

8 Testes de hipóteses

- Nos dois capítulos anteriores vimos como usar os dados observados para determinar um valor ou um conjunto de valores plausíveis para um parâmetro cujo valor numa população é desconhecido.
- Agora o procedimento inverte-se: escolhido um valor ou um conjunto de valores possíveis para um parâmetro, vamos avaliar se os dados observados contrariam ou não essa escolha.
- Por fim, iremos analisar uma questão de uma natureza diferente – o ajustamento de uma distribuição a uma variável aleatória.

8.1 Testes de hipóteses paramétricas

Seja $\mathbf{X} = (X_1, \dots, X_n)$ uma amostra aleatória de uma população com distribuição $f_X(x; \theta)$, $\theta \in \Theta$.

Hipóteses paramétricas

Hipótese nula: $H_0 : \theta \in \Theta_0$

Hipótese alternativa: $H_1 : \theta \in \Theta_1$

com $\Theta = \Theta_0 \cup \Theta_1$ e $\Theta_0 \cap \Theta_1 = \emptyset$.

Tipos de hipóteses

Hipótese simples

- $H : \theta = \theta_0$

Hipóteses compostas

- $H : \theta \neq \theta_0$
- $H : \theta > \theta_0$ ou $H : \theta \geq \theta_0$
- $H : \theta < \theta_0$ ou $H : \theta \leq \theta_0$

Consideremos uma amostra observada de dimensão 30 com $\bar{x} = s^2 = 11$ de uma população $X \sim N(\mu, \sigma^2 = 9)$.

Pretende-se testar a hipótese de interesse

$$H_0 : \mu = 10$$

contra a alternativa

$$H_1 : \mu \neq 10$$

Mais precisamente, pretende-se avaliar se a amostra observada contém informação que contrarie significativamente a hipótese de interesse.

Já usámos atrás a variável fulcral

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Neste caso, isto quer dizer que

$$Z = \frac{\bar{X} - \mu}{\frac{3}{\sqrt{30}}} \sim N(0, 1),$$

em que μ representa o valor desconhecido da média da população.

Admitindo que H_0 é verdadeira obtemos a **estatística de teste**:

$$Z_0 = \frac{\bar{X} - 10}{\frac{3}{\sqrt{30}}} \sim N(0, 1)$$

Que valores de Z_0 se deverão encarar com surpresa, caso H_0 seja verdadeira?

A resposta à pergunta anterior permite estabelecer a forma da **região crítica** (ou região de rejeição de H_0), $RC \subset \mathbb{R}$, que é definida completamente de forma que

$$P(Z_0 \in RC \mid H_0) \leq \alpha,$$

em que α é uma probabilidade, em geral pequena, a que se chama o **nível de significância do teste**.

Para um **nível de significância** $\alpha = 0.05$ obtem-se a **região crítica**:

$$RC_{0.05} =]-\infty, -1.96[\cup]1.96, +\infty[,$$

ou, equivalentemente, a regra de decisão:

1. Rejeitar H_0 para $\alpha = 0.05$ se $|Z_0| > 1.96$;
2. Não rejeitar H_0 para $\alpha = 0.05$ no caso contrário.

Para a amostra observada tem-se

$$z_0 = \frac{11 - 10}{\frac{3}{\sqrt{30}}} = 1.826 \notin RC_{0.05},$$

e, portanto, não se rejeita H_0 ao nível de significância de 0.05 (ou menor).

Nota

Para as hipóteses consideradas, o teste de hipóteses aplicado é equivalente a averiguar se o valor 10 está ou não incluído no $IC_{0.95}(\mu)$.

$$\begin{aligned} |z_0| \leq 1.96 &\iff \left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \leq 1.96 \iff \\ &\iff \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Estatística de teste

Escolha de uma estatística de teste adequada $T(\mathbf{X})$ cuja distribuição sob H_0 seja conhecida e que, de alguma forma, permita avaliar o grau de discordância entre os dados e a hipótese H_0 .

