entendendo estatística divertidamente

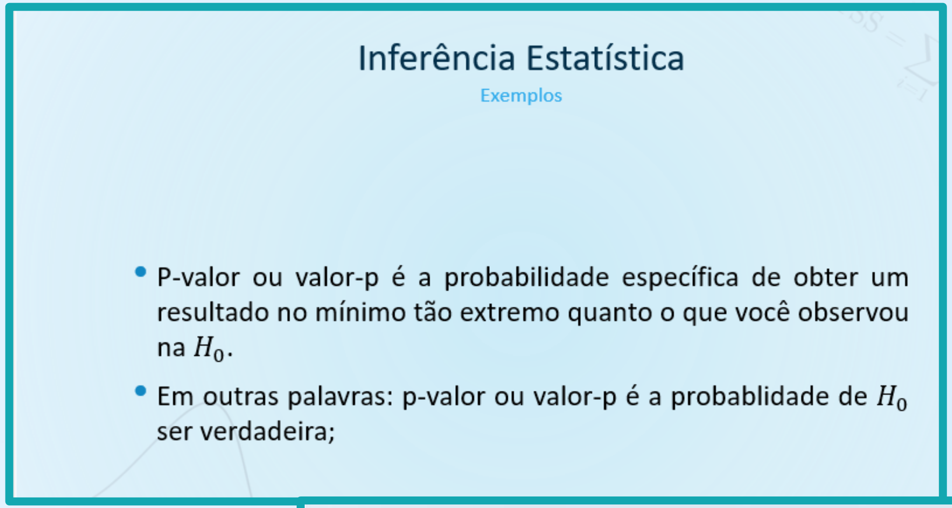
Testes Estatísticos

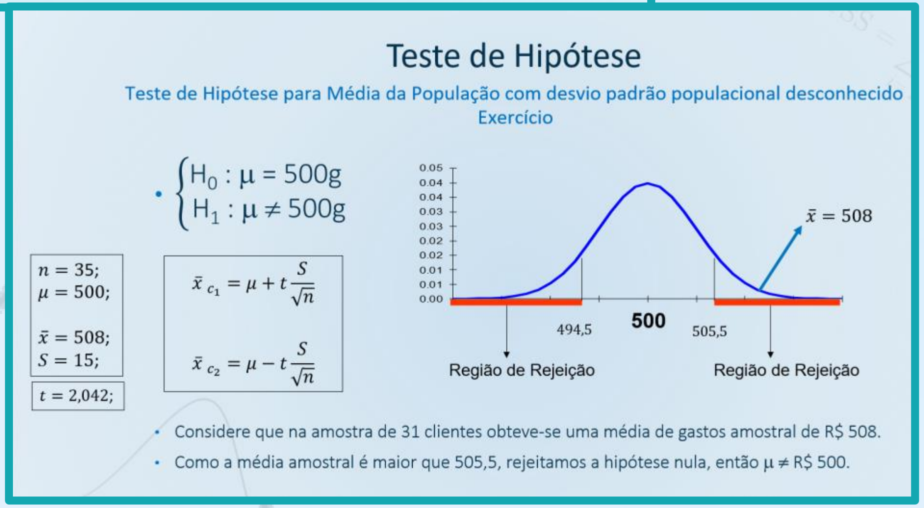
Profa. Adriana Silva

Seja bem vindo!!!
Câmera ligada e
Microfone mutado sempre que não estiver falando

Testes Estatísticos

p-valor ou valor p

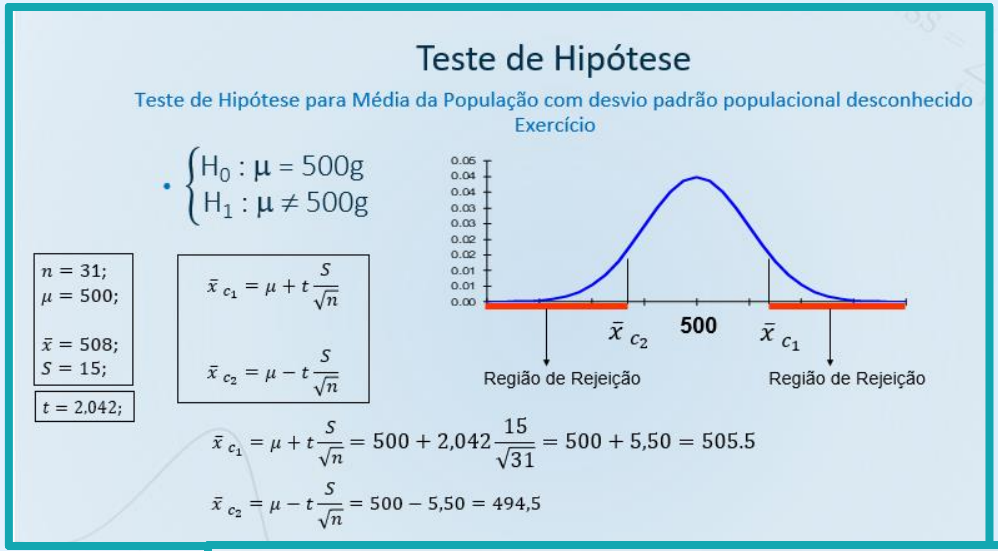




- Falácia da Probabilidade inversa
  - O p-valor é a probabilidade de H_0 ser verdadeira, dado que obtivemos esses dados
  - O erro está na inversão da probabilidade condicional
  - Exemplo:
    - P-valor é $P(\text{dados} \mid H_0)$
    - P-valor NÃO é $P(H_0 \mid \text{dados})$

