

BIOESTATÍSTICA

AULA 6

Anderson Castro Soares de Oliveira
Jose Nilton da Cruz

Departamento de Estatística/ICET/UFMT

Teste de hipótese

TESTE DE HIPÓTESE

- Teste de hipótese é uma metodologia estatística que permite tomar decisões sobre uma ou mais populações baseando-se no conhecimento de informações da amostra.
- A formulação de suposições ou de conjecturas acerca das populações são denominadas de Hipóteses Estatísticas, que podem ser:
 - HIPÓTESE NULA - É aquela Hipótese Estatística, prefixada, formulada sobre o parâmetro populacional estudado, e é sempre uma afirmativa. É representada por H_0 .
 - HIPÓTESE ALTERNATIVA - São quaisquer hipóteses que difiram da Hipótese Nula. Pode ser representada por H_1 ou H_a

TESTE DE HIPÓTESE

- Imagine que estamos interessados em responder a questão:
 - Níveis elevados de bilirrubina em recém-nascidos afetam a capacidade auditiva?
 - Dado uma amostra de n recém-nascidos, construímos dois grupos:
 - A = taxa de bilirrubina normal
 - B = taxa de bilirrubina elevada
 - Gera-se a hipótese de nula H_0 :

$$H_0 : \mu_A = \mu_B$$

- μ_A média da capacidade auditiva do grupo A
- μ_B média da capacidade auditiva do grupo B

TESTE DE HIPÓTESE

- Com a formulação dessa hipótese H_0 estamos testando se as médias auditivas dos dois grupos são iguais
- Aceitar H_0 significa afirmar que os níveis de bilirrubinas não afetam a audição!
- Rejeitar H_0 significa comprovar que os níveis de bilirrubina afetam a capacidade auditiva
- Poderíamos formular as seguintes hipóteses alternativas:
 - $H_1 : \mu_A \neq \mu_B$ as médias dos dois grupos são diferentes
 - $H_1 : \mu_A > \mu_B$ A média do grupo A é maior que a média do grupo B
 - $H_1 : \mu_A < \mu_B$ A média do grupo A é menor que a média do grupo B

TESTE DE HIPÓTESE

- Consideremos θ o parâmetro estudado e θ_0 valor inicialmente suposto para. Podemos formular as seguintes hipóteses:

$$\begin{array}{ll} \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array} \right. & \text{Teste Bilateral} \\ \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \right. & \text{Teste Unilateral} \\ \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{array} \right. & \text{Teste Unilateral} \end{array}$$

TESTE DE HIPÓTESE

- Aceitando-se H_0 rejeita-se automaticamente H_1 , e vice-versa
- Para se decidir se aceitamos ou se rejeitamos uma hipótese deve-se utilizar uma regra de decisão.
- As regras de decisão são baseadas nas distribuições amostrais.

Tabela 1: Erros possíveis de se cometer no processo de tomada de decisão

Decisões possíveis	Estados possíveis	
	Ho verdadeira	Ho falsa
Aceitação de Ho	Decisão correta	Erro do tipo II
Rejeição de Ho	Erro do tipo I	Decisão correta

TESTE DE HIPÓTESE

- Ao testar uma hipótese estabelecida, a probabilidade máxima com a qual se sujeitaria a correr o risco de um erro do tipo I é denominada de **Nível de Significância do Teste** e é representada por α .
- O nível de significância α é geralmente determinado pelo pesquisador antes da coleta dos dados.
- Em muitas aplicações da estatística, o nível de significância é tradicionalmente fixado em 0,05

TESTE DE HIPÓTESE

- Existem duas opções para expressar a conclusão final de um teste de hipóteses.
 - Comparar o valor da estatística do teste com um valor tabelado obtido a partir da distribuição teórica do teste;
 - Comparar o valor α com o nível descritivo do teste denominado *valor - p*

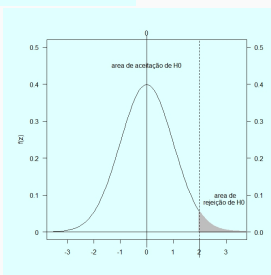
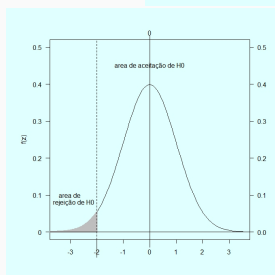
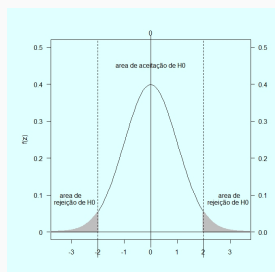
TESTE DE HIPÓTESE

