MAE0116 – Noções de Estatística
Grupos B e D – 2º semestre de 2020
Aula de revisão 2

Estimação

- **Parâmetro:** valor desconhecido da população caracterizada pela v.a. X.
- **Amostra aleatória de tamanho n de X (população):** $X_1, X_2, ...X_n$, sendo X_i a variável de interesse para o i-ésimo indivíduo selecionado, ao acaso, da população X.
- **Estimador:** função dos elementos da amostra usada para estimar o parâmetro de interesse de X.
- **Estimativa**: valor numérico do estimador, para a amostra selecionada.
- **Distribuição amostral do estimador**: Distribuição de probabilidade do estimador, utilizado para estimar o parâmetro desconhecido.

Estimação para μ

- **Estimador pontual para μ:**

$$\overline{X} = \frac{X_1 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ (média amostral)}$$

- **Estimativa pontual para** μ:

$$\overline{x} = \frac{x_1 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$x_1, ..., x_n$ são os valores observados de $X_1, ..., X_n$ para a amostra selecionada.

Resultados para $\overline{X}$

1. Seja X uma v.a. com $E(X) = \mu$ e $Var(X) = \sigma^2$, vale:

$$E(\overline{X}) = \mu \quad \text{e} \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

Obs: $DP(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ é chamado erro padrão da média amostral.

---

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

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\Downarrow$$

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

3. $X \sim N(\mu, \sigma^2)$

Vale também que

$$T = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

sendo t_{n-1} a distribuição t-Student com $n-1$ graus de liberdade e S^2 é a variância amostral.

Intervalo de confiança para μ com coeficiente de confiança γ

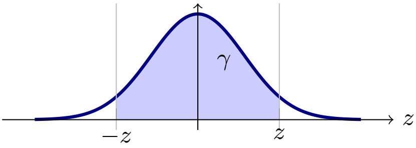
(i) $X \sim N(\mu, \sigma^2)$, σ^2 **conhecida**

$$IC(\mu,\gamma) = \left[\overline{X} - \epsilon\ ;\ \overline{X} + \epsilon\right]$$
$$= \left[\overline{X} - z\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z\frac{\sigma}{\sqrt{n}}\right]$$

sendo z tal que $\gamma = P(-z \leq Z \leq z)$ e

$\epsilon = z\dfrac{\sigma}{\sqrt{n}}$ o **erro amostral**

(ou margem de erro) .



(ii) $X \sim N(\mu, \sigma^2), \sigma^2$ **desconhecida**

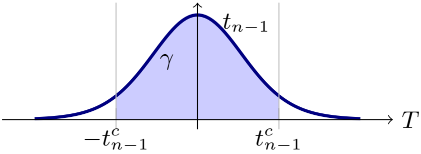
$$IC(\mu,\gamma) = \left[\overline{X} - \epsilon\ ;\ \overline{X} + \epsilon\right]$$
$$= \left[\overline{X} - t^c_{n-1}\frac{S}{\sqrt{n}} ; \overline{X} + t^c_{n-1}\frac{S}{\sqrt{n}}\right]$$

sendo t^c_{n-1} tal que

$P(-t^c_{n-1} \leq t_{n-1} \leq t^c_{n-1}) = \gamma$

$S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ a variância amostral

$\epsilon = t^c_{n-1}\dfrac{S}{\sqrt{n}}$ o **erro amostral** .



Resultado para $\overline{X}$ (cont.)

4. TEOREMA DO LIMITE CENTRAL (TLC)

Se X tem $E(X) = \mu$ e $Var(X) = \sigma^2$, para <u>n suficientemente grande</u> ,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \text{ aproximadamente.}$$

- O intervalo de confiança para μ, com coeficiente de confiança aproximadamente γ, é
  - **para σ conhecido:** $IC(\mu;\gamma) = \left[\overline{X} - z\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z\frac{\sigma}{\sqrt{n}}\right]$
  - **para σ desconhecido:** $IC(\mu;\gamma) = \left[\overline{X} - z\frac{S}{\sqrt{n}}\ ;\ \overline{X} + z\frac{S}{\sqrt{n}}\right]$

  z é tal que $\gamma = P(-z \leq Z \leq z)$, $Z \sim N(0,1)$.

