Testes de hipóteses

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística Laboratório de Estatística e Geoinformação





Um pesquisador deseja estudar o efeito de certa substância no tempo de reação de seres vivos a um certo tipo de estímulo. Um experimento foi desenvolvido com 49 cobaias que recebem a substância e logo em seguida são submetidas a um estímulo elétrico. Para cada cobaia foi avaliado o tempo de reação em segundos.

Os valores observados forneceram uma média de 9,1. Admite-se que o tempo de reação segue o modelo Normal com média 8 e desvio padrão 2 segundos. O pesquisador desconfia que o tempo médio sofre alteração por influência da substância. Efetue um teste de hipótese para verificar se o pesquisador tem razão. Considere um nível de significância de 5% e interprete os resultados.

- 1. Hipóteses nula e alternativa.
 - $H_0: \mu = 8 \times H_1: \mu \neq 8.$
- 2. Teste a ser efetuado.
 - ► Teste para a média com variância conhecida.
- 3. Estatística de teste.

$$ightharpoonup Z = rac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

4. Quantidades necessárias.

$$\bar{y} = 9.1.$$

$$\sigma = 2$$

$$\mu_0 = 8$$

$$n = 49$$

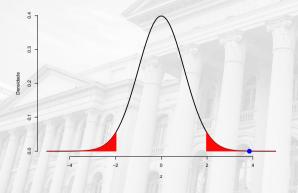
• $\mu_0 = 8$. 5. Nível de significância.

$$\alpha = 0.05$$

6. Valores críticos.

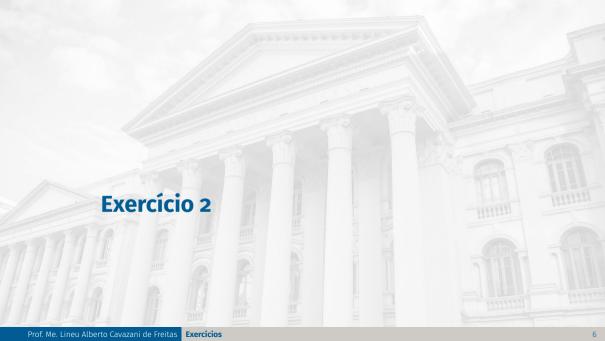
$$ightharpoonup z_{critico} = \pm 1,96$$

$$z_{crítico} = \pm 1,96$$



$$z_{calculado} = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{9.1 - 8}{2/\sqrt{49}} = 3.85$$

- ► A um nível de significância de 5% Rejeita-se a hipótese nula.
- ► Existe evidência suficiente nos dados que permita afirmar que o tempo de reação das cobaias em que foi aplicada a substância é diferente do tempo considerado habitual (8 segundos).



Deseja-se investigar se uma certa moléstia que ataca o rim altera o consumo de oxigênio desse orgão. Para indivíduos sadios, admite-se que esse consumo tem distribuição Normal com média $12cm^3/min$.

Os valores medidos em pacientes com a moléstia estão na tabela:

12.5	13.2	12.3	14.3	13.3
12.3	13.4	13.6	13.5	12.7

Qual seria a conclusão ao nível de significância de 1%?

- 1. Hipóteses nula e alternativa.
 - $H_0: \mu = 12 \times H_1: \mu \neq 12.$
- 2. Teste a ser efetuado.
 - ► Teste para a média com variância desconhecida.
- 3. Estatística de teste.