Região crítica ou de rejeição de H_0

Admitindo que H_0 é verdadeira, define-se uma região $C \subset \mathbb{R}$ tal que a $P(T(\mathbf{X}) \in C \mid H_0) \leq \alpha$, em que α é uma probabilidade, em geral pequena, a que se chama o **nível de significância do teste**.

Concretização do teste de hipóteses

Observada uma amostra $\mathbf{x} = (x_1, \dots, x_n)$ calcula-se $T(\mathbf{x})$ e é tomada uma decisão:

- se $T(\mathbf{x}) \in C$ **rejeita-se** H_0 para o nível de significância α ;
- caso contrário **não se rejeita** H_0 para o mesmo nível de significância α .

Avaliação de um teste de hipóteses

Numa decisão sobre uma hipótese H_0 há dois erros possíveis:

Erro de tipo I: rejeitar H_0 quando H_0 é verdadeira

Erro de tipo II: não rejeitar H_0 quando H_0 é falsa

As respectivas probabilidades são:

$$\begin{aligned}\alpha(\theta) &= P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) \\ &= P(T(\mathbf{X}) \in C \mid \theta \in \Theta_0)\end{aligned}$$

$$\begin{aligned}\beta(\theta) &= P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa}) \\ &= P(T(\mathbf{X}) \notin C \mid \theta \in \Theta_1)\end{aligned}$$

Definição

A **potência** de um teste de hipóteses é definida por

$$P(\text{rejeitar } H_0 \mid \theta) = \begin{cases} \alpha(\theta), & \theta \in \Theta_0 \\ 1 - \beta(\theta), & \theta \in \Theta_1 \end{cases}$$

Uma vez que o teste foi definido à custa de

$$Z = \frac{\bar{X} - \mu}{\frac{3}{\sqrt{30}}} \sim N(0, 1)$$

podemos avaliar as probabilidades de cada um dos erros:

$$\alpha = P(\text{Rejeitar } H_0 \mid H_0 \text{ verdadeira})$$

$$\beta(\mu) = P(\text{Não rejeitar } H_0 \mid H_0 \text{ falsa})$$

Como deve ser a função potência do teste ideal?