Exercício 1

A FIESP está preocupada com o tempo perdido com acidentes de trabalho. Sabe-se que o tempo perdido em acidentes de trabalho tem uma distribuição normal com desvio padrão de 20 horas/homem por ano. Algumas novas medidas de segurança foram protocoladas e deseja-se estimar o tempo médio perdido com acidentes de trabalho após 1 ano da adoção das novas medidas, considerando que o desvio padrão não se alterou. Para tanto, tomou-se uma amostra de 16 indústrias e observaram-se,para essa amostra, os seguintes números de horas/homem perdidas por acidentes de trabalho:

$$55;\ 40;\ 52;\ 19;\ 17;\ 56;\ 43;\ 42;\ 56;\ 20;\ 18;\ 80;\ 58;\ 77;\ 52;\ 68.$$

(a) Construa um intervalo de confiança para o número médio de horas/homem perdidas por acidentes de trabalho nas indústrias do Estado de São Paulo, após novo protocolo de segurança. Use coeficiente de confiança de 95%. Qual é a margem de erro desse intervalo?

X: tempo perdido em acidentes de trabalho $\Rightarrow X \sim N(\mu, 20^2)$, μ desconhecido (σ conhecido)

Objetivo: estimar μ, o tempo médio perdido com acidentes de trabalho nas indústrias de São Paulo.

Estimador pontual para μ é $\overline{X}$

Intervalo de confiança para μ com 95% de confiança é dado por:

$$IC(\mu\ ;\ 95\%) = \left[\overline{X} - \epsilon\ ;\ \overline{X} + \epsilon\right] = \left[\overline{X} - z\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z\frac{\sigma}{\sqrt{n}}\right]$$

Na amostra:

$n = 16 \Rightarrow \overline{x} = \dfrac{55 + 40 + ... + 68}{16} = 47,06$ (estimativa pontual para μ).

$\gamma = 0,95 \Rightarrow z = 1,96 \quad (A(z) = 0,975)$

$\sigma = 20$

Erro amostral:

$$\epsilon = 1,96 \frac{20}{\sqrt{16}} = 1,96 \times 5 = 9,8$$

Assim, temos

$$\begin{aligned} IC(\mu \ ; \ 95\%) &= [47,06 - 9,80 \ ; \ 47,06 + 9,80] \\ &= [37,26 \ ; \ 56,86] \quad , \end{aligned}$$

com margem de erro $\epsilon = 9,80$ horas/homem/ano.

Exercício 1

(b) Qual deveria ser o tamanho da amostra para que o intervalo de confiança de 95% tivesse amplitude de 5 unidades?

Dimensionamento da amostra

Do erro amostral $\epsilon = z\dfrac{\sigma}{\sqrt{n}}$,

segue que, para γ e margem de erro ϵ fixados,

$$\boxed{n = \left(\frac{z}{\epsilon}\right)^2 \times \sigma^2 \ ,}$$

sendo z tal que $P(-z \leq Z \leq z) = \gamma$.

Obs.: A **amplitude** (ou comprimento) do intervalo de confiança é 2ϵ.

Amplitude $= 2\epsilon \Rightarrow 5 = 2\epsilon \Rightarrow \epsilon = 2,5$ horas/homem.

Então,

$$n = \left(\frac{1,96}{2,5}\right)^2 \times 20^2 = 245,86 \cong 246 \ \text{ indústrias.}$$

Observar que quando o valor de ϵ decresce, n aumenta (em comparação ao item (a): $n = 16 \Rightarrow \epsilon = 9,8$).

Estimação da proporção p populacional

- Estimador pontual para p:

$$\widehat{p} = \frac{X_1 + ... + X_n}{n} \text{ (proporção amostral)}$$

sendo $X_1, X_2, ..., X_n$ uma amostra aleatória de tamanho n de $X \sim Bernoulli(p)$.

- Estimativa pontual para p:

$$\widehat{p}_{obs} = \frac{k}{n} \; ,$$

sendo k o número de elementos na amostra com a característica de interesse.