4. Quantidades necessárias. $\bar{y} = 13,11$. s = 0

$$\bar{y} = 13,11.$$

$$\mu_0 = 12$$

$$n = 10$$

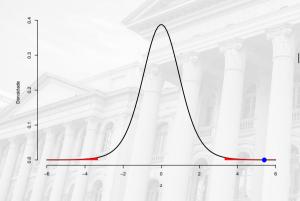
▶ $\mu_0 = 12$. 5. Nível de significância.

$$\alpha = 0.01$$

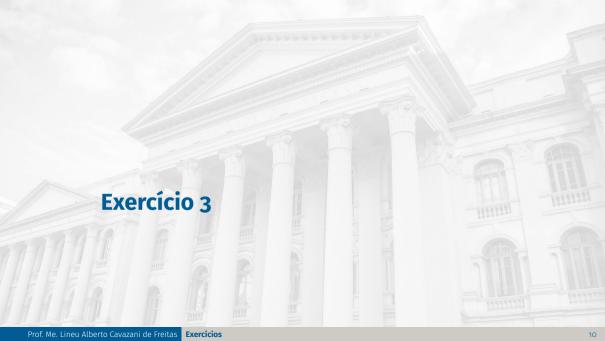
- 6. Valores críticos.
 - $t_{critico} = \pm 3,2498$

$$t_{critico} = \pm 3,2498$$

$$t_{calculado} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{13,11 - 12}{0,65/\sqrt{10}} = 5,4$$



- ► A um nível de significância de 1%, rejeita-se a hipótese nula.
- ► A moléstia tem influência no consumo renal médio de oxigênio.



Um antigo relatório de uma companhia afirma que 40% de toda a água obtida por meio de poços artesianos em determinada região é imprópria para consumo. Devido a uma série de mudanças na região existe uma forte suspeita de que este percentual seja maior do que 40%. Para dirimir as dúvidas, 400 poços foram sorteados e observou-se que em 184 deles a água era imprópria. Qual seria a conclusão, ao nível de 1% de significância?

- 1. Hipóteses nula e alternativa.
 - $H_0: p = 0.4 \times H_1: p > 0.4.$
- 2. Teste a ser efetuado.
 - ► Teste para a proporção.
- 3. Estatística de teste.

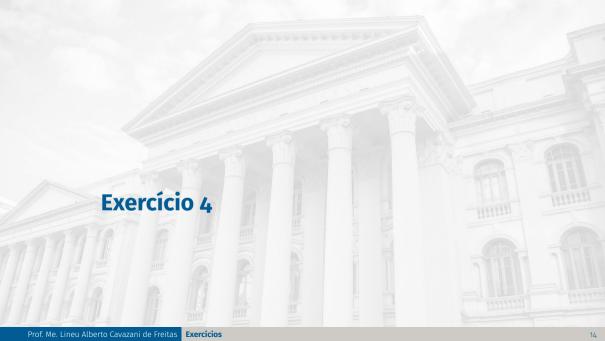
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

- 4. Quantidades necessárias.
 - $\hat{p} = 184/400 = 0.46.$
 - $p_0 = 0.4.$
 - n = 400
- 5. Nível de significância.
 - $\alpha = 0.01$
- 6. Valores críticos.
 - $ightharpoonup Z_{critico} = 2,33$

$$z_{cr(tico} = 2.33$$

$$z_{calculado} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.46 - 0.4}{\sqrt{\frac{0.4(1 - 0.4)}{400}}} = 2.45$$

- ► A um nível de significância de 1%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que mais de 40% de toda a água obtida por meio de poços artesianos da região é imprópria para consumo.



Um analista da qualidade está avaliando a variabilidade do tamanho de peças na produção de um componente. Caso a variância exceda 2 unidades de medida, existirá uma proporção inaceitável de peças que que causarão problemas. Para avaliar se a variabilidade está dentro do controle, uma amostra foi tomada. Os dados estão na tabela:

103.23	98.31	99.02	99.42	98.63	98.66	101.06	99.83
100.22	103.1	100.5	103.84	103.23	100.46	102.68	

Assumindo que o tamanho das peças segue distribuição normal e considerando um nível de significância de 5%, pode-se afirmar que a variabilidade está acima do aceitável? No cálculo da variância considere 3 casas decimais.

1. Hipóteses nula e alternativa.

$$H_0: \sigma^2 = 2 \times H_1: \sigma^2 > 2.$$

- 2. Teste a ser efetuado.
 - ► Teste para a variância.
- 3. Estatística de teste.