Potência do teste ideal



$$P(\text{Rejeitar } H_0 \mid \mu) =$$

$$= P(|Z_0| > 1.96 \mid \mu) =$$

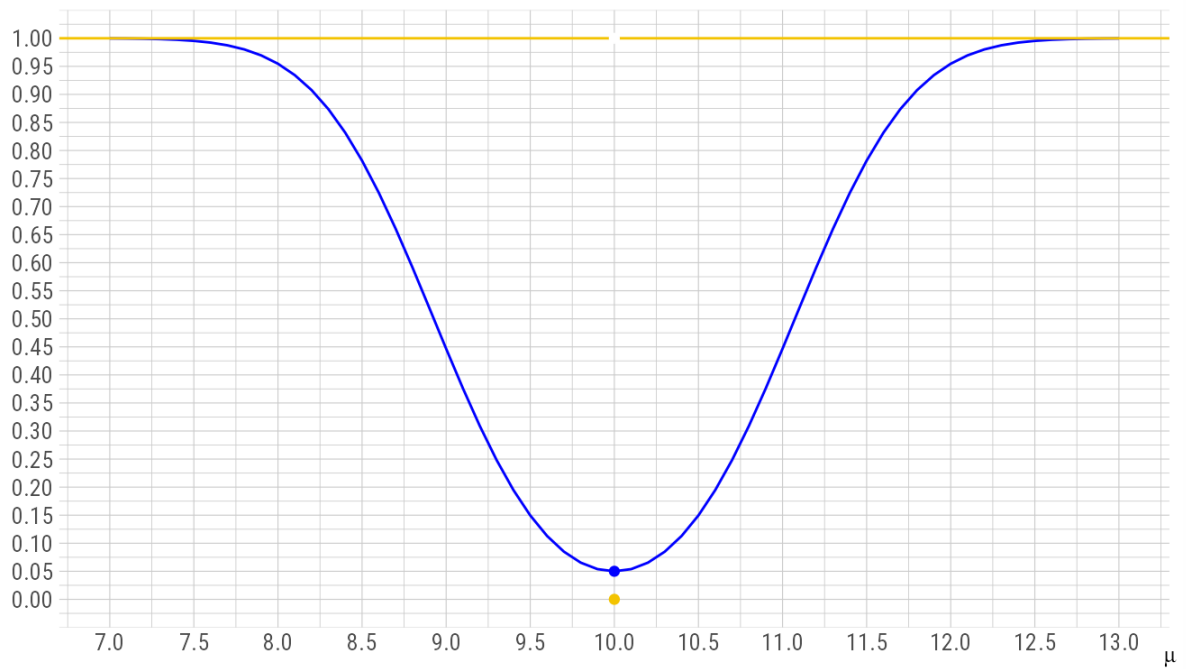
$$= 1 - P(|Z_0| \leq 1.96 \mid \mu) =$$

$$= 1 - \Phi\left(1.96 + \frac{10-\mu}{\frac{3}{\sqrt{30}}}\right) + \Phi\left(-1.96 + \frac{10-\mu}{\frac{3}{\sqrt{30}}}\right)$$

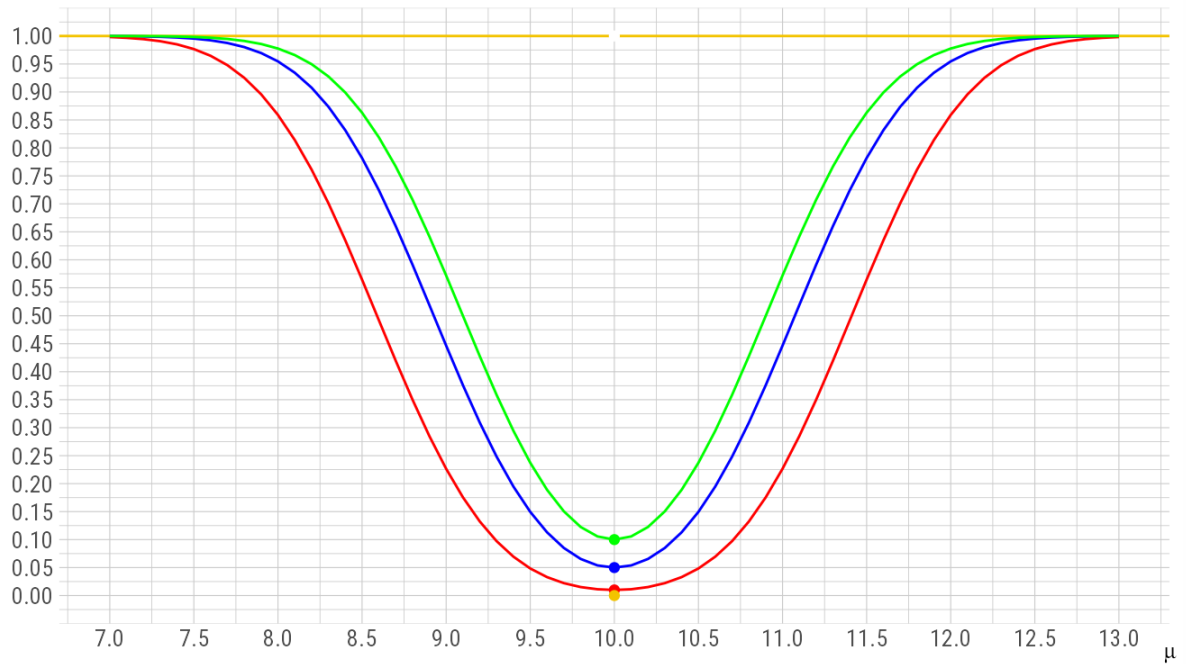
Nota

$$P(\text{Rejeitar } H_0 \mid \mu = 10) = 0.05$$

Potência do teste ($\alpha=0.05$)



Potência do teste (variando α)



Cálculo do valor-p

Definição

O **valor-p** é a probabilidade sob H_0 de a estatística de teste tomar valores mais desfavoráveis a H_0 do que o seu próprio valor observado.