Dimensionamento da amostra
- proporção p -

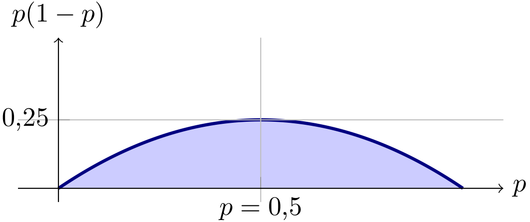
Do erro amostral $\epsilon = z\sqrt{\frac{p(1-p)}{n}}$, segue que, dado γ e margem de erro ϵ fixados,

$$\boxed{n = \left(\frac{z}{\epsilon}\right)^2 p(1-p)}$$

Observação:

- p é desconhecido
- Usar comportamento da função $p(1-p)$ para obter o valor de $\underline{n}$.

Dimensionamento da amostra
- proporção p -



- Sem informação de p:

  valor máximo de $p(1-p)$ é $0,25$ $(p = 0,5) \Rightarrow n = \left(\frac{z}{\epsilon}\right)^2 \times 0,25$

- Com informação de p: usar o valor máximo no intervalo informado.

  Exemplos:

$$p < 0,20 \Rightarrow p(1-p) \leq 0,2 \times 0,8 = 0,16 \Rightarrow n = \left(\frac{z}{\epsilon}\right)^2 \times 0,16$$

$$p > 0,70 \Rightarrow p(1-p) \leq 0,7 \times 0,3 = 0,21 \Rightarrow n = \left(\frac{z}{\epsilon}\right)^2 \times 0,21$$

$$0,35 < p < 0,65 \Rightarrow p(1-p) \leq 0,25 \ (p = 0,5 \text{ pertence ao intervalo})$$

Exercício 2

Um administrador de empresas, em fase de quedas nas vendas, está interessado em estimar a proporção p de funcionários de suas indústrias que são favoráveis à redução da jornada de trabalho, com alguma redução de salário, afim de evitar possíveis demissões.

(a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja de no máximo $0,03$, com probabilidade de $0,92$.

Seja p: proporção de funcionários das indústrias favoráveis à redução da jornada de trabalho.

Objetivo: estimar p

$n = ?$ tal que $\epsilon = 0,03$

$$\gamma = 0,92 \Rightarrow z = 1,75 \ (A(z) = 0,96)$$

$$n = \left(\frac{z}{\varepsilon}\right)^2 \times p(1-p)$$

Como nada se sabe sobre p, usa-se o valor máximo de $p(1-p)$, ou seja, $p = 0,5$.

Assim,

$$n = \left(\frac{1,75}{0,03}\right)^2 \times 0,25 = 850,69.$$

Logo, o tamanho da amostra deve ser $n = 851$ funcionários.

Exercício 2

(b) Suponha que o administrador acredite que essa proporção esteja entre 70% e 80%. Com essa informação seria possível diminuir o tamanho amostral? Se sim, qual deve ser o novo tamanho da amostra?

$0,7 \leq p \leq 0,8 \Rightarrow p(1-p) \leq 0,7 \times 0,3 = 0,21$

Assim, utilizaremos o valor máximo de $p(1-p)$, no intervalo informado.

Então,

$$n = \left(\frac{1,75}{0,03}\right)^2 \times 0,21 = 714,6.$$

Logo, com a informação sobre p, o novo tamanho de amostra necessário é $n = 715$ funcionários, produzindo uma redução de $851 - 715 = 136$ funcionários a serem selecionados para consulta.

Exercício 2

(c) Por limitação de tempo e custo, uma pesquisa foi realizada com 180 funcionários sorteados das indústrias, sendo que 130 foram favoráveis à nova política proposta. Qual é uma estimativa pontual e intervalar para a proporção de funcionários p dessas indústrias que são favoráveis à redução da jornada de trabalho, com redução de salário? Utilize coeficiente de confiança de 92%

Intervalo de confiança para a proporção p

$$X \sim Bernoulli(p), \text{ com } \sigma^2 = p(1-p) \text{ e } \widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$IC(p\ ;\ \gamma) = \left[\widehat{p} - z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\ ;\ \widehat{p} + z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right],$$

sendo z tal que $\gamma = P(-z \leq Z \leq z)$, com $Z \sim N(0,1)$.

Na amostra:

$n = 180 \Rightarrow k = 130$ funcionários declaram ser favoráveis à redução da jornada de trabalho.