- O valor-p é a probabilidade de que a estatística do teste tenha valor extremo em relação ao valor observado quando a hipótese H_0 é verdadeira.
- Alguns enganos em relação ao valor-p
 - O valor-p não é a probabilidade da hipótese nula de um teste ser verdadeira.
 - O valor-p não é a probabilidade de um dado resultado ter sido obtido de um "acaso".
 - O valor-p não é a probabilidade da hipótese nula ter sido enganosamente rejeitada.
 - Valor-p e nível de significância não são sinônimos:
 - O valor-p é sempre obtido a partir da amostra
 - O nível de significância é geralmente fixado antes da coleta dos dados.

TESTE DE HIPÓTESE

- A regra de decisão do teste de hipóteses a partir do valor-p será da seguinte forma:
 - Se o valor-p for maior que α então a hipótese nula será aceita.
 - Se o valor-p for menor ou igual a α então a hipótese nula será rejeitada.

TESTE DE HIPÓTESE



TESTE DE HIPÓTESE

Escolha do teste de hipótese

- Quantas variáveis serão analisadas ?
- Qual seu objetivo?
 - Comparação de médias
 - Comparação de variância
 - Relação entre variáveis
- Quais o tipo de variável ?
 - Quantitativa
 - Qualitativa
- Qual a natureza da distribuição de probabilidade da variável?
 - Distribuição Binomial
 - Distribuição Poisson
 - Distribuição Normal

Escolha do teste de hipótese

- Teste de hipóteses para variáveis categóricas
 - Usualmente, usam-se testes que comparam proporções, tais como o Teste Qui-Quadrado e o Exato de Fisher.
- Teste de hipóteses para variáveis quantitativas:
 - Para variáveis com distribuição normal -testes que comparam médias: teste-t e ANOVA (Análise de Variância);
 - Para variáveis com distribuição não normal- testes que comparam medianas, tais como Testes de Mann-Witney, Wilcoxon e Kruskal-Wallis (testes não-paramétricos)

TESTE DE HIPÓTESE

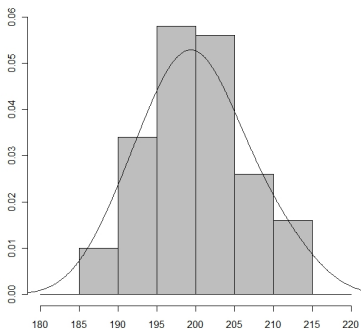
Etapas para um teste de hipótese

- Determinar as hipóteses nula e alternativa.
- Selecionar a estatística de teste que será usada para decidir rejeitar ou não a hipótese nula.
- Especificar o nível de significância α para o teste.
- Usar o nível de significância α para desenvolver regra de decisão que indica os valores críticos da estatística de teste que levará a rejeição de H_0 .
- Coletar os dados amostrais e calcular a estatística de teste.
- Comparar o valor da estatística do teste com o(s) valor(es) crítico(s) especificado(s) na regra de decisão para determinar se H_0 deve ser rejeitado.

Teste de Normalidade

TESTE DE NORMALIDADE

- Os testes de normalidade são usados para verificar se a distribuição de probabilidade associada a uma dada variável é a distribuição normal



Teste de Shapiro-Wilk

- O teste de Shapiro-Wilk W tem sido o mais utilizado para testar a normalidade e tem uma melhor performance em amostras reduzidas $n < 30$
- As hipóteses geradas são:
$$\begin{cases} H_0 : \text{A amostra provém de uma população com distribuição Normal} \\ H_1 : \text{A amostra não provém de uma população com distribuição Normal} \end{cases}$$

TESTE DE NORMALIDADE

Teste de Shapiro-Wilk

- Para aplicar o teste as observações são ordenadas em ordem decrescente:

$$x(i) : x(1) \leq x(2) \leq x(3) \leq \dots \leq x(n)$$

- A estatística W de teste para normalidade é definida como

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- Os valores a_i 's são obtidos de tabelas

Teste de Shapiro-Wilk

- Se o valor calculado de W é estatisticamente significativo (para $\alpha = 0,05$) rejeita-se a hipótese que a distribuição estudada é normal
- Ou seja Para a Distribuição ser considerada Normal o valor- p deve ser maior que 0,05.