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

4. Quantidades necessárias.

$$n = 15.$$

$$s^2 = 3,713.$$

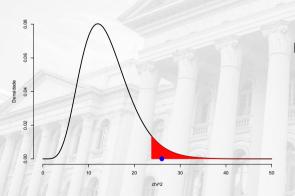
$$\sigma_0^2 = 2$$

5. Nível de significância.

$$\alpha = 0.05$$

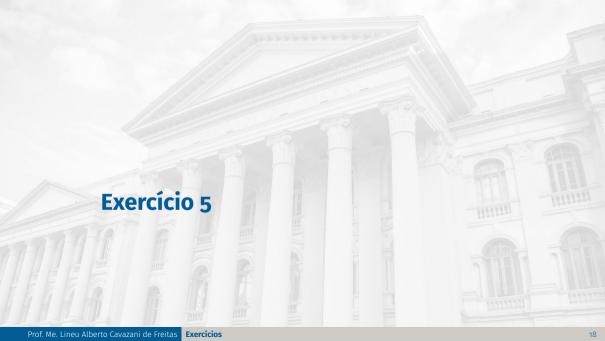
- 6. Valor crítico.
 - $\chi^2_{critico} = 23.685$

$$\chi^2_{critico} = 23.685$$



$$\chi^2 = \frac{(15-1)s^2}{\sigma_0^2} = \frac{(15-1)3,713}{2} = 25,991$$

- ► A um nível de significância de 5%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que a variância é maior que 2 unidades de medida.



Um grupo de pesquisadores conduziu um estudo para avaliar o efeito de atividade física nos batimentos cardíacos (em minutos). Foram definidos 2 grupos, ambos com 10 indivíduos. O primeiro grupo era composto por indivíduos ativos e o segundo grupo composto por indivíduos sedentários. Os dados estão na tabela:

ativos	92	76	76	90	97	90	86	93	100	115
sedentarios	102	125	123	135	114	152	115	133	126	138

Assumindo que a distribuição dos batimentos para ambos os grupos é normal com desvio padrão igual a 10 para ativos e igual a 12 para sedentários, existe evidência suficiente nos dados que permita afirmar que há diferença entre as médias dos dois grupos? Considere um nível de significância de 1%.

1. Hipóteses nula e alternativa.

$$H_0: \mu_1 - \mu_2 = 0 \times H_1: \mu_1 - \mu_2 \neq 0.$$

- 2. Teste a ser efetuado.
 - ► Teste para médias independentes com variância conhecida.
- 3. Estatística de teste.

$$Z = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{\bar{n}_1} + \frac{\sigma_2^2}{\bar{n}_2}}}$$

4. Quantidades necessárias.

$$\bar{y}_1 = 91,5$$

$$\sigma_2^2 = 12^2$$

•
$$\bar{y}_2 = 126,3$$

$$n_1 = 10$$

$$\sigma_1^2 = 10^2$$

$$n_2 = 10$$

5. Nível de significância.

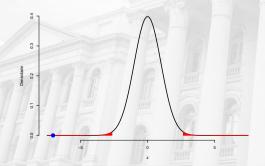
$$\alpha = 0.01$$

6. Valor crítico.

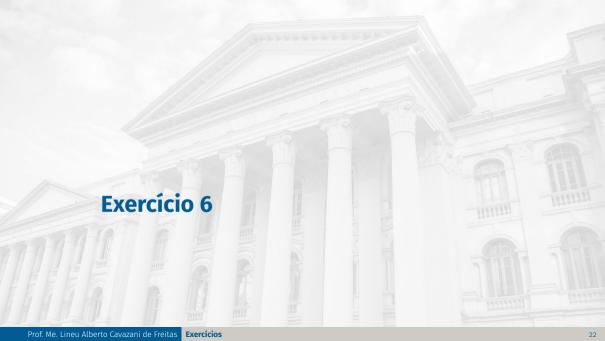
$$ightharpoonup z_{critico} = \pm 2,576$$

$$z_{critico} = \pm 2,576$$

$$z = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(91, 5 - 126, 3)}{\sqrt{\frac{10^2}{10} + \frac{12^2}{10}}} = -7,05$$



- ► A um nível de significância de 1%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que há diferença entre as médias dos grupos.