Testes Estatísticos

Estatística do Teste



Teste de Hipótese

Teste de Hipótese para Média da População com desvio padrão populacional desconhecido

Exercício

- $\begin{cases} H_0 : \mu = 500g \\ H_1 : \mu \neq 500g \end{cases}$

$n = 31;$ $\mu = 500;$ $\bar{x} = 508;$ $S = 15;$

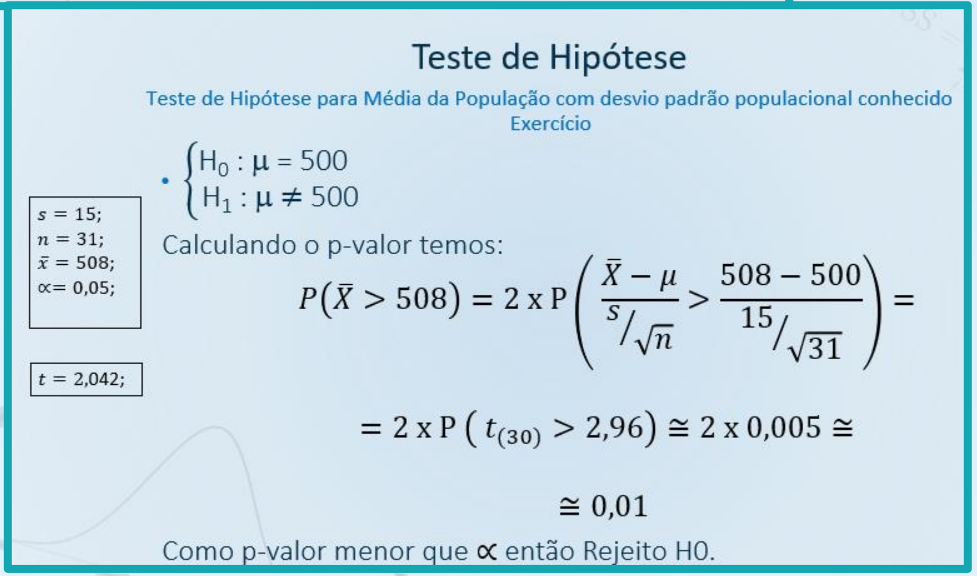
$t = 2{,}042;$

$\bar{x}_{c_1} = \mu + t\frac{S}{\sqrt{n}}$

$\bar{x}_{c_2} = \mu - t\frac{S}{\sqrt{n}}$

$$\bar{x}_{c_1} = \mu + t\frac{S}{\sqrt{n}} = 500 + 2{,}042\frac{15}{\sqrt{31}} = 500 + 5{,}50 = 505.5$$

$$\bar{x}_{c_2} = \mu - t\frac{S}{\sqrt{n}} = 500 - 5{,}50 = 494{,}5$$



Teste de Hipótese

Teste de Hipótese para Média da População com desvio padrão populacional conhecido

Exercício

- $\begin{cases} H_0 : \mu = 500 \\ H_1 : \mu \neq 500 \end{cases}$

$s = 15;$ $n = 31;$ $\bar{x} = 508;$ $\propto = 0{,}05;$

$t = 2{,}042;$

Calculando o p-valor temos:

$$P(\bar{X} > 508) = 2 \text{ x } \mathrm{P}\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} > \frac{508 - 500}{15/\sqrt{31}}\right) =$$

$$= 2 \text{ x } \mathrm{P}\left(t_{(30)} > 2{,}96\right) \cong 2 \text{ x } 0{,}005 \cong$$

$$\cong 0{,}01$$

Como p-valor menor que $\propto$ então Rejeito H0.

- Passos para um teste de hipótese (região crítica):
  1. Escolher o teste estatístico adequado, dado o objetivo
  2. Estabelecer a hipótese Nula (H_0) e a hipótese alternativa (H_1) do teste
  3. Fixar o nível de significância ($\propto$)
  4. Calcular o valor o valor observado da estatística do teste com base na amostra extraída da população
  5. Determinar a região crítica do teste em função do valor de $\propto$, fixado anteriormente
  6. Decidir se o valor da estatística pertence à região crítica, rejeita H_0, caso contrário, não rejeita H_0

entendendo estatística divertidamente

Testes Estatísticos

Estatística do Teste

Teste de Hipótese

Teste de Hipótese para Média da População com desvio padrão populacional conhecido

Exercício

- $\begin{cases} H_0 : \mu = 500 \\ H_1 : \mu \neq 500 \end{cases}$

$s = 15;$
$n = 31;$
$\bar{x} = 508;$
$\alpha = 0,05;$

$t = 2,042;$

Calculando o p-valor temos:

$$P(\bar{X} > 508) = 2 \text{ x } P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} > \frac{508 - 500}{15/\sqrt{31}}\right) =$$

$$= 2 \text{ x } P\left(t_{(30)} > 2,96\right) \cong 2 \text{ x } 0,005 \cong$$

$$\cong 0,01$$

Como p-valor menor que $\propto$ então Rejeito H0.