TESTE DE NORMALIDADE

Teste de Kolmogorov-Smirnov

- O teste de Kolmogorov-Smirnov é utilizado para verificar a normalidade e baseia-se na máxima diferença entre a distribuição acumulada da amostra e distribuição acumulada esperada e deve ser utilizado em amostras grandes ($n \geq 30$)
- As hipóteses geradas são:
$$\begin{cases} H_0 : \text{A amostra provém de uma população com distribuição Normal} \\ H_1 : \text{A amostra não provém de uma população com distribuição Normal} \end{cases}$$
- Se o valor calculado de D é estatisticamente significativo (para $\alpha = 0,05$) rejeita-se a hipótese que a distribuição estudada é normal
- Ou seja: para a Distribuição ser considerada Normal o valor-p deve ser maior que 0,05

TESTE DE NORMALIDADE

Teste de Kolmogorov-Smirnov

- Para aplicar o teste as observações são ordenadas em ordem decrescente:

$$x(i) : x(1) \leq x(2) \leq x(3) \leq \dots \leq x(n)$$

- A estatística W de teste para normalidade é definida como

$$D = \sup |F(x(i)) - F_0(x(i))|$$

- $F(x(i))$ é a probabilidade acumulada na amostra
- $F_0(x(i))$ é a probabilidade acumulada da distribuição normal
- Se o valor calculado de D é estatisticamente significativo (para $\alpha = 0,05$) rejeita-se a hipótese que a distribuição estudada é normal
- Ou seja: para a Distribuição ser considerada Normal o valor-p deve ser maior que 0,05

TESTE DE NORMALIDADE

Utilizando o R

- Para fazer o teste de Shapiro-Wilk utiliza-se
 - Estatísticas ⇒ Resumos ⇒
Teste de normalidade
Escolher a opção Shapiro-Wilk
- Para fazer o teste de Kolmogorov-Smirnov utiliza-se
 - Estatísticas ⇒ Resumos ⇒
Teste de normalidade
Escolher a opção Lilliefors (Kolmogorov-Smirnov)

Teste de Homogeneidade de Variâncias

TESTE DE HOMOGENEIDADE DE VARIÂNCIAS

- Um problema comum é o pesquisador observa dois ou mais grupos distintos com o objetivo de saber se os grupos diferem ou não em relação à resposta de interesse.
- Quando a variável de interesse é quantitativa, a estratégia pra verificar se os grupo diferem ou não é comparar suas médias.
- Antes de aplicar o teste de médias é necessário saber se este grupos possuem variância homogêneas, ou seja se a forma como variam os dados é a mesma nos grupos
- Para verificar se a variação é semelhante entre grupos utiliza-se os testes de homogeneidade de variâncias

TESTE DE HOMOGENEIDADE DE VARIÂNCIAS

Teste de Bartlett

- O teste de Bartlett é utilizado para verificar se a variância de diferentes grupos são homogêneas

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_I^2 \text{ vs } H_a : \exists i, j, \sigma_i^2 \neq \sigma_j^2$$

- O teste de Bartlett é fortemente sensível à Normalidade das observações subjacentes
- Uma regra comum é considerar que o teste apenas deve ser usado quando cada grupo tiver 5 mais repetições $n_i > 5$

TESTE DE HOMOGENEIDADE DE VARIÂNCIAS

Teste de Levene

- O teste de Levene é utilizado para verificar se a variância de diferentes grupos são homogêneas

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 \text{ vs } H_a : \exists i, j, \sigma_i^2 \neq \sigma_j^2$$

- Uma vantagem do teste de Levene é que ele não exige normalidade dos dados.

TESTE DE HOMOGENEIDADE DE VARIÂNCIAS

Utilizando o R

- Para fazer o teste de Bartlett utiliza-se
 - Estatísticas $\Rightarrow$ Variâncias $\Rightarrow$ Teste de Bartlett
- Para fazer o teste de Levene utiliza-se
 - Estatísticas $\Rightarrow$ Variâncias $\Rightarrow$ Teste de Bartlett

Teste de comparação de duas médias

TESTE PARA COMPARAÇÃO DE DUAS MÉDIAS

- Quando tem dois grupo distintos e a variável é quantitativa pode-se avaliar a igualdade entre os grupos
- Para verificar se os grupo diferem ou não é utilizado os teste de hipótese para comparação de duas médias.
- Os teste de comparação de duas médias podem ser:
 - Teste para amostras independentes - quando é medido a mesma variável sobre indivíduos diferentes após um tempo ou um tratamento.
 - Teste para amostras dependentes ou pareadas - quando é medido a mesma variável sobre o mesmo individuo após um tempo ou um tratamento.