Um estudo tinha como objetivo avaliar o efeito de uma aula no conhecimento dos alunos. Para isso, um grupo de 10 alunos de uma escola foi sorteado ao acaso e submetidos a uma avaliação de um tema que não tinham conhecimento prévio. Após esta avaliação, os alunos tiveram uma aula do tema e foram submetidos a uma segunda avaliação. As notas dos alunos estão na tabela.

aluno	1	2	3	4	5	6	7	8	9	10
antes										
depois	69	76	69	70	67	56	70	52	94	52

Considerando um nível de significância de 10%, podemos afirmar que após a aula a nota dos alunos foi superior?

1. Hipóteses nula e alternativa.

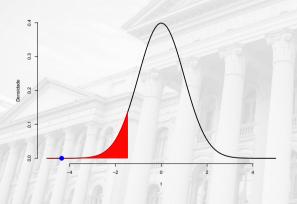
$$\blacktriangleright H_0: \mu_d = 0 \times H_1: \mu_d < 0.$$

- 2. Teste a ser efetuado.
 - ► Teste para difeerençs de médias para amostras pareadas.
- 3. Estatística de teste.

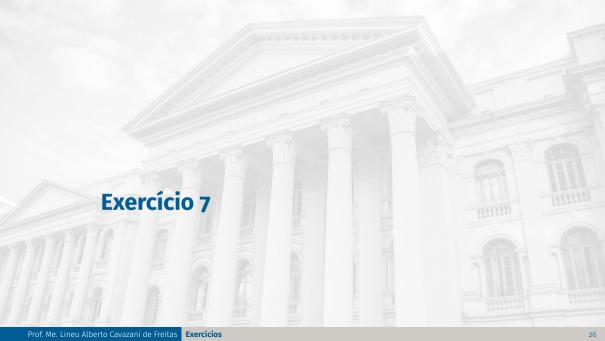
- 4. Quantidades necessárias.
 - \bullet $\bar{d} = -22.9$ (aumento depois).
 - $s_d = 16,64$
 - n = 10
- 5. Nível de significância.
 - $ightharpoonup \alpha = 0.1$
- 6. Valor crítico.
 - $t_{critico} = -1,38$

$$t_{critico} = -1,38$$

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-22,9 - 0}{16,64 / \sqrt{10}} = -4,352$$



- ► A um nível de significância de 10%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que a nota depois da aula foi superior à nota antes da aula.



Considere uma empresa que faz vendas de produtos usando uma página web. Algumas análises mostraram que uma determinada seção do site vem sendo menos acessada do que o esperado. A equipe de marketing digital da empresa quer avaliar se uma mudança de cores nesta parte específica da página faz com que o número de acessos aumente.

Para avaliar o efeito da mudança dos elementos na página os dois cenários foram colocados em produção sem que os usuários soubessem. Foi verificado que, no esquema de cores antigo, dos 1200 usuários, 564 acessaram a seção de interesse. Já no esquema de cores modificado, de 1200 usuários, 654 acessaram a seção.

Existe evidência de que a alteração no esquema de cores causou aumento no número de acessos? Proceda o teste de hipóteses adequado a um nível de significância de 5%.

1. Hipóteses nula e alternativa.

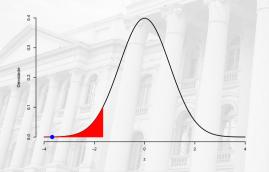
$$H_0: p_1 - p_2 = 0 \times H_1: p_1 - p_2 < 0.$$

- 2. Teste a ser efetuado.
 - ► Teste para diferença de proporções para duas populações.
- 3. Estatística de teste.

