3, 1.3, 2.9, 2.2, 5.1, 2.3, 1.9, 2.9, 5.2, 3.4, 2.6, 2.4, 3.2, 1.3, 3.1, 5.1, 1.4, 4.2, 0.9, 1.3, 2.1, 2.6, 6.2, 1.6, 2.7, 1.7, 2.3, 3.3, 2.8, 1.2, 2.6, 1.5, 2, 2.8, 2.5, 2, 1.2, 2.2, 2.6, 2.5, 6, 1.9, 3, 3.8, 1.9, 3.2, 3.1, 1.8, 2.6, 1.9, 3.5, 3.7, 1.8, 2.2, 2, 1.3, 2, 1.1, 2.2, 3.1, 2.9, 1.3, 0.2, 3.9
Os membros da equipa conjecturam que X possui distribuição de Rayleigh com parâmetro de escala σ, i.e., com função de distribuição dada por

$$F_0(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x > 0.$$

Teste $H_0 : X \sim \text{Rayleigh}(\sigma = 2.4)$ contra $H_1 : X \nsim \text{Rayleigh}(\sigma = 2.4)$, procedendo do seguinte modo:

1. Fixe a semente em 5885 e selecione ao acaso e sem reposição uma subamostra de dimensão $n = 160$ da amostra original.
2. Divida o suporte da variável aleatória X, $\mathbb{R}^+$, em $k = 5$ classes equiprováveis sob H_0.
3. Agrupe as observações da subamostra selecionada em 1. nas classes definidas em 2. e obtenha o conjunto de frequências absolutas observadas associadas a essas classes.
4. Recorra às frequências absolutas observadas obtidas em 3. e calcule o valor-p do teste de ajustamento do qui-quadrado para as hipóteses referidas.

Com base neste procedimento, indique qual das cinco decisões abaixo deverá tomar a equipa de engenheiros.

a. Rejeitar H_0 aos n.s. de 5% e 10% e não rejeitar H_0 ao n.s. de 1%.

b. Rejeitar H_0 aos n.s. de 1%, 5% e 10%.

c. Rejeitar H_0 ao n.s. de 10% e não rejeitar H_0 aos n.s. de 1% e 5%.

d. Teste é inconclusivo.

e. Não rejeitar H_0 aos n.s. de 1%, 5% e 10%.