- Passos para um teste de hipótese (região crítica):
  1. Escolher o teste estatístico adequado, dado o objetivo
  2. Estabelecer a hipótese Nula (H_0) e a hipótese alternativa (H_1) do teste
  3. Fixar o nível de significância ($\propto$)
  4. Calcular o valor o valor observado da estatística do teste com base na amostra extraída da população
  5. Determinar o p-valor que corresponde à probabilidade associada ao valor da estatística do teste calculada no passo anterior
  6. Decide se o p-valor for menor do que o nível de significância $\propto$ estabelecido anteriormente, rejeita-se H_0. Caso contrário não rejeita-se H_0

Testes Estatísticos

Testes Paramétricos x Testes Não Paramétricos

- Testes Paramétricos
  - Envolvem parâmetros populacionais, para dados <u>quantitativos</u>
  - <u>Tem pressupostos:</u>
    - A amostra seguir uma distribuição normal
    - Homogeneidade de variâncias
    - As observações serem independentes
- Testes Não Paramétricos
  - As hipóteses são formuladas sobre características <u>qualitativas</u> da população
  - Tem muito menos suposições
    - por isso são chamados de <u>testes livres de distribuição</u>

Testes Estatísticos

Testes Paramétricos

- Testes Paramétricos
  - Normalidade
    - Teste de Kolgmogorv – Smirnov (K-S)
    - Teste de Shapiro Wilk
  - Homogeneidade de variância
    - Teste F de Levene
  - Comparação de médias
    - Teste t para uma amostra
    - Teste t para duas amostras
  - ANOVA
    - de 1 fator
    - fatorial

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Normalidade – Kolmogorov Smirnov

- Testes de aderência
- Compara a distribuição de frequência acumulada com uma distribuição teórica
- Teste que perde poder para pequenas amostras, sendo recomendado para n>30.

$$\begin{cases} H_0: amostra\ provém\ de\ uma\ população\ Normal \\ H_1: a\ amostra\ não\ provém\ de\ uma\ população\ Normal \end{cases}$$

$$\begin{cases} H_0:\ F_{obs}(X) = F_{esp}(X) \\ H_1: F_{obs}(X) \neq F_{esp}(X) \end{cases}$$

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

$$D_{cal} = máx\ \{\ |\ F_{esp}(X_i)\ - F_{obs}(X_i)\ |\ ;|\ F_{esp}(X_i)\ - F_{obs}(X_{i-1})\ |\ \}$$

Se $D_{cal} < D_c$ então não rejeito H_0

Testes Estatísticos

Testes Paramétricos – Normalidade – Shapiro Wilk

- Recomendado para $4 \leq n \leq 2.000$.

$$\begin{cases} H_0: amostra\ provém\ de\ uma\ população\ Normal \\ H_1: a\ amostra\ não\ provém\ de\ uma\ população\ Normal \end{cases}$$

$$W = \frac{b^2}{\sum_{i=1}^{n}(x_{(i)} - \bar{x})^2}$$

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)}) \text{ se n é par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)} \text{ se n é ímpar} \end{cases}$$

Pequenos valores de W_{cal} indicam não normalidade, então neste caso o teste de rejeição é inverso!!
Se $W_{cal} < W_c$ então rejeito H_0

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos

- Testes Paramétricos
  - Normalidade
    - Teste de Kolgmogorv – Smirnov (K-S)
    - Teste de Shapiro Wilk
  - Homogeneidade de variância
    - Teste F de Levene
  - Comparação de médias
    - Teste t para uma amostra
    - Teste t para duas amostras
  - ANOVA
    - de 1 fator
    - fatorial

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Variância – F de Levene

- Menos sensível ao desvio de normalidade.
- Considerado um teste mais robusto!

$$\begin{cases} H_0\colon \sigma_1^2 = \sigma_2^2 = \ \ldots = \sigma_k^2 \\ H_1\colon pelo\ menos\ um\ \text{é}\ \neq \end{cases}$$

$$F_{cal} = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2}$$

Se $F_{cal} < F_c$ então não rejeito H_0

Testes Estatísticos

Testes Paramétricos

- Testes Paramétricos
  - Normalidade
    - Teste de Kolgmogorv – Smirnov (K-S)
    - Teste de Shapiro Wilk
  - Homogeneidade de variância
    - Teste F de Levene
  - Comparação de médias
    - Teste t para uma amostra
    - Teste t para duas amostras
  - ANOVA
    - de 1 fator
    - fatorial

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Comparação de médias – teste t uma amostra

- Testa se a média de uma população é um valor fixo

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

$$T_{cal} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Se $T_{cal} < T_c$ ou $T_{cal} > -T_c$ então não rejeito H_0

Testes Estatísticos

Testes Paramétricos – Comparação de médias – teste t duas amostras independentes

- Testa se a média de uma população é um valor fixo

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

Quando $\sigma_1^2 = \sigma_2^2$

$$T_{cal} = \frac{(\overline{X_1} - \overline{X_2})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$