Estimativa pontual para p: $\widehat{p}_{obs} = \dfrac{k}{n} = \dfrac{130}{180} \cong 0,722.$

Estimativa intervalar para p:

$$[\widehat{p} - \epsilon \ \ ; \ \ \widehat{p} + \epsilon] = \left[\widehat{p} - z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \ \ ; \ \ \widehat{p} + z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

$\gamma = 0,92 \Rightarrow z = 1,75 \ (A(z) = 0,96).$

Então,

$$\begin{aligned} IC(p;0,92) &= \left[0,722 - 1,75\sqrt{\frac{0,722(1-0,722)}{180}}; 0,722 + 1,75\sqrt{\frac{0,722(1-0,722)}{180}}\right] \\ &= [0,722 - 0,0584; 0,722 + 0,0584] \\ &= [0,6636; 0,7804]\,, \end{aligned}$$

sendo o erro amostral $\epsilon = 0,0584$.

Exercício 3

Em uma floricultura de São Paulo as mudas de uma espécie de orquídea são disponibilizadas para venda 5 semanas após a semeação. Quanto maior a muda, maior o preço pelo qual ela pode ser comercializada. A altura das orquídeas nesse período da vida pode ser modelada por uma distribuição Normal com média $\mu = 15$ cm. Um procedimento mais barato de adubação para mudas foi adotado e deseja-se saber se a altura das orquídeas diminui quando adubadas por meio desse procedimento. Para tanto, 10 orquídeas foram sorteadas 5 semanas após serem semeadas e adubadas pelo procedimento mais barato, e suas medidas de altura em centímetros foram as seguintes:

14,5 12,7 14,9 13,8 15,1 12,2 15,1 14,7 15,2 13,8

(a) Formule o problema como um problema de testes de hipóteses. Especifique o parâmetro a ser testado.

Denote por X a altura das orquídeas quando adubadas por meio do novo procedimento.

Pelo enunciado,

$$X \sim \mathrm{N}(\mu, \sigma^2),$$

com σ^2 desconhecido.

O parâmetro de interesse é μ, a altura média das orquídeas quando adubadas por meio do novo procedimento.

As hipóteses de interesse são:

H_0: $\mu = 15$ cm (a altura média das orquídeas não mudou com o novo procedimento de adubagem).

H_1: $\mu < 15$ cm (a altura média das orquídeas diminui com o novo procedimento de adubagem).

Exercício 3

(b) Calcule o nível descritivo do teste e conclua para um nível de significância de 5%.

Sabemos que $X \sim \mathrm{N}(\mu, \sigma^2)$, com σ^2 desconhecido. Recorde que para testar a média μ, utilizamos a média amostral $\overline{X}$ como evidência amostral.

Pelo resultado apresentado na Aula 10, a distribuição de $\overline{X}$ é dada por

$$\overline{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right) \overset{\sigma^2 \; desconh}{\Longrightarrow} T = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

Da amostra temos que $\overline{x} = 14,2$ e $s^2 = 1,1133$.

Como a hipótese alternatia é $H_1 : \mu < 15$, o nível descritivo do teste é dado por:

$$\begin{aligned} \text{valor-}p &= P(\overline{X} \leqslant \overline{x}, \text{ sendo } H_0 \text{ verdadeira }) = P(\overline{X} \leqslant 14,2, \text{ sendo } \mu = 15) \\ &= P\left(T \leqslant \frac{14,2 - 15}{\sqrt{\frac{1,1133}{10}}}\right) = P(T \leqslant -2,3976) = 0,0200276, \end{aligned}$$

sendo $T \sim t_9$.

Para calcular $P(T \leqslant -2,3976)$ há 2 maneiras:

(i) Pelo Rcmdr: clique no menu Distribuições $\rightarrow$ Distribuições Contínuas $\rightarrow$ Distribuição t $\rightarrow$ Probabilidades da distribuição t. Em Valores da variável coloque -2.3976, em Graus de liberdade coloque $9\ (= 10 - 1)$ e clique em cauda inferior.

(ii) Pela tabela $t-$Student: na linha 9 (graus de liberdade) o valor $2,398$ está bem próximo de $2,3976$ que corresponde a probabilidade, na cauda superior, a 0,98. Portanto, pela simetria da distribuição T, a cauda inferior tem probabilidade aproximada de 0,02.

Decisão e conclusão: Como o nível descritivo do teste é menor do que o nível de significância considerado de 5%, rejeitamos H_0. Temos evidências suficientes de que o novo procedimento de adubação diminui a altura média das orquídeas.