Calculado o *valor-p*:

1. rejeita-se H_0 para níveis de significância superiores ao *valor-p*;
2. não se rejeita no caso contrário.

Para a amostra observada tem-se

$$z_0 = \frac{11 - 10}{\frac{3}{\sqrt{30}}} = 1.826$$

$$\text{Valor-p} = 2(1 - \Phi(1.826)) = 0.068$$

Apenas se deve rejeitar H_0 para um nível de significância superior a 0.068.

Considere-se agora a hipótese de interesse

$$H_0 : \mu \leq 10$$

contra a alternativa

$$H_1 : \mu > 10$$

Sob H_0 a estatística de teste é a mesma:

$$Z_0 = \frac{\bar{X} - 10}{\frac{3}{\sqrt{30}}} \stackrel{\mu=10}{\sim} N(0, 1)$$

Para um nível de significância $\alpha = 0.05$ obtem-se a região de rejeição de H_0 ou região crítica:

$$RC_{0.05} =]1.645, +\infty[,$$

ou, equivalentemente, a regra de decisão:

1. Rejeitar H_0 para $\alpha = 0.05$ se $Z_0 > 1.645$;
2. Não rejeitar H_0 para $\alpha = 0.05$ no caso contrário.

Para a amostra observada tem-se

$$z_0 = \frac{11 - 10}{\frac{3}{\sqrt{30}}} = 1.826 \in RC_{0.05},$$

e, portanto, rejeita-se H_0 ao nível de significância de 0.05.

O valor-p é, neste caso, dado por $1 - \Phi(1.826) = 0.034$.

Apenas se deve rejeitar H_0 para um nível de significância superior a 0.034.

Quando o valor de σ^2 é também desconhecido passamos a usar a variável fulcral

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{(n-1)}$$

A construção dos testes de hipóteses é inteiramente análoga aos casos anteriores.

Testes de hipóteses para a variância

Neste caso recorre-se à variável fulcral usada atrás para σ^2

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

No entanto, pode-se pensar que a assimetria da distribuição da estatística de teste introduz novas dificuldades.

Consideremos a construção de um teste para as hipóteses

$$H_0 : \sigma^2 = 9 \text{ contra } H_1 : \sigma^2 \neq 9$$

A estatística de teste é

$$Q_0 = \frac{29S^2}{9} \stackrel{\sigma^2=9}{\sim} \chi_{(29)}^2$$

Qual deverá ser a forma da região crítica?

Sob H_0 é de esperar que a estatística tome valores em torno de $E[Q_0] = 29$. Logo, valores que se afastem dessa medida de localização da distribuição $\chi_{(29)}^2$, quer para valores elevados quer para valores próximos de 0, fornecem evidência contra a hipótese nula, ou seja,

$$RC_\alpha = \left[0, F_{\chi_{(29)}^2}^{-1}(\alpha/2) \right] \cup \left[F_{\chi_{(29)}^2}^{-1}(1 - \alpha/2), +\infty \right]$$

Para $\alpha = 0.01$ tem-se

$$RC_{0.01} = [0, 13.121[\cup]52.336, +\infty[,$$

$q_0 = 35.444 \notin RC_{0.01}$ e, portanto, não se rejeita H_0 ao nível de significância de 0.01.

O valor-p é dado por

$$2 \times \min\{P(Q_0 \leq 35.444), P(Q_0 > 35.444)\} = 0.381$$

Testes de hipóteses em populações não normais uniparamétricas

Tal como na construção de intervalos de confiança, também aqui só iremos considerar testes de hipóteses aproximados baseados em estatísticas de teste obtidas pela aplicação do Teorema do Limite Central, ou seja,

$$\frac{\bar{X} - E[X]}{\sqrt{\frac{Var[X]}{n}}} \stackrel{a}{\sim} N(0, 1),$$

para n suficientemente grande.