Código R

```
# Original data (200 observations)
original_data <- c(2.3, 2.7, 5.2, 0.7, 2.9, 0.6, 2.6, 2.2, 3.8, 0.5, 4.9, 5.4,
    3.7, 0.4, 4, 3.6, 2, 0.8, 2.5, 2.8, 1.7, 3.3,
```

```r
1.5, 0.4, 6.4, 1.5, 6, 2.1, 0.4, 4.6, 3.1, 4.4, 4, 2.1, 5, 3.3, 4.7, 3.4, 4.3,
    4.5, 2.3, 0.5, 4.9, 3.5,
1.8, 1.9, 2.6, 4.3, 4.6, 5.2, 1.6, 2.8, 2.4, 2.8, 1.8, 3.6, 0.8, 5.1, 1.4, 3.2,
    1, 6.3, 3.6, 3.6, 1.8,
0.9, 4.6, 2.5, 5.8, 0.6, 3.3, 3.2, 6.6, 2.6, 2.5, 1.5, 4.1, 1.7, 2.1, 1.5, 0.4,
    4.8, 0.4, 1.5, 4.2, 3.3,
1.2, 8.1, 2.4, 2.8, 2.1, 6.3, 4.2, 1.3, 6, 1.3, 3.7, 2.5, 6.6, 2.7, 1.4, 2, 0.7,
     4.3, 3.4, 4.3, 4, 4,
0.8, 2.3, 2.5, 5.4, 4.3, 0.5, 3.9, 2.2, 3.4, 1.3, 2.4, 4.7, 2, 1.3, 4.4, 2.9,
    2.1, 2.5, 1.6, 2.3, 4.4,
1.9, 1.9, 1.7, 2, 4.2, 3.4, 3.9, 4.3, 1.3, 2.9, 2.2, 5.1, 2.3, 1.9, 2.9, 5.2,
    3.4, 2.6, 2.4, 3.2, 1.3,
3.1, 5.1, 1.4, 4.2, 0.9, 1.3, 2.1, 2.6, 6.2, 1.6, 2.7, 1.7, 2.3, 3.3, 2.8, 1.2,
    2.6, 1.5, 2, 2.8, 2.5, 2,
1.2, 2.2, 2.6, 2.5, 6, 1.9, 3, 3.8, 1.9, 3.2, 3.1, 1.8, 2.6, 1.9, 3.5, 3.7, 1.8,
     2.2, 2, 1.3, 2, 1.1,
2.2, 3.1, 2.9, 1.3, 0.2, 3.9)

# Parameters
set.seed(5885)
n <- 160  # subsample size
sigma <- 2.4  # Rayleigh scale parameter
k <- 5  # number of classes

cat("Dados originais: ", length(original_data), "observacoes\n")
cat("Parametros do teste:\n")
cat("n =", n, "(tamanho da subamostra)\n")
cat("sigma =", sigma, "(parametro de escala Rayleigh)\n")
cat("k =", k, "(numero de classes)\n\n")

# 1. Select random subsample of size 160 without replacement
subsample_indices <- sample(1:length(original_data), n, replace = FALSE)
subsample <- original_data[subsample_indices]

cat("1. Subamostra selecionada (primeiros 10 valores):", head(subsample, 10), "
    ...\n")

# 2. Define k=5 equiprobable classes under H0
# For Rayleigh distribution with CDF F(x) = 1 - exp(-x^2/(2*sigma^2))
# Quantiles for equiprobable classes: 0.2, 0.4, 0.6, 0.8, 1.0

probabilities <- seq(0.2, 1.0, by = 0.2)
# Rayleigh quantile function: Q(p) = sigma * sqrt(-2 * log(1-p))
class_boundaries <- sigma * sqrt(-2 * log(1 - probabilities))

cat("2. Limites das classes equiprovaveis:\n")
cat("   (0,", round(class_boundaries[1], 3), "]\n")
for (i in 2:length(class_boundaries)) {
  cat("   (", round(class_boundaries[i-1], 3), ",", round(class_boundaries[i],
    3), "]\n")
}

# 3. Group observations into classes and get observed frequencies
observed_freq <- numeric(k)

for (i in 1:length(subsample)) {
  x <- subsample[i]
  if (x <= class_boundaries[1]) {
    observed_freq[1] <- observed_freq[1] + 1
  } else if (x <= class_boundaries[2]) {
    observed_freq[2] <- observed_freq[2] + 1
  } else if (x <= class_boundaries[3]) {
    observed_freq[3] <- observed_freq[3] + 1
```

```
  } else if (x <= class_boundaries[4]) {
    observed_freq[4] <- observed_freq[4] + 1
  } else {
    observed_freq[5] <- observed_freq[5] + 1
  }
}

cat("\n3. Frequencias absolutas observadas:\n")
for (i in 1:k) {
  cat("   Classe", i, ":", observed_freq[i], "\n")
}

# Expected frequency for each class (equiprobable under H0)
expected_freq <- n / k

cat("\nFrequencia esperada por classe:", expected_freq, "\n")

# 4. Chi-square goodness-of-fit test
chi_square_stat <- sum((observed_freq - expected_freq)^2 / expected_freq)
df <- k - 1  # degrees of freedom
p_value <- 1 - pchisq(chi_square_stat, df)

cat("\n4. Teste qui-quadrado:\n")
cat("Estatistica qui-quadrado:", chi_square_stat, "\n")
cat("Graus de liberdade:", df, "\n")
cat("Valor-p:", p_value, "\n")

# Decision at different significance levels
alpha_levels <- c(0.01, 0.05, 0.10)
decisions <- character(length(alpha_levels))

for (i in 1:length(alpha_levels)) {
  if (p_value < alpha_levels[i]) {
    decisions[i] <- "Rejeitar H0"
  } else {
    decisions[i] <- "Nao rejeitar H0"
  }
}

cat("\nDecisoes:\n")
cat("Nivel 1%:", decisions[1], "\n")
cat("Nivel 5%:", decisions[2], "\n")
cat("Nivel 10%:", decisions[3], "\n")

# Determine which option matches the results
if (all(decisions == "Nao rejeitar H0")) {
  answer <- "e"
  cat("\nResposta: e) Nao rejeitar H0 aos n.s. de 1%, 5% e 10%.\n")
} else if (all(decisions == "Rejeitar H0")) {
  answer <- "b"
  cat("\nResposta: b) Rejeitar H0 aos n.s. de 1%, 5% e 10%.\n")
} else if (decisions[1] == "Nao rejeitar H0" && decisions[2] == "Rejeitar H0" &&
     decisions[3] == "Rejeitar H0") {
  answer <- "a"
  cat("\nResposta: a) Rejeitar H0 aos n.s. de 5% e 10% e nao rejeitar H0 ao n.s.
     de 1%.\n")
} else if (decisions[1] == "Nao rejeitar H0" && decisions[2] == "Nao rejeitar H0
    " && decisions[3] == "Rejeitar H0") {
  answer <- "c"
  cat("\nResposta: c) Rejeitar H0 ao n.s. de 10% e nao rejeitar H0 aos n.s. de
    1% e 5%.\n")
} else {
  answer <- "d"
```

```
  cat("\nResposta: d) Teste e inconclusivo.\n")
}
```

Resultados

Solução

Para o teste de ajustamento da distribuição de Rayleigh com $\sigma = 2.4$:

Parâmetro/Resultado	**Valor**
Tamanho da subamostra (n)	160
Número de classes (k)	5
Frequência esperada por classe	32
Frequências Observadas por Classe	
Classe 1: (0, 1.603]	38
Classe 2: (1.603, 2.426]	37
Classe 3: (2.426, 3.249]	33
Classe 4: (3.249, 4.306]	27
Classe 5: $(4.306, +\infty)$	25
Teste Qui-Quadrado	
Estatística qui-quadrado	4.25
Graus de liberdade	4
Valor-p	0.373228

Decisões por Nível de Significância	
Nível 1%	Não rejeitar H_0
Nível 5%	Não rejeitar H_0
Nível 10%	Não rejeitar H_0

Resposta final: Com base no procedimento descrito, a equipa de engenheiros deverá escolher a opção $\boxed{\texttt{e}}$.

Justificação: O valor-p = 0.373 é superior a todos os níveis de significância testados (1%, 5% e 10%), pelo que não se rejeita H_0 em nenhum dos casos. Os dados são consistentes com a distribuição de Rayleigh proposta.