Exercício 3

(c) Construa um intervalo de confiança para a altura média de orquídeas adubadas segundo o novo procedimento, 5 semanas após a semeação, com coeficiente de confiança de 90%.

O intervalo de confiança para μ com coeficiente de confiança γ, sendo σ^2 desconhecido, é dado por:

$$\mathrm{IC}(\mu;\gamma) = \left[\overline{X} - t^c_{n-1}\frac{S}{\sqrt{n}};\;\; \overline{X} + t^c_{n-1}\frac{S}{\sqrt{n}}\right].$$

sendo t^c_{n-1} tal que $P(-t^c_{n-1} \leq T \leq t^c_{n-1}) = \gamma$, $T \sim t_{n-1}$.

Temos $\gamma = 0,90 \rightarrow t_9 = 1,833$. Logo,

$$\begin{aligned}
\mathrm{IC}(\mu;90\%) &= \left[14,2 - 1,833\frac{1,0551}{\sqrt{10}};\;\; 14,2 + 1,833\frac{1,0551}{\sqrt{10}}\right] \\
&= [14,2 - 0,612;\;\; 14,2 + 0,612] \\
&= [13,588;\;\; 14,813]
\end{aligned}$$

Exercício 4

Um fabricante de inseticida afirma no rótulo da embalagem que sua eficiência é de 75% contra todos os tipos de insetos. Para verificar essa afirmação, um agente da Secretaria de Defesa do Consumidor encomenda a um laboratório um teste de toxicidade do inseticida. O laboratório examinou 60 insetos expostos ao inseticida.

(a) Formule o problema como um teste de hipóteses. Especifique o parâmetro a ser testado.

Seja p a proporção de insetos mortos quando expostos ao inseticida. As hipóteses de interesse são:

H_0: $p = 0,75$ (a eficiência do inseticida contra todos os tipos de insetos, ou seja, a proporção de insetos mortos quando expostos ao inseticida, é igual a informada no rótulo da embalagem).

H_1: $p < 0,75$ (a eficiência do inseticida contra todos os tipos de insetos, ou seja, a proporção de insetos mortos quando expostos ao inseticida, é menor do que a informada no rótulo da embalagem).

Exercício 4

(b) Interprete os erros tipo I e tipo II.

Erro tipo I: Rejeitar H_0 quando H_0 é verdadeira, ou seja, concluir que a eficiência do inseticida contra todos os tipos de insetos é menor do que 0,75, quando na verdade é 0,75.

Erro tipo II: Não rejeitar H_0 quando H_0 é falsa, ou seja, concluir que a eficiência do inseticida contra todos os tipos de insetos é 0,75, quando na verdade é menor do que 0,75.

Exercício 4

(c) Se 40 dos 60 insetos expostos ao inseticida foram mortos, qual é a conclusão em relação a afirmação do fabricante, ao nível de 3% de significância, com base no cálculo do nível descritivo?

A estimativa pontual para p é dada por

$$\widehat{p}_{obs} = \frac{40}{60} = 0,67.$$

Para testar as hipóteses

$$H_0 : p = 0,75 \quad \text{versus} \quad H_1 : p < 0,75,$$

o nível descritivo do teste é dado por

$$\text{valor-}p = P(\widehat{p} < \widehat{p}_{obs}, \text{ sendo } H_0 \text{ verdadeira }) = P(\widehat{p} < 0,67, \text{ sendo } p = 0,75).$$

Pelo Resultado 4 (aulas 8 e 9) a proporção amostral $\widehat{p}$ tem distribuição aproximadamente Normal:

$$\widehat{p} \approx \text{Normal}\left(p, \frac{p(1-p)}{n}\right) \quad \Longrightarrow \quad Z = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx \text{Normal}(0,1).$$

Assim,

$$\begin{aligned}
\text{valor-}p &= P(\widehat{p} \leqslant 0,67, \text{ sendo } p = 0,75) \approx P\left(Z \leqslant \frac{0,67 - 0,75}{\sqrt{\frac{0,75\times(1-0,75)}{60}}}\right) \\
&= P\left(Z \leqslant \frac{-0,08}{0,055902}\right) = P(Z \leqslant -1,43018) \\
&= 1 - A(1,43) = 1 - 0,9236 = 0,0764
\end{aligned}$$

Como o nível descritivo do teste é maior do que o nível de significância considerado de 3%, não rejeitamos H_0. Ou seja, concluímos que não temos evidências suficientes de que a proporção de insetos mortos quando expostos ao inseticida seja diferente da informada no rótulo da embalagem.