TESTE PARA COMPARAÇÃO DE DUAS MÉDIAS

- Antes de aplicar um teste de comparação de duas médias é necessário verificar se os dados seguem distribuição normal ou não.
 - Se os dados tem distribuição normal utiliza-se teste paramétrico (teste t)
 - Se os dados não tem distribuição normal utiliza-se teste não paramétrico
- Para verificar se os dados seguem uma distribuição normal aplica-se um teste de normalidade.

Teste t

TESTE T

- Se a variável de interesse tem distribuição normal ou aproximadamente normal, utiliza-se o teste t para comparar duas médias.
- O teste t para duas médias tem duas abordagens distintas
 - Amostras independentes
 - Amostras dependentes ou pareadas

TESTE T - AMOSTRAS INDEPENDENTES

- Para aplicar o teste t para duas médias independentes é necessário verificar se as variâncias dos grupos são homogêneas, pode-se utilizar os teste de Bartlett ou Levene.
- O teste de hipótese pode ser formulado como segue:

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right. \quad \left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array} \right. \quad \left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array} \right.$$

TESTE T - AMOSTRAS INDEPENDENTES

- Quando as variâncias são consideradas homogêneas, para testar a hipótese de igualdade entre duas médias, se calcula a estatística

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_d^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- em que S_d^2 é uma variância combinada dada por

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

- Rejeita-se H_0 se $\text{valor} - p \leq \alpha$ ou $|t_c| > t_\alpha$, em que t_α é obtido da distribuição t com $n_1 + n_2 - 2$ graus de liberdade

TESTE T - AMOSTRAS INDEPENDENTES

- Quando as variâncias são consideradas heterogêneas, para testar a hipótese de igualdade entre duas médias, se calcula a estatística

$$t_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}}$$

- Rejeita-se H_0 se $\text{valor} - p \leq \alpha$ ou $|t_c| > t_\alpha$, em que t_α é obtido da distribuição t com $\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$ graus de liberdade

TESTE T - AMOSTRAS INDEPENDENTES

Exemplo

- Um estudo foi realizado para verificar a influencia da adição de cálcio na alimentação de frangos sobre o peso e parâmetros densitometria óssea.
- Foram utilizados 60 pintos de um dia distribuídos em dois grupos, alimentado durante 45 dias com ração com e sem adição de cálcio.
- Aos 45 dias os frangos foram abatidos pesados e foram medidos os diâmetros das epífises proximais e distais e diáfises dos tibiotarsos de cada grupo foram avaliados e posteriormente destinados à análise densitométrica.
- Os dados obtidos estão contidos no arquivo "calcio.xlsx".

TESTE T - AMOSTRAS INDEPENDENTES

Exemplo

- Para a variável diáfise
 - teste de Kolmogorov-Smirnov: aceita-se a hipótese de normalidade (valor-p=0,0978)
 - teste de Bartlett: aceita-se a hipótese de homogeneidade de variância (valor-p=0,1069)
- Utiliza-se o Teste t com variâncias homogêneas
- Temos:

$$\bar{X}_{com} = 2,26 \quad \bar{X}_{sem} = 1,76$$

$$S_{com}^2 = 0,56 \quad S_{sem}^2 = 0,31$$

$$S_d^2 = \frac{(29)0,56 + (29)0,31}{58} = 0.435$$

TESTE T - AMOSTRAS INDEPENDENTES

Exemplo

- Vamos testar

$$\begin{cases} H_0 : \mu_{com} = \mu_{sem} \\ H_1 : \mu_{com} \neq \mu_{sem} \end{cases}$$

- Assumindo $\alpha = 0,05$ e a estatística de teste

$$t_c = \frac{2,26 - 1,76}{\sqrt{0,435 \left(\frac{1}{30} + \frac{1}{30} \right)}} = 2,93$$

- Com grau de liberdade $\nu = 58$, temos $\text{valor} - p = 0,0048$.
- Temos que como $\text{valor} - p \leq \alpha$, assim rejeita-se H_0 ao nível de 5% de significância.
- Conclusão: Portanto, temos evidências o cálcio afeta a diá-fise, pois as médias de com e sem calcio são diferentes.