Quando $\sigma_1^2 \neq \sigma_2^2$

$$T_{cal} = \frac{(\overline{X_1} - \overline{X_2})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$gl = n_1 + n_2 - 2$$

Se $T_{cal} < T_c$ ou $T_{cal} > -T_c$ então não rejeito H_0

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Comparação de médias – teste t duas amostras emparelhadas

- Testa se a média de uma população é um valor fixo

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

$$\mu_d = \mu_{antes} - \mu_{depois}$$

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$S_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

$$T_{cal} = \frac{(\bar{d} - \mu_d)}{S_d / \sqrt{n}}$$

$$gl = n - 1$$

Se $T_{cal} < T_c$ ou $T_{cal} > -T_c$ então não rejeito H_0

Testes Estatísticos

Testes Paramétricos

- Testes Paramétricos
  - Normalidade
    - Teste de Kolgmogorv – Smirnov (K-S)
    - Teste de Shapiro Wilk
  - Homogeneidade de variância
    - Teste F de Levene
  - Comparação de médias
    - Teste t para uma amostra
    - Teste t para duas amostras
  - ANOVA
    - de 1 fator
    - fatorial

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Comparação de médias de mais de uma população – ANOVA 1 fator

- Objetivo: comparar médias de 3 ou mais populações, por meio da análise de variância amostrais.
- O teste se baseia em uma amostra extraída de cada população, com intuito de determinar se as diferenças entre as médias populacionais, ou se tais diferenças entre as médias populacionais, ou se tais diferenças são decorrentes apenas da variabilidade implícita da amostra.
- Existe as suposições: 1 – amostras independentes entre si, 2 – as populações são normais, 3 – variância homogêneas
- É uma extensão do teste t para duas amostras, mas só que ANOVA é para 3 ou mais

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1: pelo\ menos\ um\ é\ \neq \end{cases}$$

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Fonte de variação	Soma dos quadrados	Graus de liberdade	Quadrados médios	F
Entre os grupos	$SQF = \sum_{i=1}^{k} n_i \left(\bar{Y}_i - \bar{Y}\right)^2$	$k-1$	$QMF = \frac{SQF}{k-1}$	$F = \frac{QMF}{QME}$
Dentro dos grupos	$SQU = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$	$N-k$	$QME = \frac{SQU}{N-k}$	
Total	$SQT = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$	$N-1$		

Fonte: Fávero et al, 2017

Se $F_{cal} < F_{c\,(k-1, N-k, \alpha)}$ então não rejeito H_0

entendendo estatística divertidamente

Testes Estatísticos

Testes Paramétricos – Comparação de médias de mais de uma população – ANOVA fatorial

- Extensão da ANOVA de um fator, porém com dois ou mais fatores. Sendo Y numérico e X´s categóricos. Presume-se que Y é influenciado por um ou mais fatores e aproveita e teste também possível interações entre fatores (resultante da combinação do nível *i* do fator A e *j* do fator B)
- Testa se a média de cada nível do fator são iguais (efeito isolado de cada fator) e testa se há interação entre os fatores (efeito conjunto dos fatores).

$$\begin{cases} H_0^A: \mu_1 = \mu_2 = \cdots = \mu_a \\ H_1^A: pelo\ menos\ um\ é\ \neq \end{cases}$$

$$\begin{cases} H_0^B: \mu_1 = \mu_2 = \cdots = \mu_b \\ H_1^B: pelo\ menos\ um\ é\ \neq \end{cases}$$

$$\begin{cases} H_0^A: \gamma_{ij} = 0\ (não\ há\ interação\ entre\ A\ e\ B) \\ H_1^A: \gamma_{ij} \neq 0 (há\ interação\ entre\ A\ e\ B) \end{cases}$$

Se F_{cal} < F_c então não rejeito H_0

Fonte de variação	Soma dos quadrados	Graus de liberdade	Quadrados médios	F
Fator *A*	$SQF_A = b \cdot n \cdot \sum_{i=1}^{a} (\bar{Y}_i - \bar{Y})^2$	$a-1$	$QMF_A = \frac{SQF_A}{a-1}$	$F_A = \frac{QMF_A}{QME}$
Fator *B*	$SQF_B = a \cdot n \cdot \sum_{j=1}^{b} (\bar{Y}_j - \bar{Y})^2$	$b-1$	$QMF_B = \frac{SQF_B}{b-1}$	$F_B = \frac{QMF_B}{QME}$
Interação	$SQ_{AB} = n \cdot \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y})^2$	$(a-1) \cdot (b-1)$	$QM_{AB} = \frac{SQ_{AB}}{(a-1)\cdot(b-1)}$	$F_{AB} = \frac{QM_{AB}}{QME}$
Erro	$SQU = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{ij})^2$	$(n-1) \cdot ab$	$QME = \frac{SQU}{(n-1)\cdot ab}$	
Total	$SQT = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y})^2$	$N-1$		

Fonte: Fávero et al, 2017

Testes Estatísticos

Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

Testes Não Paramétricos – para uma amostra – Teste binomial

- Para variável binária
- Testa se a proporção observada é igual ao valor questionado

$$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$$

$$Z_{cal} = \frac{|N_{\hat{p}} - N_p| - 0{,}5}{\sqrt{N.p.q}}$$

Se $Z_{cal} < Z_c$ ou $Z_{cal} > -Z_c$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos – para uma amostra – Teste qui-quadrado