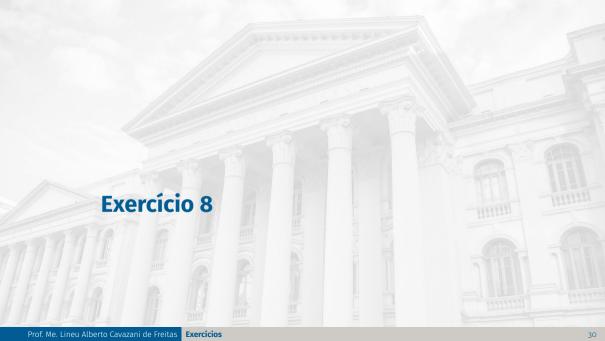
- 4. Quantidades necessárias.
 - $\hat{p}_1 = 0.47.$
 - $\hat{p}_2 = 0.545.$
 - $\bar{p} = 0.5075.$
 - $n_1 = 1200$
 - $n_2 = 1200$
- 5. Nível de significância.
 - $\alpha = 0.05$
- 6. Valor crítico.
 - $ightharpoonup Z_{critico} = -1,645$

$$z_{critico} = -1,645$$

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\bar{p}(1-\bar{p})(1/n_1 + 1/n_2)}} = \frac{0.47 - 0.545}{\sqrt{0.5075(1-0.5075)(1/1200 + 1/1200)}} = -3.675$$



- ► A um nível de significância de 5%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar as alterações feitas no site causaram aumento no número de acessos.



Queremos verificar se duas máquinas produzem peças com a mesma homogeneidade quanto à resistência à tensão. Para isso, sorteamos duas amostras de 6 peças de cada máquina e obtivemos as seguintes resistências. Os dados estão na tabela:

Máquina 1	145	127	136	142	141	137
Máquina 2	143	128	132	138	142	132

Considerando que as resistências seguem distribuição normal, existe evidência para afirmar que a variabilidade das duas máquinas difere a um nível de significância de 10%?

- 1. Hipóteses nula e alternativa.
 - $\blacktriangleright H_0: \sigma_A^2/\sigma_B^2 = 1 \times H_1: \sigma_A^2/\sigma_B^2 \neq 1.$
- 2. Teste a ser efetuado.
 - ► Teste para razão de variâncias.
- 3. Estatística de teste.

$$F = \frac{s_1^2}{s_s^2} \sim F_{\nu_1, \nu_2}$$

4. Quantidades necessárias.

$$ightharpoonup s_1^2 = 40.$$
 $ightharpoonup n_1 = 6$

$$n_1 = 6$$

$$ightharpoonup s_2^2 = 36,97.$$
 $ightharpoonup n_2 = 6$

$$n_2 = 6$$

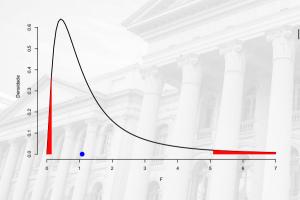
5. Nível de significância.

$$\alpha = 0.1$$

6. Valor crítico.

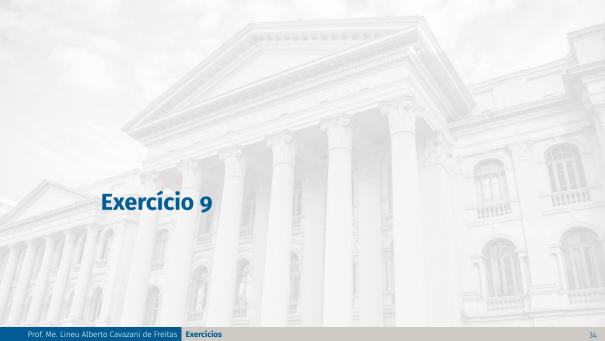
$$F_{critico} = 1/5,05;5,05$$

$$F_{critico} = 1/5,05;5,05$$



$$F = \frac{s_1^2}{s_s^2} = \frac{40}{36,97} = 1,08$$

- ► A um nível de significância de 10%, não rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que as máquinas produzem com a mesma homogeneidade.