TESTE T - AMOSTRAS INDEPENDENTES

Exemplo

- Para a variável peso:
 - teste de Kolmogorov-Smirnov: aceita-se a hipótese de normalidade (valor-p=0,3135)
 - teste de Bartlett: rejeita a hipótese de homogeneidade de variância (valor-p=<0,0001)
- Utiliza-se o Teste t com variâncias heterógeneas
- Temos:

$$\bar{X}_{com} = 1961,87 \quad \bar{X}_{sem} = 1952,87$$

$$S_{com}^2 = 335,56 \quad S_{sem}^2 = 2028,60$$

$$\nu = \frac{\left(\frac{335,56}{30} + \frac{2028,60}{30} \right)^2}{\frac{\left(\frac{335,56}{30} \right)^2}{29} + \frac{\left(\frac{2028,60}{30} \right)^2}{29}} = 38,33$$

TESTE T - AMOSTRAS INDEPENDENTES

Exemplo

- Vamos testar

$$\begin{cases} H_0 : \mu_{com} = \mu_{sem} \\ H_1 : \mu_{com} \neq \mu_{sem} \end{cases}$$

- Assim,

$$t_c = \frac{1961,87 - 1952,87}{\sqrt{\frac{335,56}{30} + \frac{2028,60}{30}}} = 1,01$$

- Considerando grau de liberdade $\nu = 38,33$, temos *valor - p* = 31.88.
- Temos que como *valor - p* $\leq \alpha$, assim rejeita-se H_0 ao nível aceita-se H_0 ao nível de 5% de significância.
- Conclusão: Portanto, não temos evidências que o cálcio afeta o peso, pois as médias de com e sem calcio são iguais.

TESTE T - AMOSTRAS INDEPENDENTES

Utilizando o R

- Para fazer o Teste t para duas médias independentes
 - Estatísticas $\Rightarrow$ Médias $\Rightarrow$
Teste t para amostras independentes
 - Em seguida escolher a variável a ser analisada e os grupos.
 - Na aba opções deve-se informar:
 - Se o teste é unilateral ou bilateral
 - Nível de confiança do teste
 - Se o teste é com variância homogênea ou heterogênea

TESTE T - AMOSTRAS PAREADAS

- Num estudo pareado, temos duas amostras mas cada observação da primeira amostra é pareada com uma observação da segunda amostra.
- Neste tipo de estudo o teste é baseado nas diferenças d_i dos valores em cada par

$$d_i = x_{2i} - x_{1i}$$

- Assim a diferença entre os valores de cada par deve seguir uma distribuição normal

TESTE T - AMOSTRAS PAREADAS

- Para testar a hipótese de igualdade entre duas médias, se calcula a estatística

$$t_c = \frac{\bar{d}}{\sqrt{\frac{S_d^2}{n}}}$$

em $\bar{d}$ e S_d^2 são respectivamente a média e a variância das diferenças

- Rejeita-se H_0 se *valor* – $p \leq \alpha$ ou $|t_c| > t_\alpha$, em que t_α é obtido da distribuição t com $\nu = n - 1$ graus de liberdade

TESTE T - AMOSTRAS PAREADAS

Exemplo

- Um estudo observou o efeito de uma determinada dieta sobre o colesterol em 40 cães. Os cães foram submetidos a dieta durante o período de 60 dias e mediu-se o colesterol antes e depois da aplicação da dieta.

animal	CA	CB	animal	CA	CB
1	54,72	42,06	21	63,23	42,53
2	54,90	49,64	22	54,32	39,18
3	58,59	47,66	23	47,27	51,65
4	53,79	49,76	24	62,84	42,71
5	65,90	39,69	25	51,51	43,16
6	54,98	50,68	26	56,31	43,79
7	52,89	54,23	27	54,46	37,63
8	63,42	47,23	28	58,57	31,38
9	60,93	47,15	29	59,63	47,78
10	60,56	44,15	30	60,98	46,40
11	62,66	44,27	31	55,94	39,48
12	47,05	43,48	32	51,96	45,53
13	56,05	47,21	33	53,89	28,98
14	62,60	36,86	34	62,55	43,97
15	48,77	51,38	35	53,16	28,86
16	65,22	40,25	36	55,15	45,80
17	74,72	48,25	37	62,39	42,30
18	64,64	46,64	38	45,08	48,67
19	58,91	45,39	39	49,76	37,43
20	57,13	36,74	40	54,61	45,28