Exemplo – Haja luz!

Um fabricante de lâmpadas afirma que o tempo médio de vida das suas lâmpadas é de 1000 ($= 10^3$) horas, no mínimo. Numa amostra de 120 lâmpadas retiradas ao acaso da produção desse fabricante observou-se um tempo total de vida de 112×10^3 horas.

Admitindo que o tempo de vida de uma lâmpada, em milhares de horas, segue uma distribuição exponencial, pretende-se avaliar a afirmação do fabricante.

Seja $\mathbf{X} = (X_1, \dots, X_n)$ uma amostra aleatória de uma população com distribuição $Exp(\lambda)$, $\lambda > 0$.

Hipóteses

$$H_0 : 1/\lambda \geq 1 \text{ contra } H_1 : 1/\lambda < 1$$

Estatística de teste

Como a dimensão da amostra é suficientemente grande podemos utilizar

$$T_0 = \sqrt{n} (\bar{X} - 1) \stackrel{a}{\sim} N(0, 1),$$

Região crítica

Tendo em conta as hipóteses e a estatística de teste, a região crítica deverá ter a forma

$$RC_{\alpha} =]-\infty, \Phi^{-1}(\alpha)[$$

Assim o *valor-p* é dado por $P(T_0 \leq t_0)$.

Concretização do teste de hipóteses

Como $\sum_{i=1}^{120} x_i = 112$ tem-se $t_0 = \sqrt{120} \left(\frac{112}{120} - 1 \right) \simeq -0.73$.

O *valor-p* é igual a $P(T_0 \leq -0.73) \simeq 1 - \Phi(0.73) = 1 - 0.7673 = 0.2327$.

Conclusões

1. deve-se rejeitar H_0 para níveis de significância superiores a 0.2327 e não rejeitar no caso contrário;
2. não há evidência suficiente para rejeitar a afirmação do fabricante de lâmpadas aos níveis de significância mais usuais ($\alpha \in [0.01, 0.1]$).

8.2 Teste de ajustamento do qui-quadrado

Nos procedimentos estatísticos que vimos até aqui admitiu-se que a distribuição da variável aleatória de interesse era conhecida a menos do valor de um ou mais parâmetros. Iremos agora encarar esse pressuposto como uma hipótese estatística cuja plausibilidade se pretende avaliar.



Serão as 4 tartarugas ninjas igualmente populares?

| i | TN_i | o_i |
|-----|---------------------|-----------|
| 1 | <i>Raphael</i> | 30 |
| 2 | <i>Donatello</i> | 15 |
| 3 | <i>Leonardo</i> | 20 |
| 4 | <i>Michelangelo</i> | 35 |
| | | $n = 100$ |

Defina-se p_i = “proporção de pessoas na população que preferem a TN_i ”, $i = 1, \dots, 4$.

$$H_0 : p_i = 1/4, \forall i$$

$$H_1 : \exists i : p_i \neq 1/4$$

| i | TN_i | o_i | p_i^0 | $e_i = np_i^0$ |
|-----|---------------------|-----------|---------|----------------|
| 1 | <i>Raphael</i> | 30 | 1/4 | 25 |
| 2 | <i>Donatello</i> | 15 | 1/4 | 25 |
| 3 | <i>Leonardo</i> | 20 | 1/4 | 25 |
| 4 | <i>Michelangelo</i> | 35 | 1/4 | 25 |
| | | $n = 100$ | 1 | $n = 100$ |

Para se comparar as frequências observadas, o_i , com as frequências esperadas sob H_0 , e_i , iremos recorrer à estatística do qui-quadrado de Pearson:

$$Q^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim} \chi_{(k-1)}^2.$$

Efetuada os cálculos, obtém-se o valor observado da estatística de teste, $q_0^2 = 10$.

$$\text{valor-p} = P(Q_0^2 > 10) = 1 - F_{\chi_{(3)}^2}(10) = 0.019$$

\therefore rejeita-se H_0 para n. s. superiores a 0.019 e não se rejeita no caso contrário.