Exercício 5

O gerente de uma indústria de carnes enlatadas tem estabelecido um padrão que diz que, de um novilho de 400 kg (peso vivo), a indústria obteria uma quantidade média de 250 kg de carne (cortada e enlatada). Para determinar se o padrão está sendo mantido, o gerente seleciona uma amostra aleatória de 100 novilhos pesando 400 kg e obtém uma quantidade média de carne de 253 kg com um desvio padrão igual a 18 kg.

(a) Faça o teste conveniente para verificar se o padrão de carne fornecida por novilhos de 400 kg está sendo mantido, utilizando um nível de significância de 10%.

Defina a seguinte variável aleatória

$$X : \text{Peso da carne cortada e enlatada obtido de um novilho de 400 kg},$$

e denote por μ e σ^2 a média e a variância de X, respectivamente. Pelo enunciado, não conhecemos a distribuição de X.

O parâmetro de interesse é μ, o peso médio da carne cortada e enlatada obtido de um novilho de 400 kg.

As hipóteses de interesse são:

H_0: $\mu = 250$ kg (o padrão está sendo mantido).

H_1: $\mu \neq 250$ kg (o padrão foi alterado).

Do enunciado, o tamanho amostral foi de $n = 100$ novilhos pesando 400kg, a média amostral foi igual a $\overline{x} = 253$ kg, e o desvio padrão foi de $s = 18$ kg. Assumindo que $n = 100$ é um tamanho amostral suficientemente grande, podemos aproximar a distribuição da média amostral pela distribuição normal utilizando o Teorema Central do Limite, de forma que

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{aproximadamente,}$$

e

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim \mathrm{N}\left(0, 1\right), \quad \text{aproximadamente.}$$

Neste caso, como o teste é bilateral e $\overline{x} = 253 > \mu = 250$, o nível descritivo é dado por

$$
\begin{aligned}
\text{valor-}p &= 2 \times P(\overline{X} \geqslant \overline{x}, \text{ sendo } H_0 \text{ verdadeira }) \\
&= 2 \times P(\overline{X} \geqslant 253, \text{ sendo } \mu = 250) = 2 \times P\left(Z \geqslant \frac{253 - 250}{\sqrt{\frac{18^2}{100}}}\right) \\
&= 2 \times P\left(Z \geqslant \frac{3}{1,8}\right) = 2 \times P(Z \geqslant 1,67) \\
&= 2 \times (1 - A(1,67)) = 2 \times (1 - 0,9525) = 2 \times 0,0475 = 0,095.
\end{aligned}
$$

Conclusão: Como o nível descritivo do teste é menor do que o nível de significância adotado de 10%, rejeitamos H_0, ou seja, temos evidências de que o padrão foi alterado.

Exercício 5

(b) Se a hipótese nula for rejeitada, construa um IC para μ utilizando um coeficiente de confiança de 90%.

O intervalo de confiança para μ com coeficiente de confiança γ, utilizando o Teorema Central do Limite, é dado por:

$$\mathrm{IC}(\mu;\gamma) = \left[\overline{X} - z\frac{S}{\sqrt{n}};\;\; \overline{X} + z\frac{S}{\sqrt{n}}\right],$$

em que z tal que $P(-z \leq Z \leq z) = \gamma$, sendo $Z \sim \mathrm{N}(0,1)$.

Temos da amostra que $\overline{x} = 253$, $s = 18$ e $\gamma = 0,90 \rightarrow z = 1,64$. Logo, o intervalo de confiança para μ com 90% de confiança é dado por:

$$\begin{aligned}\mathrm{IC}(\mu;90\%) &= \left[253 - 1,64\frac{18}{\sqrt{100}};\;\; 253 + 1,64\frac{18}{\sqrt{100}}\right]\\ &= [253 - 2,952;\;\; 253 + 2,952]\\ &= [250,048;\;\; 255,952]\end{aligned}$$

Portanto, com confiança de 90%, o padrão foi alterado para cima, ou seja, agora a industria obtém uma quantidade maior de carne por novilho de 400 kg.