TESTE T - AMOSTRAS PAREADAS

Exemplo

- Para testar o efeito da dieta, primeiro vamos obter as diferenças.

animal	CA	CB	d	animal	CA	CB	d
1	54,72	42,06	12,66	21	63,23	42,53	20,70
2	54,90	49,64	5,26	22	54,32	39,18	15,14
3	58,59	47,66	10,93	23	47,27	51,65	-4,38
4	53,79	49,76	4,03	24	62,84	42,71	20,13
5	65,90	39,69	26,21	25	51,51	43,16	8,35
6	54,98	50,68	4,30	26	56,31	43,79	12,52
7	52,89	54,23	-1,34	27	54,46	37,63	16,83
8	63,42	47,23	16,19	28	58,57	31,38	27,19
9	60,93	47,15	13,78	29	59,63	47,78	11,85
10	60,56	44,15	16,41	30	60,98	46,40	14,58
11	62,66	44,27	18,39	31	55,94	39,48	16,46
12	47,05	43,48	3,57	32	51,96	45,53	6,43
13	56,05	47,21	8,84	33	53,89	28,98	24,91
14	62,60	36,86	25,74	34	62,55	43,97	18,58
15	48,77	51,38	-2,61	35	53,16	28,86	24,30
16	65,22	40,25	24,97	36	55,15	45,80	9,35
17	74,72	48,25	26,47	37	62,39	42,30	20,09
18	64,64	46,64	18,00	38	45,08	48,67	-3,59
19	58,91	45,39	13,52	39	49,76	37,43	12,33
20	57,13	36,74	20,39	40	54,61	45,28	9,33

TESTE T - AMOSTRAS PAREADAS

Exemplo

- Para a variável diferença:
 - teste de Kolmogorov-Smirnov: aceita-se a hipótese de normalidade (valor-p=0,8014)
- A partir das diferenças obtém-se:

$$\bar{d} = 13,67 \quad S^2 = 74,82 \quad \overline{CA} = 57,30 \quad \overline{CB} = 43,63$$

- Vamos testar

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases}$$

TESTE T - AMOSTRAS PAREADAS

Exemplo

- Assim,

$$t_c = \frac{13,67}{\sqrt{\frac{74,82}{40}}} = 9,99$$

- Considerando grau de liberdade $\nu = 39$, temos $\text{valor} - p < 0,0001$.
- Temos que como $\text{valor} - p \leq \alpha$, assim rejeita-se H_0 ao nível de 5% de significância.
- Conclusão: Portanto, temos evidências que a dieta afeta o colesterol médio dos cães.

TESTE T - AMOSTRAS PAREADAS

Usando o R

- Para fazer o Teste t para duas médias pareadas
 - Estatísticas $\Rightarrow$ Médias $\Rightarrow$
Teste t (dados pareados)
 - Em seguida escolher a variável na primeira amostra e segunda amostra
 - Na aba opções:
 - Se o teste é unilateral ou bilateral
 - Nível de confiança do teste

Teste Não Parametrico

TESTE NÃO PARAMÉTRICO

- Quando queremos comparar dois grupos e a hipótese de normalidade dos dados não é atendida podemos utilizar teste não paramétricos.
- Alguns testes não paramétricos são baseados no posto (score ou rank) de um valor de um conjunto de n valores, que corresponde a sua posição no conjunto ordenado (posto, score ou rank). Exemplo
 - $X = \{3, 5, 7, 12, 15, 20\} \Rightarrow \text{Posto} = \{1, 2, 3, 4, 5, 6\}$
- Para comparar dois grupos pode-se utilizar dois teste distintos:
 - Amostras independentes - teste de Mann-Whitney
 - Amostras dependentes ou pareadas - teste de Wilcoxon

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

- O teste de Mann-Whitney é usado para a comparação de dois grupos independentes.
- O teste de hipótese pode ser formulado como segue:

$$\begin{cases} H_0 : & \text{As duas amostras têm distribuições idênticas} \\ H_1 : & \text{As duas amostras têm distribuições diferentes} \end{cases}$$
$$\begin{cases} H_0 : & \mu_1 = \mu_2 \\ H_1 : & \mu_1 \neq \mu_2 \end{cases}$$

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

- Para testar a hipótese de igualdade entre duas médias, se calcula a estatística

$$U = \min\{U_1, U_2\}$$

- sendo

$$U_1 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - w_2 \quad U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - w_1$$