Em um estudo para verificar a relação entre crises de asma e incidência de gripe, 101 crianças foram escolhidas ao acaso dentre aquelas acompanhadas por um posto de saúde. Os dados estão na tabela:

Asma/Gripe	Sim	Não		
Sim	15	22		
Não	30	34		

A um nível de significância de 5%, existe evidência de que as ocorrências de asma e gripe são independentes?

1. Hipóteses nula e alternativa.

$$H_0: \chi^2 = 0 \times H_1: \chi^2 > 0$$

- 2. Teste a ser efetuado.
 - ► Teste Qui-quadrado para independência.
- 3. Estatística de teste.

•
$$Q = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{\nu}^2$$

$$v = (r-1)(c-1)$$

4. Matriz de valores observados e esperados:

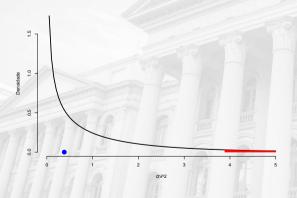
Observados

Esperados

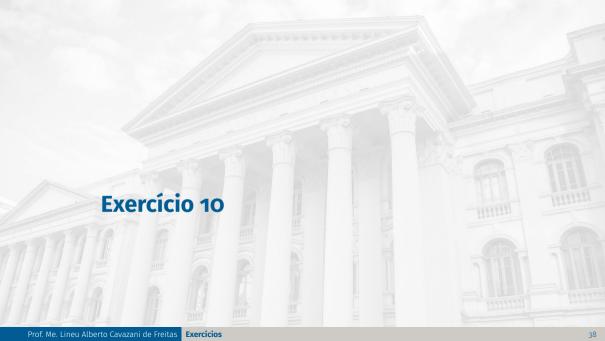
- 5. Nível de significância.
 - $\alpha = 0.05$
- 6. Valor crítico.
 - $x_{critico}^2 = 3.841$

$$\chi^2_{critico} = 3.841$$

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 0,3808$$



- A um nível de significância de 5%, não rejeita-se a hipótese nula.
- Não existe evidência suficiente nos dados que permite afirmar que as ocorrências de asma e gripe são eventos dependentes.



Um dado foi fabricado com o centro em madeira leve e cada face com uma chapa metálica porém de diferentes características (espessura/densidade) em cada face. Este dado foi lançado diversas vezes e número de vezes que cada face apareceu está na tabela:

Toronto.						
face	1	2	3	4	5	6
n	56	63	48	49	61	55

A um nível de significância de 5%, existe evidência para crer que o dado é viciado?

1. Hipóteses nula e alternativa.

$$H_0: \chi^2 = 0 \times H_1: \chi^2 > 0$$

- 2. Teste a ser efetuado.
 - ► Teste Qui-quadrado para aderência.
- 3. Estatística de teste.

$$\chi^2 = \sum_{j=1} \frac{(O_j - E_j)^2}{E_j} \sim \chi_v^2$$

$$v = k - 1$$

4. Vetor de valores observados e esperados:

$$O = \begin{bmatrix} 56 & 63 & 48 & 49 & 61 & 55 \end{bmatrix}$$

 $E = \begin{bmatrix} 55,33 & 55,33 & 55,33 & 55,33 & 55,33 \end{bmatrix}$

5. Nível de significância.

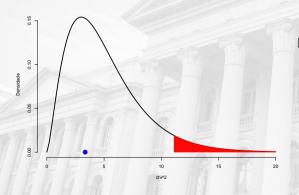
$$\alpha = 0.05$$

6. Valor crítico.

$$\chi^2_{critico} = 11,070$$

$$\chi^2_{critico} = 11,070$$

$$\chi^2 = \sum_{j=1} \frac{(O_j - E_j)^2}{E_j} = 3,35$$



- ► A um nível de significância de 5%, não rejeita-se a hipótese nula.
- Não existe evidência suficiente nos dados que permite afirmar o dado é viciado.