Exercício 6

Uma emissora de TV desconfia da qualidade do método utilizado por um instituto para medir a audiência de programas de TV. Tal instituto aponta que em um determinado horário a emissora A tem 37% da audiência, enquanto que a emissora B tem 30%, a C tem 13% e as demais têm 20%. A emissora contratou uma empresa de pesquisa de mercado que selecionou uma amostra de 300 residências. Em cada uma, perguntou-se em qual canal a principal TV da casa estava sintonizada, na última semana, no horário determinado. Dos 300, 95 declararam estar assistindo a emissora A, 87 a emissora B, 51 a C e 67 uma das demais emissoras.

(a) Dê uma estimativa para a proporção de residências em que a principal TV da casa estava sintonizada na emissora A ou na emissora C, na última semana, no horário determinado.

A proporção de residências em que a principal TV da casa estava sintonizada na emissora A ou na emissora C é

$$\frac{95 + 51}{300} = 0,4867.$$

Exercício 6

(b) Se o percentual de audiência apontado pelo instituto estiver correto, quantas residências seriam esperadas estar sintonizadas na emissora B, na última semana, no horário determinado, entre as 300 pesquisadas?

Segundo apontado pelo instituto, é esperado que $0{,}30 \times 300 = 90$ das 300 residências estariam sintonizadas na emissora B.

Exercício 6

(c) Qual seria o teste de hipóteses adequado para verificar se os dados da amostra estão ou não de acordo com as porcentagens apontadas pelo instituto? Escreva as hipóteses H_0 e H_1 e calcule o número de graus de liberdade da estatística do teste.

Podemos descrever as hipóteses da seguinte maneira:

$$\begin{aligned}
\mathrm{H}_0 &: \text{A audiência das emissoras de TV em um determinado horário} \\
&\quad \text{se distribuem conforme afirma o instituto,} \\
\mathrm{H}_1 &: \text{A audiência das emissoras de TV em um determinado horário} \\
&\quad \text{não se distribuem conforme afirma o instituto,}
\end{aligned}$$

Mais especificamente,

$$\begin{aligned}
&\mathrm{H}_0 : p_1 = 0,37,\ p_2 = 0,30,\ p_3 = 0,13 \ \text{ e } p_4 = 0,20, \\
&\mathrm{H}_1 : \text{Existe ao menos uma diferença,}
\end{aligned}$$

em que p_1, p_2, p_3 e p_4 denotam as audiências das emissoras A, B, C e demais, respectivamente.

Para testar estas hipóteses, o teste adequado é o **teste de aderência** no qual precisamos calcular o valor da estatística do teste

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

em que O_i e E_i, $i = 1, 2, 3, 4$, são as frequências observadas e esperadas de cada possível categoria (A, B, C e demais emissoras).

Sob H_0, $E_i = n \times p_i$ e

$$\chi^2 \sim \chi^2_q$$

em que $q = k - 1$ graus de liberdade, para $k = 4$.

Logo, o número de graus de liberdade da estatística de teste é $q = 3$.

Exercício 6

(d) Calcule o nível descritivo (valor-p) do teste e, com base nele, dê sua conclusão ao nível de significância de 10%.

As frequências observadas e esperadas são

Emissora	Frequência Observada (O_i)	Frequência esperada sob H_0 (E_i)
A	95	$E_1 = n \times p_1 = 300 \times 0,37 = 111$
B	87	$E_2 = n \times p_2 = 300 \times 0,30 = 90$
C	51	$E_3 = n \times p_3 = 300 \times 0,13 = 39$
Demais	67	$E_4 = n \times p_4 = 300 \times 0,20 = 60$
Total	300	300

Observe que $E_i > 5$ para todo $i = 1, 2, 3, 4$.

Assim,

$$\begin{aligned}\chi^2_{obs} &= \frac{(95-111)^2}{111} + \frac{(87-90)^2}{90} + \frac{(51-39)^2}{39} + \frac{(67-60)^2}{60} \\ &= 2,3063 + 0,1000 + 3,6923 + 0,8167 = 6,9153\end{aligned}$$

Pela distribuição Qui-quadrado com $q = 4 - 1 = 3$ graus de liberdade, o nível descritivo é,

$$\text{valor-}p = P(X^2 \geq \chi^2_{obs}) = P(X^2 \geq 6,9153) = 0,0746,$$

sendo $X^2 \sim \chi^2_3$.