- Extensão do teste binomial, mas para 2 ou mais categorias, que podem ser nominais ou ordinais
- Testa se a média de uma população é um valor fixo

$$\begin{cases} H_0: não\ há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas \\ H_1: há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas \end{cases}$$

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

Se $\chi^2{}_{cal} < \chi^2{}_{c}$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

Testes Não Paramétricos – para 2 amostras emparelhadas – Teste McNemar

- Verifica se duas amostras relacionadas (antes e depois / emparelhadas) que são binárias tem alguma relação

	Depois	
Antes	**+**	**-**
+	A	B
-	C	D

$$\begin{cases} H_0: P(B \to C) = P(C \to B) & -\ não\ tem\ mudança \\ H_1: P(B \to C) \neq P(C \to B) & -\ tem\ mudança \end{cases}$$

$$\chi^2{}_{cal} = \frac{(|B - C| - 1)^2}{B + C}$$

Se $\chi^2{}_{cal} < \chi^2{}_c$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos – para 2 amostras emparelhadas – Teste Wilcoxon

- Extensão do teste do sinal, mas é mais poderoso! É uma alternativa para o teste t quando os pressupostos não são sanados.
- Este é mais poderoso que o test do sinal, pois leva em consideração, além da direção das diferenças para cada par, a magnitude da diferença dentro de um par!

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

$$\mu_d = Mediana_{antes} - Mediana_{depois}$$

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

$$\mu_d = \mu_{antes} - \mu_{depois}$$

$$Z_{cal} = \frac{\min(S_p, S_n) - {}^{N(N+1)}/_{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24} - \frac{\sum_{j=1}^{g} t_j^3 + \sum_{j=1}^{g} t_j}{48}}}$$

Se $Z_{cal} < Z_c$ ou $Z_{cal} > -Z_c$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

Testes Não Paramétricos – para 2 amostras independentes – Teste qui-quadrado

- Testa se existe alguma mudança no comportamento médio (ou seja, se tem associação)

$$\begin{cases} H_0: não\ há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas\ \ (não\ há\ associação) \\ H_1: há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas (há\ associação) \end{cases}$$

$$\chi^2{}_{cal} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$gl = (I-1)(J-1)$$

Se $\chi^2{}_{cal} < \chi^2{}_c$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos – para 2 amostras independentes – Teste dos Mann-Whitney

- Um dos teste não paramétricos mais poderosos, aplicado em variáveis quantitativa e qualitativa em escala ordinal.
- Alternativa para o teste t quando não tem normalidade.
- Os dados originais são transformados em postos (rank/ordenação), sendo não tão poderoso quanto o teste t. Diferente do teste t, ele testa igualdade de medianas

$$\begin{cases} H_0: mediana_1 = mediana_2 \\ H_1: mediana_1 \neq mediana_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$Z_{cal} = \frac{U - {}^{N_1.N_2}/_2}{\sqrt{\frac{N_1.N_2}{N(N-1)}\left(\frac{N^3 - N}{12} - \frac{\sum_{j=1}^{g} t_j^3 + \sum_{j=1}^{g} t_j}{12}\right)}}$$

Se $Z_{cal} < Z_c$ ou $Z_{cal} > -Z_c$ então não rejeito H_0

entendendo estatística divertidamente

Testes Estatísticos

Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

Testes Não Paramétricos – para *k* amostras emparelhadas – Teste de Friedman

- Aplicado para variáveis quantitativas ou qualitativas em escala ordinal.
- Extensão do teste de Wilcoxon para 3 ou mais amostras emparelhadas.
- Uma alternativa à ANOVA quando as suposições foram violadas.

$$\begin{cases} H_0: mediana_1 = mediana_2 = \cdots = mediana_k \\ H_1: pelo\ menos\ um\ \text{é}\ \neq \end{cases}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1: pelo\ menos\ um\ \text{é}\ \neq \end{cases}$$

$$F_{cal} = \frac{12}{N.K.(K+1)} \sum_{J=1}^{K} (R_J)^2 - 3.N.(K+1)$$

Se $F_{cal} < F_c$ então não rejeito H_0

entendendo estatística divertidamente

Testes Estatísticos

Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

Testes Não Paramétricos – para *k* amostras independentes – Teste qui-quadrado

- Teste muito conhecido e muito utilizado como teste de associação entre variáveis nominais ou ordinais

$$\begin{cases} H_0: \textit{não há diferença significativa entre as frequências observadas e esperadas} \ (\textit{não há associação}) \\ H_1: \textit{há diferença significativa entre as frequências observadas e esperadas} (\textit{há associação}) \end{cases}$$

$$\chi^2{}_{cal} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$gl = (I-1)(J-1)$$

Se $\chi^2{}_{cal} < \chi^2{}_{c}$ então não rejeito H_0

Testes Estatísticos

Testes Não Paramétricos – para *k* amostras independentes – Teste de Kruskal Wallis

- Alternativa à ANOVA quando não tem normalidade e homogeneidade de variância ou quando a variável é ordinal.
- Para k=2 o Kruskal Wallis é o mesmo do Mann-Whitney

$$\begin{cases} H_0: mediana_1 = mediana_2 = \cdots = mediana_k \\ H_1: pelo\ menos\ um\ é\ \neq \end{cases}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1: pelo\ menos\ um\ é\ \neq \end{cases}$$

