

ESTATÍSTICA II

Michelle Hanne Soares de Andrade

michellehanne@cefetmg.br

1º. SEMESTRE 2018

Teste de Hipótese para a Variância e Desvio

Padrão

- Ao testar uma hipótese sobre o desvio-padrão σ ou a variância σ^2 de uma população, admitimos que os valores da população sejam distribuídos normalmente.

Notação

n = tamanho da amostra

s^2 = variância amostral

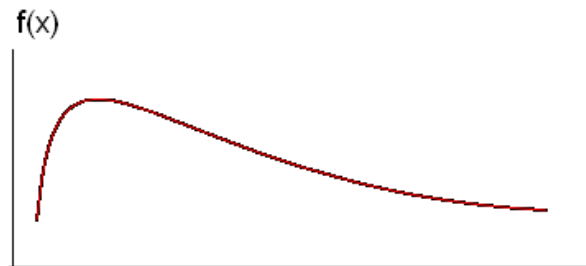
σ^2 = variância populacional

Teste de Hipótese para a Variância e Desvio

Padrão

- Estatística do Teste

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



- Tabela Qui-Quadrado (**Tabela IV: Distribuição Qui-Quadrado**) x
- Número de Graus de Liberdade = $n - 1$

Exemplo

- O tempo para transmitir 10 MB em determinada rede de computadores varia segundo um modelo normal, com média 7,4 segundos e variância 1,3 segundos. Depois de algumas mudanças na rede, acredita-se numa redução no tempo de transmissão de dados. Além de uma possível mudança na variabilidade. Foram realizados 10 ensaios independentes com um arquivo de 10 MB e foram coletados os tempos de transmissão, em segundos:

6,8	7,1	5,9	7,5	6,3	6,9	7,2	7,3	6,6	6,3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Exemplo 1

- Existe evidência suficiente de que as mudanças na rede de computadores alteraram a variabilidade no tempo de transmissão de dados? Ao nível de 0,05.
- Solução:

As hipóteses são:
$$\begin{cases} H_0 : \sigma^2 = 1,3 \\ H_1 : \sigma^2 \neq 1,3 \end{cases}$$

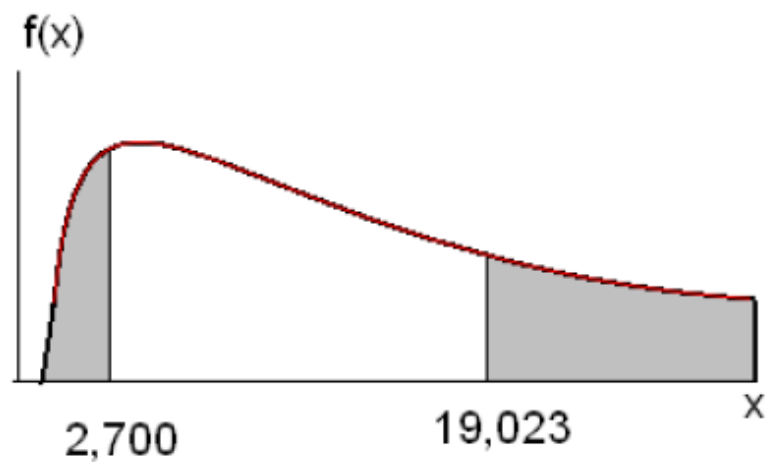
O teste é bilateral direito, pois o sinal de H_1 é \neq ;

O nível de significância é $\alpha = 0,05$;

Exemplo 1

O valor do grau de liberdade é de $10-1 = 9$. Logo os valores críticos são

$\chi^2 = 2,700$ e $\chi^2 = 19,023$; Logo temos:



Exemplo 1

Os dados amostrais indicam: $s^2 = 0,261$;

a estatística de teste é dada por:

$$\chi^2 = \frac{(10-1) \cdot 0,261}{1,3} = 1,807$$

Conclusão: Como a estatística de teste está na dentro da região crítica, então rejeitamos H_0 .