O valor-p pode ser obtido no *Rcmdr* através dos seguintes passos:

Clique no menu `Distribuições` → `Distribuições Contínuas` → `Distribuição Qui-Quadrado` → `Probabilidades da Qui-Quadrado`. Em `Valores da variável` coloque 6.9153, em `Graus de liberdade` coloque 3 e clique em `cauda superior`.

Decisão e conclusão: Como o nível descritivo do teste é menor do que o nível de significância considerado de 10%, então decidimos por rejeitar H_0. Ou seja, concluímos que temos evidências suficientes de que a audiência das emissoras de TV em um determinado horário não se distribuem conforme afirma o instituto.

Exercício 7

Uma escola de ensino médio deseja verificar se os desempenhos em Matemática e Física de estudantes desse ciclo estão associados. A tabela a seguir mostra o aproveitamento em Física e Matemática para uma amostra de alunos do ensino médio.

	Matemática			
Física	Notas Altas	Notas Regulares	Notas Baixas	Total
Notas Altas	48	75	22	145
Notas Regulares	52	145	58	255
Notas Baixas	32	74	40	146
Total	132	294	120	546

Exercício 7

(a) Especifique as hipóteses H_0 e H_1 de um teste de hipóteses estatístico adequado a esta situação.

H_0: O desempenho em Matemática é independente do desempenho em Física.

H_1: O desempenho em Matemática é dependente do desempenho em Física.

Exercício 7

(b) Se as variáveis não estão associadas, quantos estudantes seriam esperados com nota baixa em Física e nota alta em Matemática? E com nota baixa em Física e nota baixa em Matemática?

A tabela a seguir apresenta as frequências observadas e esperadas (entre parentêses) de estudantes de acordo com os desempenhos em Matemática e em Física:

	Matemática			
Física	Notas Altas	Notas Regulares	Notas Baixas	Total
Notas Altas	48 (35,05)	75 (78,08)	22 (31,87)	145
Notas Regulares	52 (61,65)	145 (137,31)	58 (56,04)	255
Notas Baixas	32 (35,30)	74 (78,62)	40 (32,09)	146
Total	132	294	120	546

Logo, o número esperado de alunos com nota baixa em Física e nota alta em Matemática é aproximadamente igual a

$$E_{31} = \frac{O_{.1} \times O_{3.}}{n} = \frac{132 \times 146}{546} = 35 \text{ alunos},$$

e o número esperado de alunos com nota baixa em Física e nota baixa em Matemática é aproximadamente igual a

$$E_{33} = \frac{O_{.3} \times O_{3.}}{n} = \frac{120 \times 146}{546} = 32 \text{ alunos}.$$

Exercício 7

(c) Com base nessa amostra, dê uma estimativa para a proporção de alunos dessa escola:

(i) que têm notas altas em matemática

(ii) com nota alta em matemática e física

(iii) com notas altas em matemática, entre aqueles com notas altas em física.

A proporção de alunos dessa escola que têm notas altas em matemática é $132/546 = 0,2418$.

A proporção de alunos dessa escola com nota alta em matemática e física é $48/546 = 0,0879$.

A proporção de alunos dessa escola com notas altas em matemática, entre aqueles com notas altas em física, é $48/145 = 0,3310$.

Exercício 7

(d) Por meio do nível descritivo, conclua sobre suas hipóteses utilizando um nível de significância de 5%.

Pelo *Rcmdr*, obtemos (Estatísticas → Tabelas de contingência → Digitar Tabela)

```
> .Table  # Counts
       matematica
fisica  1   2  3
     1 48  75 22
     2 52 145 58
     3 32  74 40
> .Test <- chisq.test(.Table, correct=FALSE)
> .Test
Pearson's Chi-squared test
data:  .Table
X-squared = 12.496, df = 4, p-value = 0.01402
```

Como o nível descritivo (valor-p=0,01402) é menor do que o nível de significância considerado de 5%, rejeitamos H_0, ou seja, temos evidências de que o desempenho em física é dependente do desempenho em matemática, nessa escola.