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

$$H'_{cal} = \frac{H}{1 - \frac{\sum_{j=1}^{g}(t_j^3 - t_j)}{(N^3 - N)}}$$

Se $H'_{cal} > \chi^2_c$ rejeito H_0

Testes Estatísticos

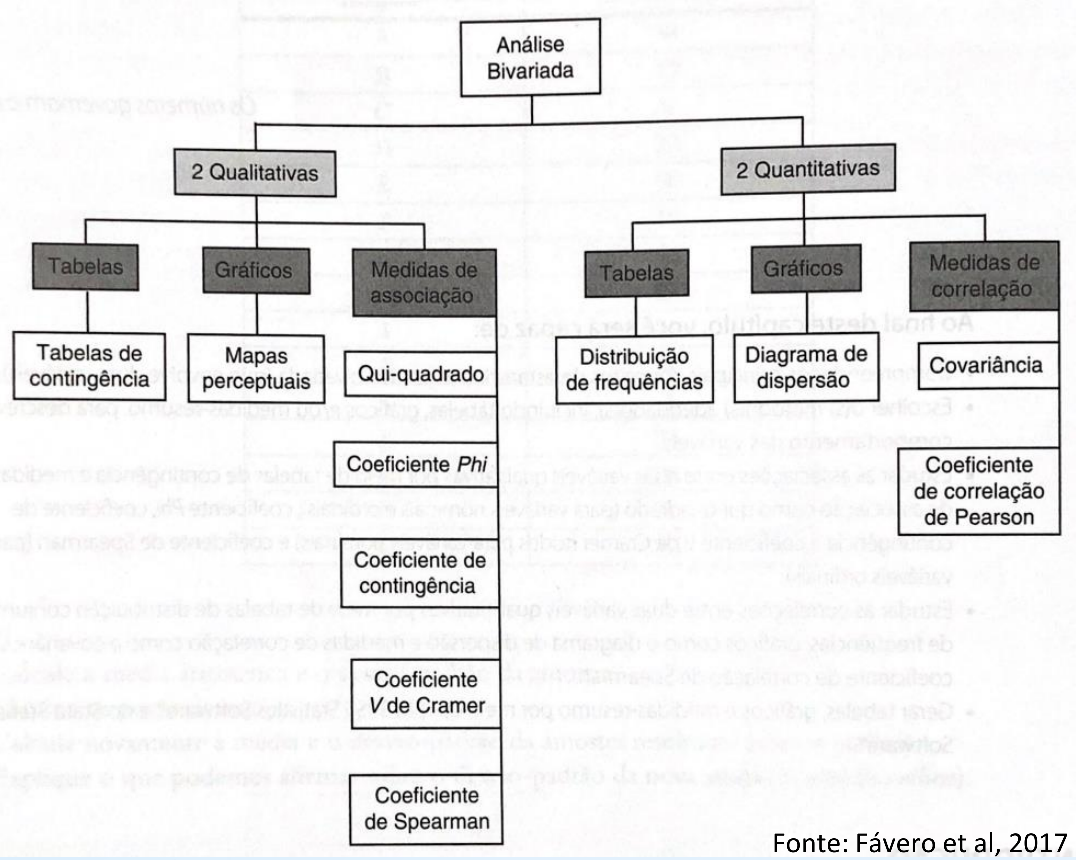
Testes Não Paramétricos

- Testes Não Paramétricos
  - Teste para uma amostra
    - Teste binomial (binario)
    - Teste qui-quadrado (nominal ou ordinal)
  - Teste para 2 amostras emparelhadas
    - Teste McNemar (binario)
    - Teste Wilcoxon (ordinal)
  - Teste para 2 amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste dos Mann-Whitney (ordinal)
  - Teste para k amostras emparelhadas
    - Teste de Friedman (ordinal)
  - Teste para k amostras independentes
    - Teste qui-quadrado (nominal ou ordinal)
    - Teste de Kruskal Wallis (ordinal)

Testes Estatísticos

e Medidas de Associação

- Associação entre duas variáveis qualitativas
  - Tabela de contingência
  - χ^2 (nominais e ordinais)
  - Coeficiente de contingência (nominais)
  - Coeficiente phi (nominais)
  - Coeficiente Cramers' V (nominais)
  - Coeficiente de Spearman (ordinais)
- Associação entre duas variáveis quantitativas
  - Gráfico de Dispersão
  - Covariância
  - Correlação de Pearson



Fonte: Fávero et al, 2017

Medidas de Associação

Estatística χ^2

- Medida baseada no teste qui-quadrado, apresentado anteriormente. Medida baseada no p-valor.
- P-valor representa a probabilidade associada ao valor esperado da amostra, p-valor representa um índice decrescente de confiabilidade de um resultado; quanto mais baixo seu valor, menos se pode acreditar na hipótese suposta.

$$\begin{cases} H_0: não\ há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas\ \ (não\ há\ associação) \\ H_1: há\ diferença\ significativa\ entre\ as\ frequências\ observadas\ e\ esperadas (há\ associação) \end{cases}$$

$$\chi^2{}_{cal} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

P-valor é calculado a partir do $\chi^2{}_{cal}$.