Exemplo 2

- Numa determinada empresa, empregados que desempenham a mesma função têm salários diferentes em função do tempo de casa e bonificações por desempenho. Segundo a empresa, o desvio-padrão para o salário de uma certa função é de R\$ 150,00. Entrevistando 5 funcionários que desempenham essa mesma função, verificou-se que seus salários eram, respectivamente, R\$ 1.000; R\$ 1.200; R\$ 1.500; R\$ 1300; R\$ 900. Testar a afirmação da empresa com significância de 5%, supondo que os salários sejam normalmente distribuídos.

Exemplo 2

Resolução

$$H_o : \sigma^2 = \sigma_o^2$$

$$H_A : \sigma^2 \neq \sigma_o^2$$

Sendo:

$$\sigma^2 = (150)^2 = 22.500$$

$$H_o : \sigma^2 = 22.500$$

$$H_A : \sigma^2 \neq 22.500$$

Exemplo 2

Calculando a estatística do teste:

$$\chi^2_{(n-1)} = \frac{S^2(n-1)}{\sigma^2}$$

dados

$$\alpha = 0,05$$

$$n = 5$$

Precisamos encontrar S^2

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

sendo

$$\bar{X} = 1.180$$

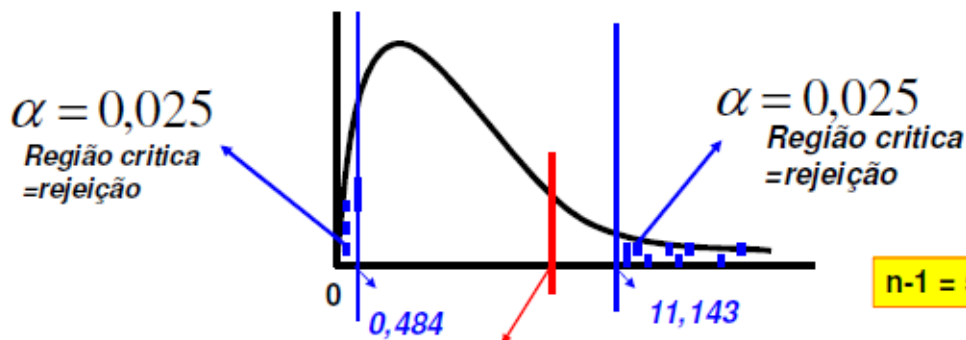
$$S^2 = 57.000$$

Exemplo 2

$$\chi^2_{(n-1)} = \frac{S^2(n-1)}{\sigma^2}$$

Substituindo....

$$\chi^2_{(n-1)} = \frac{57.000(5-1)}{(150)^2} = 10,13 \text{ approx.}$$



$$\chi^2_{(n-1)} = 10,13$$

Não Rejeito H0 – afirmação da empresa não é contestada

Exemplo 3

- Uma caixa de fósforos de certa marca vem com a inscrição: “ contém, em média, 40 palitos”. Segundo o fabricante, o desvio-padrão é de, no máximo, dois palitos. Em uma amostra com 31 caixas, entretanto, foi encontrado um desvio padrão amostral de 3 palitos. Supondo que o número de palitos por caixa seja uma variável normal, teste a afirmativa do fabricante utilizando um nível de significância de 1%.

Exemplo 3

$$H_o : \sigma^2 \leq 4$$

$$H_A : \sigma^2 > 4$$

Calculando a estatística do teste:

$$\chi^2_{(n-1)} = \frac{S^2(n-1)}{\sigma^2}$$

$$\chi^2_{(n-1)} = \frac{3^2(31-1)}{2^2} \cong 67,5$$

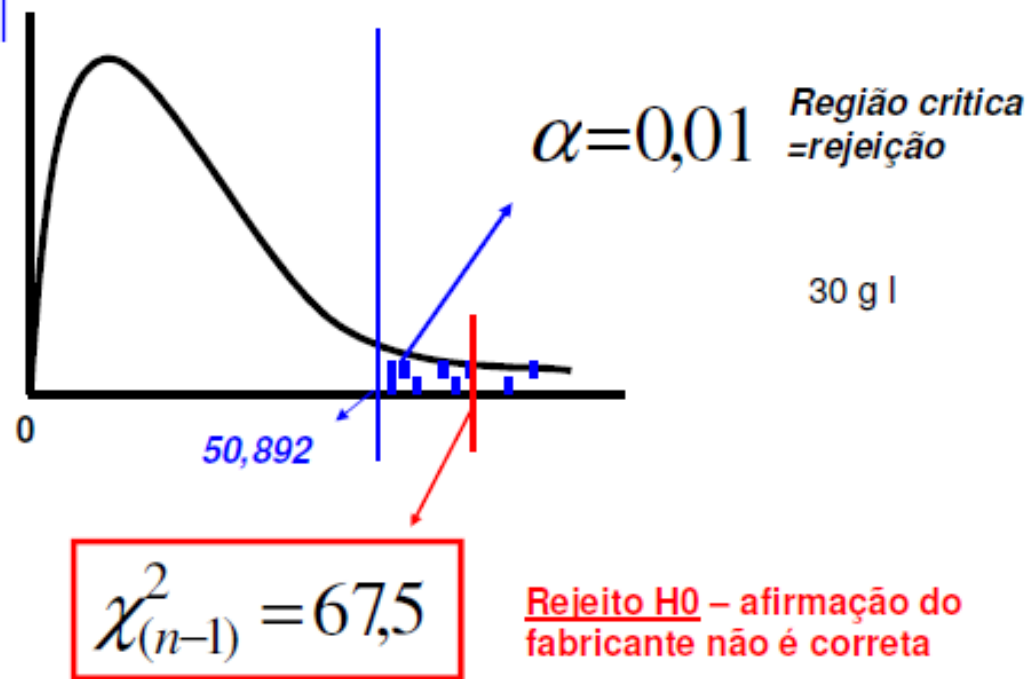
$$\alpha = \overset{\text{dados}}{0,01}$$

$$n = 31$$

$$\sigma_0^2 = 2^2 = 4$$

$$S = \sigma_{amostral} = 3$$

Exemplo 3



Testes Não-Paramétricos

- Os testes não-paramétricos são utilizados quando não temos informação sobre a distribuição da população.
- **Vantagens-** Menos suposições são necessárias. Em muitos casos, apenas dados nominais (categóricos) ou ordinais (ranks) são necessários, ao invés de numéricos (intervalares).
- **Desvantagens-** Frequentemente preferimos ter um modelo bem definido com parâmetros importantes tais como média e variância incluídas para melhor interpretação.

Testes Não-Paramétricos - Tipos

- Teste dos sinais;
- Teste de postos com sinais de Wilcoxon para duas amostras dependentes e independentes;
- Teste de kruskal-Wallis;
- Correlação por postos;
- Teste de repetições para aleatoriedade;
- Teste do qui-quadrado;
- Teste do qui-quadrado para independência ou associação;
- Teste de Mann-Whitney;
- Teste da mediana;

Testes Não-Paramétricos - Tipos

Não se refere à distribuição da estatística de teste, mas ao fato de que os métodos podem ser aplicados a amostras de populações de qualquer distribuição.

Esta deve ser especificada apenas em termos gerais (ser contínua, simétrica, idêntica) sem precisar pertencer a alguma família (como normal, uniforme, exponencial, etc).

Teste de Correlação por Postos

- O teste de correlação por postos pode ser utilizado para verificar se existe alguma associação entre duas variáveis.
- A taxa de eficiência do teste é de 91%.

Notação:

r_s : coeficiente de correlação por postos para dados amostrais emparelhados;

ρ_s : coeficiente de correlação por postos para todos os dados populacionais emparelhados;

n : número de pares de dados;

d : diferença entre postos para as duas observações dentro de um par.

Teste de Correlação por Postos

O índice s é utilizado em homenagem a Charles Spearman (1863 - 1945).

Ao testar se há ou não correlação, testamos as seguintes hipóteses:

$$\begin{cases} H_0 : \rho_s = 0 \\ H_1 : \rho_s \neq 0 \end{cases}$$

Teste de Correlação por Postos

- A estatística de teste, para o caso em que não há empate entre os postos, é dada por:

– onde cada valor de d é um $r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$ postos para um par de dados amostrais.