Uma empresa que fabrica automóveis está fazendo testes em um modelo para avaliar modificações a serem feitas na próxima geração. Havia uma suspeita de que ocorria uma alteração no consumo médio de acordo com a quilometragem dos veículos, de tal modo que quanto maior a quilometragem, maior o consumo médio. Para avaliar se isso ocorria, uma amostra de 10 veículos foi selecionada ao acaso e anotou-se a quilometragem e o consumo médio de combustível. Os dados coletados estão na tabela:

km	1239	2198	2870	3231	3670	5003	5223	5874	6427	6832
consumo	8	9	10	11	11	13	13	14	15	15

Com base nos dados, a suspeita do fabricante é válida ou não? Considere um nível de significância de 1%.

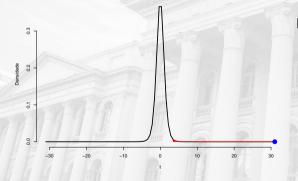
- 1. Hipóteses nula e alternativa.
 - $H_0: \rho = 0$
 - ▶ $H_1: \rho > 0$
- 2. Teste a ser efetuado.
 - ► Teste para correlação linear de Pearson.
- 3. Estatística de teste.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

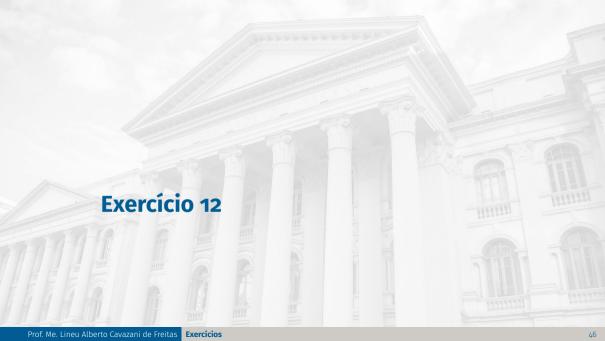
- 4. Quantidades necessárias.
 - ightharpoonup cov(km, consumo) = 4646,3.
 - ightharpoonup var(km) = 3568078.
 - \triangleright var(consumo) = 6,1
 - r = 0.9959
- 5. Nível de significância.
 - $\alpha = 0.01$
- 6. Valor crítico.
 - $t_{critico} = 2,8965$

$$t_{critico} = 2,8965$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,9959\sqrt{10-2}}{\sqrt{1-0,9959^2}} = 31,01$$



- A um nível de significância de 1%, rejeita-se a hipótese nula.
- Existe evidência suficiente nos dados que permite afirmar que quanto maior a quilometragem, maior é o consumo médio dos veículos.



Em uma indústria é realizado um curso de treinamento após 1 ano de admissão dos operários. Historicamente, o coeficiente de correlação entre a nota final do curso e a produtividade dos operários após 6 meses do curso é 0,4.

Para uma nova versão, foram introduzidas modificações no curso, com o objetivo de aumentar a produtividade. Em uma amostra de 28 operários submetidos ao novo modelo, o coeficiente de correlação entre nota e produtividade foi igual a 0,54. Você diria que os objetivos das alterações no curso foram atingidos? Proceda o teste de hipóteses adequado considerando um nível de significância de 10%.