Nota

Definindo a v. a. X = “TN preferida por uma pessoa escolhida ao acaso” temos que

$$H_0 \iff H_0^* : X \sim U(\{1, 2, 3, 4\})$$

Consideremos uma amostra aleatória de dimensão n extraída de uma população X com distribuição desconhecida. Pretende-se testar as hipóteses

$$H_0^* : X \sim f_X^0(x) \text{ contra } H_1^* : X \sim f_X^0(x)$$

Seja A_1, \dots, A_k uma partição do suporte de f_X^0 na qual se agrupam os dados observados.

Desse agrupamento obtém-se um vetor de frequências observadas (O_1, \dots, O_k) , em que O_i representa o número de observações na classe A_i , $i = 1, \dots, k$.

Sendo $p_i^0 = P(X \in A_i \mid H_0^*)$ temos que $O_i \sim Bi(n, p_i^0)$, sob H_0^* .

As frequências esperadas sob H_0^* são então

$$E_i = E[O_i \mid H_0^*] = np_i^0$$

Note-se que $\sum_{i=1}^k O_i = \sum_{i=1}^k E_i = n \sum_{i=1}^k p_i^0 = n$.

As hipóteses em confronto passam agora a ser:

$$H_0 : p_i = p_i^0, \forall i$$

$$H_1 : \exists i : p_i \neq p_i^0$$

Nota

Em geral, e ao contrário do exemplo inicial, tem-se apenas que $H_0^* \implies H_0$.

Uma forma de avaliar a plausibilidade de H_0 consiste em comparar as frequências observáveis, O_i , com as frequências esperadas sob H_0 , E_i . Para isso utiliza-se a estatística do qui-quadrado de Pearson

$$Q^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim} \chi_{(k-1)}^2$$

Para um nível de significância α a região crítica é, naturalmente,

$$RC_\alpha = \left[F_{\chi_{(k-1)}^2}^{-1}(1 - \alpha), +\infty \right[$$

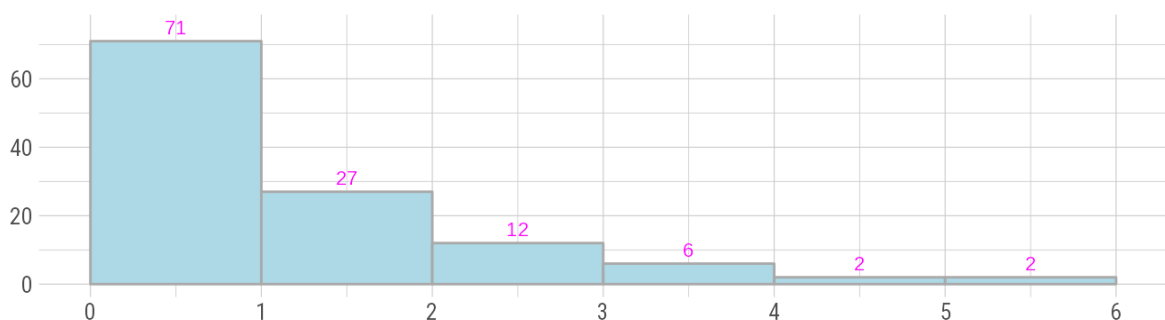
Uma vez que se trata de um teste aproximado é necessário que se verifiquem as seguintes condições:

1. todas as classes com $E_i \geq 1$;
2. pelo menos 80% das classes com $E_i \geq 5$.

Quando isto não acontece procede-se ao agrupamento de classes adjacentes.

Exemplo – Haja luz outra vez!