- Caso haja empate entre os postos então a estatística de teste é dada por:

– onde x = posto d $r_s = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}}$

Posto:
ordem dos
elementos

Teste de Correlação por Postos

- **Valores críticos:**
- Se $n \leq 30$, consulte tabela T.
- Se $n > 30$, use a fórmula:

$$r_s = \frac{\pm z}{\sqrt{n-1}}$$

- onde o valor de z corresponde ao nível de significância

Exemplo 1

- A tabela a seguir apresenta 9 dados do volume desgastado do aço e da viscosidade do óleo.
- Há correlação entre as duas variáveis? Use $\alpha = 0,05$.

Volume desgastado Y (10^{-4} mm^3)	Viscosidade X
240	1,6
181	9,4
193	15,5
155	20
172	22
110	35,5
113	43
75	40,5
94	33

Exemplo 1

- Solução:

Passo 1: As hipóteses são $\begin{cases} H_0 : \rho_s = 0 \\ H_1 : \rho_s \neq 0 \end{cases}$

Passo 2: Nível de significância é $\alpha = 0,05$

Passo 3: Utilizaremos estatística não-paramétrica pois não temos informação sobre a população original.

Exemplo 1

- Passo 4: Estatística de teste:

Volume desgastado Y (10^{-4} mm^3)	Viscosidade X	Posto Y	Posto X	d	d^2
240	1,6	9	1	8	64
181	9,4	7	2	5	25
193	15,5	8	3	5	25
155	20	5	4	1	1
172	22	6	5	1	1
110	35,5	3	7	4	16
113	43	4	9	5	25
75	40,5	1	8	7	49
94	33	2	6	4	16

$$\begin{aligned} r_s &= 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \\ &= 1 - \frac{6(222)}{9(81 - 1)} \\ &= 1 - \frac{1332}{720} \\ &= 1 - 1,85 \\ &= -0,85 \end{aligned}$$

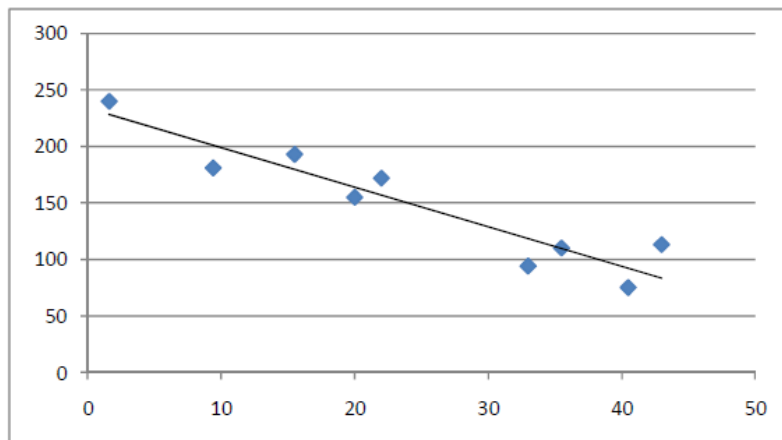
Assim temos $\sum d^2 = 222$

Exemplo 1

Passo 5: valores críticos:

Pela tabela A-9, os valores críticos são $\pm 0,700$

Como $r_s = -0,85$ está dentro da região crítica então rejeitamos H_0 . Logo existe correlação.



Exemplo 1

Tabela de
Coeficiente
g.l=9 e $\alpha =$
0,05

Tabela de valores críticos para o teste bilateral
da correlação de postos de Spearman

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,02$	$\alpha = 0,01$
5	0,900	—	—	—
6	0,829	0,886	0,943	—
7	0,714	0,786	0,893	0,929
8	0,643	0,738	0,833	0,881
9	0,600	0,700	0,783	0,833
10	0,564	0,648	0,745	0,794
11	0,536	0,618	0,709	0,755
12	0,503	0,587	0,678	0,727
13	0,484	0,560	0,648	0,703
14	0,464	0,538	0,626	0,679
15	0,446	0,521	0,604	0,654
16	0,429	0,503	0,582	0,635
17	0,414	0,485	0,566	0,615
18	0,401	0,472	0,550	0,600
19	0,391	0,460	0,535	0,584
20	0,380	0,447	0,520	0,570
21	0,370	0,435	0,508	0,556
22	0,361	0,425	0,496	0,544
23	0,353	0,415	0,486	0,532
24	0,344	0,406	0,476	0,521
25	0,337	0,398	0,466	0,511
26	0,331	0,390	0,457	0,501
27	0,324	0,382	0,448	0,491
28	0,317	0,375	0,440	0,483
29	0,312	0,368	0,433	0,475
30	0,306	0,362	0,425	0,467