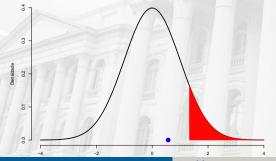
- 1. Hipóteses nula e alternativa.
 - $\vdash H_0: \rho = 0.4$
 - $\vdash H_1: \rho > 0.4$
- 2. Teste a ser efetuado.
 - ► Teste para correlação linear de Pearson.
- 3. Estatística de teste.
 - $ightharpoonup z = (arctan(r) arctan(\rho_0))\sqrt{n-3}$

- 4. Quantidades necessárias.
 - r = 0.54.
 - $\rho_0 = 0.4.$
 - n = 28.
- 5. Nível de significância.
 - $ightharpoonup \alpha = 0.1$
- 6. Valor crítico.
 - $ightharpoonup Z_{critico} = 1,28$

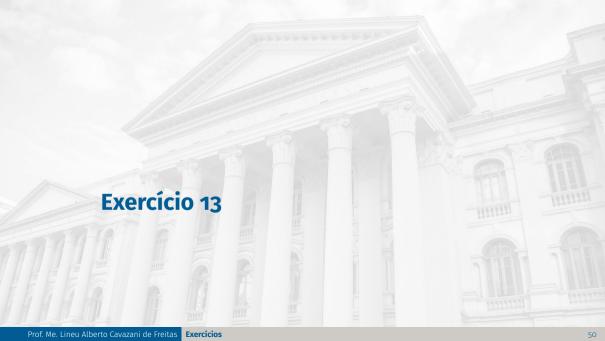
$$z_{critico} = 1,28$$

$$z = (arctan(r) - arctan(\rho_0))\sqrt{(n-3)} =$$
$$(arctan(0,54) - arctan(0,4))\sqrt{(28-3)} =$$

0,573



- A um nível de significância de 10%, não rejeita-se a hipótese nula.
- Não existe evidência suficiente nos dados que permite afirmar que o coeficiente de correlação aumentou.



Um laboratório está interessado em medir o efeito da temperatura sobre a potência de um antibiótico. Dez amostras de 50 gramas cada foram guardadas a diferentes temperaturas, e após 15 dias mediu-se a potência. Os resultados estão na tabela:

Temperatura	30	30	50	50	50	70	70	70	90	90
Potência	38	43	32	26	33	19	27	23	14	21

Existe efeito da temperatura sobre a potência? Qual o valor esperado de potência para uma temperatura de 30°?

Quantidades necessárias

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$ightharpoonup e_i \sim N(0, \sigma^2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

▶
$$n = 10$$

•
$$\bar{x} = 60$$

$$y = 27,6$$

$$\sum_{i} x_i y_i = 14960$$

$$\sum_{i} x^2 = 40200$$

$$\sum_{i} x^2 = 40200$$

Estimativas dos parâmetros

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} = \frac{14960 - (10 \times 60 \times 27,6)}{40200 - (10 \times 60^2)} = -0,381$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 27.6 - (-0.381 \times 60) = 50.46$$

Equação do modelo: $\hat{y}_i = 50,46 - 0,381x_i$

$$SQreg = \hat{\beta}_1^2 \sum_i (x_i - \bar{x})^2 = -0.381^2 \times 4200 = 609.6762$$

$$SQtot = \sum_{i} (y_i - \bar{y})^2 = 720,4$$

$$S_e^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2} = 13,859$$

$$R^2 = \frac{SQreg}{SQtot} = 0.8463$$

Variabilidade das estimativas dos parâmetros

$$s_{\beta_1} = \sqrt{\frac{S_e^2}{\sum_i (x_i - \bar{x})^2}} = \sqrt{\frac{13,859}{4200}} = 0,05744$$

$$s_{\beta_0} = \sqrt{S_e^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2} \right]} = \sqrt{13,859 \left[\frac{1}{10} + \frac{60^2}{4200} \right]} = 3,6421$$

Intervalos de confiança para os parâmetros

$$IC(\beta_i) = \hat{\beta}_i \pm t_{n-2} s_{\beta_i}$$

$$IC(\beta_1) = -0.381 \pm 2.3060 \times 0.05744 = [-0.5134; -0.248]$$

$$IC(\beta_0) = 50,46 \pm 2.3060 \times 3,6421 = [42,06;58,85]$$

Valor esperado de potência para 30°

$$\hat{y}_i = 50,46 - 0,381 \times 30 = 39,03$$