- n_1 é o tamanho de amostra grupo 1 e n_2 é o tamanho de amostra do grupo 2;
- w_1 é a soma dos postos do grupo 1 e nw_2 é a soma dos postos do grupo 2;

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

- Se ambas as amostras tiverem tamanhos superiores a 10 observações, pode fazer-se a aproximação à distribuição normal, com parâmetros

- Média $\mu_U = \frac{n_1 n_2}{2}$

- Desvio Padrão sem empate $\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

- Desvio Padrão com empate $\sigma_U = \sqrt{\frac{n_1 n_2 (n^3 - n - \sum (u_i^3 - u_i))}{12(n^2 - n)}}$

em que u_i é o numero de elementos dentro do empate

- Calcula-se $z_c = \frac{U - \mu_U}{\sigma_U}$

- Rejeita-se H_0 se $\text{valor} - p \leq \alpha$ ou $|z_c| > z_\alpha$

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

Exemplo

- Um estudo foi realizado com o objetivo de estimar os efeitos da inalação prolongada de óxido de cádmio.
- Os animais foram divididos em dois grupos, o primeiro grupo composto por 15 animais sujeitas em laboratório a um ambiente contaminado com este óxido, e o segundo grupo com 10 animais estiveram num ambiente normal sem essa contaminação (grupo de controle).
- A variável de interesse é a concentração de hemoglobina após a exposição.

Controle	15,0	15,0	16,0	16,2	16,3	16,8	16,9	17,1	17,4	17,5
Exposto	13,7	13,8	14,0	14,1	14,1	14,2	14,4	15,3	15,3	15,6
	15,7	15,9	16,5	16,6	16,7					

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

Exemplo

- Primeiro ordenar como se fosse um só grupo

grupo	exposto	posto	grupo	exposto	posto
exposto	13,7	1	exposto	15,9	14
exposto	13,8	2	controle	16,0	15
exposto	14,0	3	controle	16,2	16
exposto	<u>14,1</u>	4,5	controle	16,3	17
exposto	<u>14,1</u>	4,5	exposto	16,5	18
exposto	14,2	6	exposto	16,6	19
exposto	14,4	7	exposto	16,7	20
controle	<u>15,0</u>	8,5	controle	16,8	21
controle	<u>15,0</u>	8,5	controle	16,9	22
exposto	<u>15,3</u>	10,5	controle	17,1	23
exposto	<u>15,3</u>	10,5	controle	17,4	24
exposto	15,6	12	controle	17,5	25
exposto	15,7	13			

- Nos valores sublinhados ocorrem empates, por exemplo o valor 14,1, ocorre no 4º e 5º elemento, assim o posto é obtido da média dos dois valores ficando 4,5.

TESTE DE MANN-WHITNEY - AMOSTRAS INDEPENDENTES

Exemplo

- A partir dos dados e dos postos temos:

$$n_1 = 15 \quad n_2 = 10 \quad w_1 = 145 \quad w_2 = 180$$

$$U_1 = 15 \times 10 + \frac{10 \times 11}{2} - 180 = 25$$

$$U_2 = 15 \times 10 + \frac{15 \times 16}{2} - 145 = 125$$

- Assim, a estatística de teste é $U = \min\{25, 125\} = 25$
- Usando a aproximação à distribuição normal, temos:

- Média: $\mu_U = \frac{15 \times 10}{2} = 75$

- Desvio Padrão, temos 3 empates 14,1 com 2 elementos, 15,0 com 2 elementos e 15,3 15,0 com 2 elementos

$$\sigma_U = \sqrt{\frac{15 \times 10(25^3 - 25 - [(2^3 - 2) + (2^3 - 2) + (2^3 - 2)])}{12(25^2 - 25)}} = 18,02$$

Exemplo

- Assim

$$z_c = \frac{25 - 72}{18,02} = -2,77$$

- Temos $\text{valor} - p = 0,0056$.
- Temos que como $\text{valor} - p \leq \alpha$, assim rejeita-se H_0 ao nível de 5% de significância.
- Conclusão: Portanto, existe diferença entre os dois grupos, assim a exposição a óxido de cádmio afeta a hemoglobina.