Teste do Qui-Quadrado

A prova χ^2 de uma amostra é aplicada quando o pesquisador está interessado no número de indivíduos, objetos ou respostas que se enquadram em várias categorias que podem ser duas ou mais.

Usa-se a técnica do tipo de prova de aderência, ou seja, deve comprovar se existe diferença significativa entre o número observado de indivíduos, ou de respostas, em determinada categoria, e o respectivo número esperado, baseado na hipótese de nulidade.

Teste do Qui-Quadrado

- Como calcular Karl Pearson propôs a seguinte fórmula para medir as possíveis discrepâncias entre proporções observadas e esperadas:

$$\chi^2 = \Sigma [(o - e)^2 / e]$$

em que

- o = frequência observada para cada classe,
- e = frequência esperada para aquela classe.

Note-se que $(o - e)$ = desvio (d), portanto a fórmula também pode ser escrita como

$$\chi^2 = \Sigma (d^2 / e)$$

Teste do Qui-Quadrado

- Percebe-se que as frequências observadas são obtidas diretamente dos dados das amostras, enquanto que as frequências esperadas são calculadas a partir destas.
- É importante notar que $(o - e)$ é a diferença entre a frequência observada e a esperada em uma classe. Quando as frequências observadas são muito próximas às esperadas, o valor de χ^2 é pequeno. Mas, quando as divergências são grandes $(o - e)$ passa a ser também grande e, conseqüentemente, χ^2 assume valores altos.

Teste do Qui-Quadrado - Hipóteses a serem testadas

- O pesquisador trabalha com duas hipóteses:
 - **Hipótese nula:** As frequências observadas não são diferentes das frequências esperadas. Não existe diferença entre as frequências (contagens) dos grupos. Portanto, não há associação entre os grupos.
 - **Hipótese alternativa:** As frequências observadas são diferentes das frequências esperadas, portanto existe diferença entre as frequências. Portanto, há associação entre os grupos.

Teste do Qui-Quadrado – Exemplo 1

- Se uma moeda não viciada for jogada 100 vezes, espera-se obter 50 caras e 50 coroas, já que a probabilidade de cair cara (p) é $= \frac{1}{2}$ e a de cair coroa (q) também é $= \frac{1}{2}$.
- Entretanto, na prática, é muito difícil obter valores observados, idênticos aos esperados, sendo comum encontrar valores que se desviam dos teóricos.
- **Supondo que uma moeda foi jogada 100 vezes e se obteve 60 caras e 40 coroas.**
 - a. Qual será o valor de χ^2 ?
 - b. Como se pode interpretar esse valor?