Medidas de Associação

Coeficiente *Phi*

- Medida de associação mais simples para variáveis nominais. É baseada no χ^2, podendo ser expressa por:

$$Phi = \sqrt{\frac{\chi^2}{n}}$$

Para que *phi* varie apenas entre 0 e 1, é necessário que a tabela de contingência seja 2x2.

Quanto maior o valor de *phi* maior a associação entre dias variáveis.

Medidas de Associação

Coeficiente *de contingência*

- Também conhecido como coeficiente de contingência de Pearson
- Medida de associação para variáveis nominais, baseada na estatística χ^2, podendo ser expressa por:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

O coeficiente de contingência tem como limite inferior o valor 0, indicando que não existe relação entre as variáveis, porém o limite superior varia em função do número de categorias, ou seja:

$$0 \leq C \leq \sqrt{\frac{q-1}{q}} \qquad q = \min(I, J)$$

Quando C máximo significa que existe associação perfeita entre as variáveis, porém, esse limite nunca assume o valor 1. Dois coeficientes de contingência só podem ser comparados se forem feitos de uma tabela de contingência do mesmo tamanho.

entendendo estatística divertidamente

Medidas de Associação

Coeficiente *V de Cramer*

- Outra medida de associação para variáveis nominais, baseada na estatística χ^2, podendo ser expressa por:

$$V = \sqrt{\frac{\chi^2}{n(q-1)}} \qquad q = \min(I, J)$$

Quando a tabela de contingência é 2x2 o V acaba sendo phi.
Este é uma alternativa ao coeficiente de phi, ao coeficiente de contingência e seu valor está sempre limitado no intervalo [0,1], independente da quantidade de categorias nas linhas e colunas.

O valor 0 indica que as variáveis não tem nenhum tipo de associação e o valor 1 revela que elas são perfeitamente associadas.

Medidas de Associação

Coeficiente *de Spearman*

- É uma medida de associação para variáveis ordinais.
- É baseada na ordenação, onde cria-se rankings

$$r_{sp} = 1 - \frac{6.\sum_{k=1}^{n} d_k^2}{n(n^2-1)} \qquad\qquad q = \min(I, J)$$

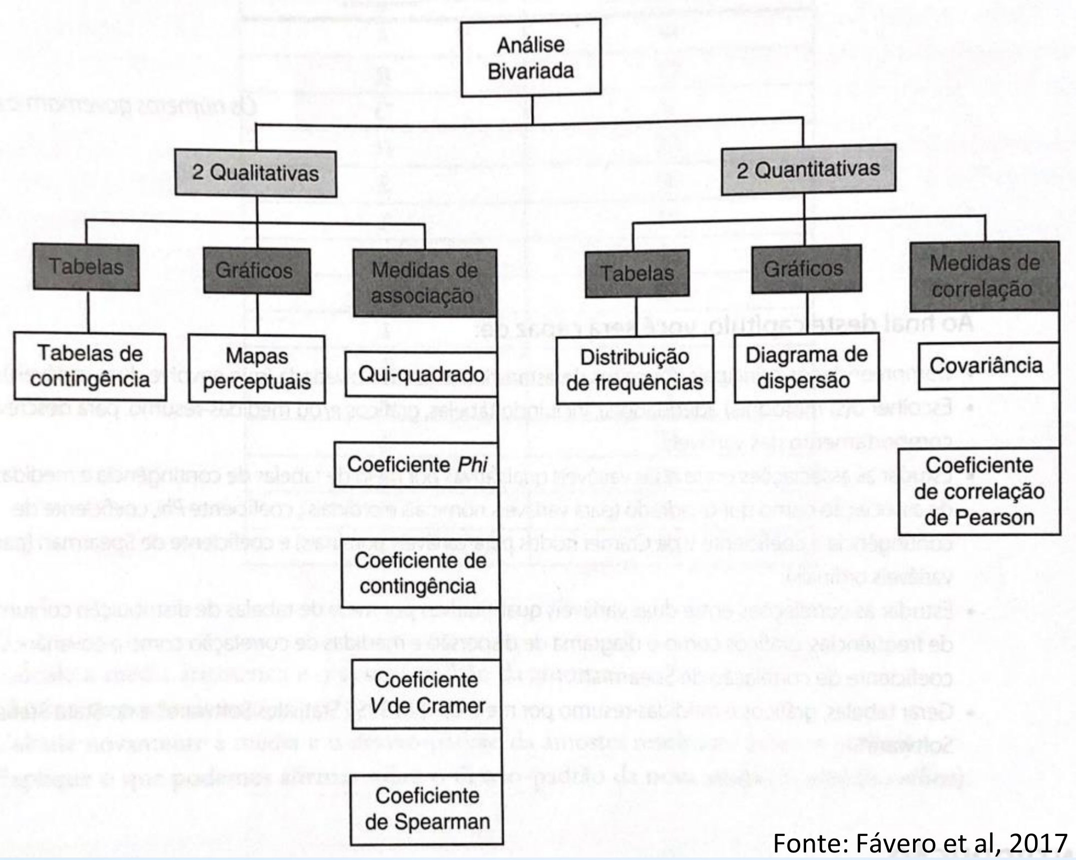
É uma medida que varia entre -1 e 1. Se:

- $r_{sp} = 1$ então todos os valores de d_k são nulos, indicando que todos os postos são iguais para as variáveis X e Y (associação positiva perfeita)
- $r_{sp} = -1$ é encontrado quando $\sum_{k=1}^{n} d_k^2 = \frac{n(n^2-1)}{3}$ atingir seu valor máximo (há a inversão nos valores dos postos das variáveis), indicando uma associação negativa perfeita.
- $r_{sp} = 0$ não há associação entre as variáveis X e Y

Testes Estatísticos

e Medidas de Associação

- Associação entre duas variáveis qualitativas
  - Tabela de contingência
  - χ^2 (nominais e ordinais)
  - Coeficiente de contingência (nominais)
  - Coeficiente phi (nominais)
  - Coeficiente Cramers' V (nominais)
  - Coeficiente de Spearman (ordinais)
- Associação entre duas variáveis quantitativas
  - Gráfico de Dispersão
  - Covariância
  - Correlação de Pearson



Fonte: Fávero et al, 2017

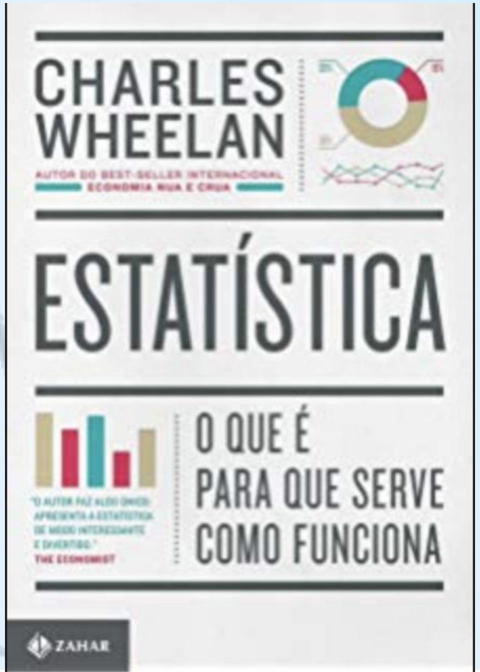
Não esqueça de deixar seu feedback!

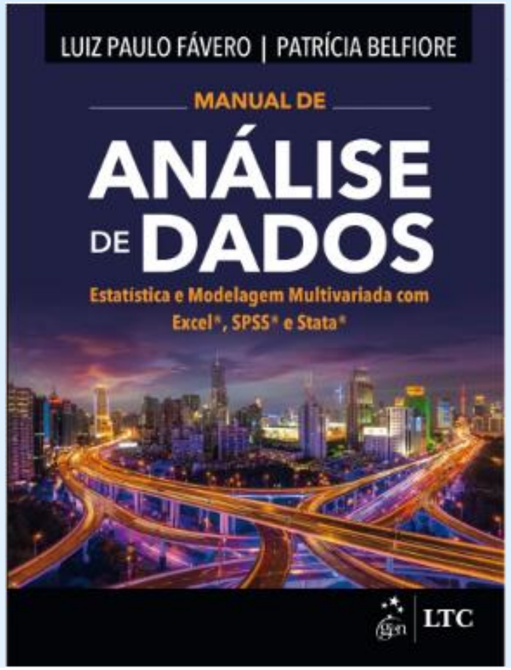
=]

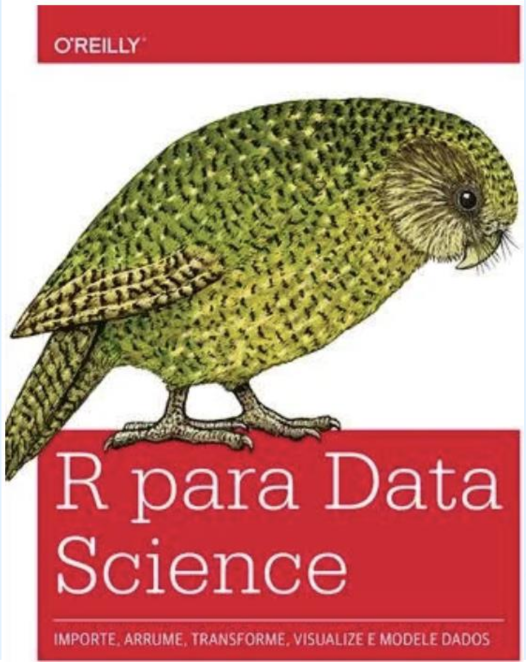
entendendo estatística divertidamente

Referência

- Moore, D., McCabe, G., Duckworth, W., Sclove, S. *A prática da Estatística Empresarial*. LTC, Rio de Janeiro, 2006.
- Anderson, D., Sweeney, D., Williams, T. *Estatística Aplicada à Administração e Economia*. Segunda Edição. Cengage Learning, São Paulo, 2011.
- www.asn.rocks
- www.curso-r.com







entendendo estatística divertidamente

It´s kind of fun to do the IMPOSSIBLE

dri@asn.rocks

/in/adrianamms

/in/asn.rocks

asn.rocks

www.asn.rocks