Utilizando o R

- Para fazer o Teste de Mann-Whitney para duas médias independentes
 - Estatísticas $\Rightarrow$ Teste Não Paramétricos $\Rightarrow$ Teste de Wilcoxon (2 amostras)
 - Em seguida escolher a variável a ser analisada e os grupos.
 - Também deve-se informar:
 - Se o teste é unilateral ou bilateral
 - Nível de confiança do teste

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

- O teste de Wilcoxon é usado para a comparação de dois grupos quando os dados são obtidos através do esquema de pareamento.
- O teste de hipótese pode ser formulado como segue:

$$\begin{cases} H_0 : & \text{As duas amostras têm distribuições idênticas} \\ H_1 : & \text{As duas amostras têm distribuições diferentes} \end{cases}$$
$$\begin{cases} H_0 : & \mu_1 = \mu_2 \\ H_1 : & \mu_1 \neq \mu_2 \end{cases}$$

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

- Para testar a hipótese de igualdade entre duas médias, é necessário primeiro obter uma amostra de diferenças d_i

$$d_i = x_{2i} - x_{1i}$$

As diferenças iguais a zero devem ser excluídas.

- Em seguida ignorar os sinais das diferenças e atribuir postos a elas.
- Para testar a hipótese de igualdade entre duas médias, se calcula a estatística

$$T = \min\{T_+, T_-\}$$

- T_+ é a soma dos postos das diferenças positivas
- T_- é a soma dos postos das diferenças negativas

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

- Se a amostra tiver tamanho superiores a 10 observações, pode fazer-se a aproximação à distribuição normal, com parâmetros

- Média $\mu_T = \frac{n(n+1)}{4}$

- Desvio Padrão sem empate $\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

- Desvio Padrão com empate

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum u_i^3 - \sum u_i}{48}}$$

em que u_i é o número de elementos dentro do empate

- Calcula-se $z_c = \frac{U - \mu_U}{\sigma_U}$

- Rejeita-se H_0 se $\text{valor} - p \leq \alpha$ ou $|z_c| > z_\alpha$

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

Exemplo

- Um estudo foi realizado com o objetivo de verificar o efeito de uma ração sobre o peso dos animais.
- Dez animais adultos foram submetidas ao tratamento com a ração durante uma semana.
- Os pesos, em gramas, no princípio e no fim da semana, são designados respectivamente por peso 1 e peso 2.

Animal	1	2	3	4	5	6	7	8	9	10
peso1	635	704	662	560	603	745	698	575	633	669
peso2	640	712	681	558	610	740	707	585	635	682

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

Exemplo

- Obter os postos

Animal	peso1	peso2	d	d	posto	-	+
9	633	635	2	<u>2</u>	1,5		1,5
4	560	558	-2	<u>2</u>	1,5	1,5	
1	635	640	5	<u>5</u>	3,5		3,5
6	745	740	-5	<u>5</u>	3,5	3,5	
5	603	610	7	7	5		5
2	704	712	8	8	6		6
7	698	707	9	9	7		7
8	575	585	10	10	8		8
10	669	682	13	13	9		9
3	662	681	19	19	10		10
Total						5	50

- Nos valores sublinhados ocorrem empates

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

Exemplo

- Assim, a estatística de teste é $U = \min\{5, 50\} = 5$
 - Usando a aproximação à distribuição normal, temos:
 - Média $\mu_T = \frac{10(10+1)}{4} = 27,5$
 - Desvio Padrão sem empate $\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$
 - Média: $\mu_U = \frac{15 \times 10}{2} = 75$
 - Desvio Padrão, temos 2 empates 2 com 2 elementos e 5 com 2 elementos
- $$\sigma_T = \sqrt{\frac{10(10+1)(2 \times 10 + 1)}{24} - \frac{2^3 + 2^3 - (2 + 2)}{48}} = 9,8$$

Exemplo

- Assim

$$z_c = \frac{5 - 27,5}{9,8} = -2,229$$

- Temos $\text{valor} - p = 0,0264$.
- Temos que como $\text{valor} - p \leq \alpha$, assim rejeita-se H_0 ao nível de 5% de significância.
- Conclusão: Portanto, existe diferença entre os dois grupos, assim o uso da ração contribui para o aumento do peso médio dos animais.

TESTE DE WILCOXON - AMOSTRAS DEPENDENTES

Utilizando o R

- Para fazer o Teste t para duas médias pareadas
 - Estatísticas $\Rightarrow$ Teste Não Paramétricos $\Rightarrow$ Teste de Wilcoxon (amostras pareadas)
 - Em seguida escolher a variável na primeira amostra e segunda amostra
 - Também deve-se informar:
 - Se o teste é unilateral ou bilateral
 - Nível de confiança do teste