Voltemos ao exemplo das lâmpadas. Para os 120 tempos de vida de lâmpadas observados construiu-se o seguinte histograma de frequências absolutas:



Será razoável admitir que X tem uma distribuição $Exp(1)$?

| Classes | (0, 1] | (1, 2] | (2, 3] | (3, 4] | (4, 5] | (5, 6] | |
|---------|--------|--------|--------|--------|--------|--------|-----------|
| o_i | 71 | 27 | 12 | 6 | 2 | 2 | $n = 120$ |

$$H_0 : X \sim Exp(1)$$

$$H_1 : X \sim \text{Exp}(1)$$

As classes em que os dados estão agrupados cobrem o suporte da distribuição sob H_0 ?

Não, o suporte é \mathbb{R}^+ e é preciso ampliar a tabela.

| Classes | (0, 1] | (1, 2] | (2, 3] | (3, 4] | (4, 5] | (5, 6] | (6, +∞) | |
|---------|--------|--------|--------|--------|--------|--------|---------|-----------|
| o_i | 71 | 27 | 12 | 6 | 2 | 2 | 0 | $n = 120$ |

Seja

$$p_i^0 = P(X \in (i-1, i] \mid H_0), i = 1, \dots, 6 \text{ e}$$

$$p_7^0 = P(X > 6 \mid H_0) = 1 - \sum_{i=1}^6 p_i^0.$$

$$\text{Sob } H_0, F_X(x) = 1 - e^{-x}, x \geq 0.$$

Então

$$\begin{aligned} p_i^0 &= F_X(i) - F_X(i-1) = \\ &= e^{-i+1} - e^{-i}, i = 1, \dots, 6 \end{aligned}$$

| Classes | o_i | p_i^0 | $e_i = 120p_i^0$ |
|---------|-------|---------|------------------|
| (0, 1] | 71 | 0.6321 | 75.854 |
| (1, 2] | 27 | 0.2325 | 27.905 |
| (2, 3] | 12 | 0.0855 | 10.266 |
| (3, 4] | 6 | 0.0315 | 3.777 |
| (4, 5] | 2 | 0.0116 | 1.389 |
| (5, 6] | 2 | 0.0043 | 0.511 |
| (6, +∞) | 0 | 0.0025 | 0.297 |
| | 120 | 1 | 120 |

Há classes com frequências demasiado pequenas!

É necessário agrupar essas classes.

| Classes | o_i | p_i^0 | $e_i = 120p_i^0$ | $(o_i - e_i)^2/e_i$ |
|---------|-------|---------|------------------|---------------------|
| (0, 1] | 71 | 0.6321 | 75.852 | 0.3104 |
| (1, 2] | 27 | 0.2325 | 27.900 | 0.0290 |
| (2, 3] | 12 | 0.0855 | 10.260 | 0.2951 |
| (3, +∞) | 10 | 0.0499 | 5.988 | 2.6881 |
| | 120 | 1 | 120 | $q^2 = 3.3226$ |

Para um nível de significância $\alpha = 0.05$, a região crítica é $RC_{0.05} =]7.815, +\infty[$.

Como $q^2 = 3.3226 \notin C_{0.05}$, não há evidência para rejeitar a hipótese nula para $\alpha \leq 0.05$.

Alternativa

valor-p = $P(Q^2 > 3.3226) = 0.3445$

($0.30 < \text{valor-p} < 0.40$, usando a tabela)

8.3 Testes de hipóteses no R

A maior parte dos testes de hipóteses que vimos são pouco mais do que simples exemplos educativos e, por isso, no R apenas podemos encontrar:

1. o teste para o valor esperado de uma população normal

```
t.test(x, ...)
```

2. o teste de ajustamento do χ^2

```
chisq.test(x, p = rep(1/length(x), length(x)), ...)
```

Exemplo – Haja luz ainda outra vez!

```
obs <- c(71, 27, 12, 10)
p0 <- c(0.6321, 0.2325, 0.0855, 0.0499)

chisq.test(obs, p = p0)
```

Chi-squared test for given probabilities

```
data: obs
X-squared = 3.323, df = 3, p-value = 0.345
```

De notar que a função `chisq.test` devolve um objeto com mais informação sobre o teste de hipóteses.