Teste do Qui-Quadrado – Exemplo 1

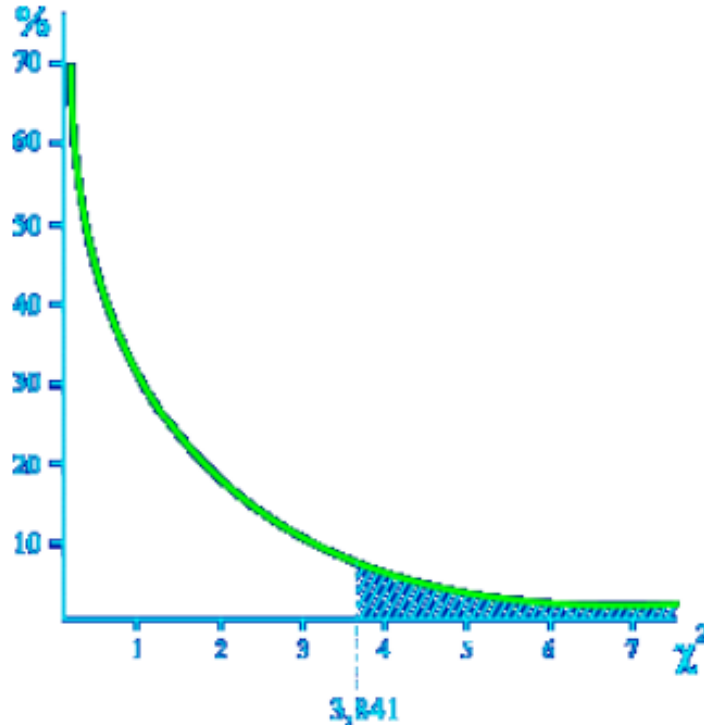
- $E(\text{cara}) = \frac{1}{2} \cdot 100$ e $E(\text{coroa}) = \frac{1}{2} \cdot 100$.
- Assim, os valores esperados são: cara: 50 e coroa: 50 e os observados são: cara: 60 e coroa: 40.

$$\chi^2 = [(60 - 50)^2 / 50] + [(40 - 50)^2 / 50]$$

a. Valor de $\chi^2 = 2 + 2 = 4$

- O que significa esse número? Ou seja, como se analisa um teste de χ^2 ? Supondo que em vez de lançarmos 100 moedas uma única vez, tivéssemos feito inúmeros lançamentos de 100 moedas. Se calcularmos o χ^2 a cada 100 lançamentos, e, depois, colocarmos todos os resultados em um gráfico, teria sido obtida a figura abaixo:

Teste do Qui-Quadrado – Exemplo 1



Tomando a área total sob a curva como 100%, sabe-se que o valor 3,841 delimita 5% dela. Este é o valor crítico de qui quadrado conhecido como χ_c^2 . Portanto, espera-se em experimentos semelhantes, que valores de χ^2 , e menores que 3,841 tenham 95% de probabilidade de ocorrência

Teste do Qui-Quadrado – Exemplo 1

- Sempre que o valor de χ^2 for menor que 3,841 aceita-se a hipótese de igualdade estatística entre os números de observados e de esperados (H_0). Ou seja, admite-se que os desvios não são significativos.
- No exemplo dado, como o valor de Qui-Quadrado obtido (4) para 2 classes foi maior que o esperado ao acaso (3,841), aceita-se a hipótese alternativa e admite-se que a moeda seja viciada.

Teste do Qui-Quadrado – Exemplo 2

- Se um dado não viciado for jogado 6 vezes, espera-se obter 1 vez cada face (1, 2, 3, 4, 5 e 6) já que a probabilidade de cair qualquer face é $1/6$. Supondo que um dado foi jogado 186 vezes e se obteve:

Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
34	29	30	32	28	33

- a. Qual será o valor de χ^2 ?
- b. Como se pode interpretar esse valor?

Teste do Qui-Quadrado – Exemplo 2

As frequências esperadas em cada classe são calculadas por: $p.N$. Portanto:

$$E_{(\text{face } 1)} = E_{(\text{face } 2)} = E_{(\text{face } 3)} = E_{(\text{face } 4)} = E_{(\text{face } 5)} = E_{(\text{face } 6)} = p.N = 1/6 . 186 = 31$$

a. Qual será o valor de χ^2 ?

Assim, os valores parciais são somados: e chega-se ao valor de χ^2 :

observado	34	29	30	32	28	33
esperado	31	31	31	31	31	31
χ^2 parcial	0,2903	0,1290	0,0322	0,0322	0,2903	0,1290

$$\chi^2 = (0,2903 + 0,1290 + 0,0322 + 0,0322 + 0,2903 + 0,1290) = 0,903$$

Teste do Qui-Quadrado – Exemplo 2

- b. Como se pode interpretar esse valor?

Lembrando que G.L. = número de classes -1, como há há 6 classes, G.L. = 5. Verificando-se a tabela de χ^2 na linha em G.L. = 5 encontra-se χ_c^2 igual a 11,070.

Como o valor de Qui Quadrado obtido (0,903) foi menor que o esperado ao acaso (11,070) admite-se que o dado seja honesto.