Teste T para amostras pareadas

Bianca Duarte
Ingrid Lorrane
Taciana Zerger

Amostras pareadas

- Cada observação da primeira amostra é pareada com uma observação da segunda amostra
- Ex
  - Cada indivíduo de um grupo tem duas respostas – uma antes e outra depois de uma intervenção.
    - Mesmo indivíduo, antes e depois de um tratamento
  - Cada indivíduo de um grupo recebe um par de outro grupo, que seja parecido com ele em relação às variáveis de interesse.
    - Gêmeos monozigóticos

Teste T para amostras pareadas

- A distribuição da variável desejada é gaussiana

- A intervenção feita entre as duas medidas interfere apenas no aumento ou no decréscimo da média da distribuição, e não na variabilidade das medidas

- Amostras pequenas
  - Com graus de liberdade muito grandes, a distribuição t de student se assemelha à distribuição de gauss

Teste T para amostras pareadas

- $H_0 : \mu_A = \mu_D$
  - Média antes do tratamento = média depois do tratamento
- $H_1 : \mu_A \neq \mu_D$
  - Média antes do tratamento ≠ média depois do tratamento

- Sendo as amostras dependentes, vamos verificar se a média das diferenças entre as amostras se afasta muito de zero, ou seja, verificar onde essa média das diferenças está em uma distribuição com média zero.

Teste T para amostras pareadas

- Considerando duas amostras pareadas
  - $x_1,\ldots,x_n$ e $y_1,\ldots,y_n$
- Temos n pares de observações
  - $(x_1, y_1), (x_2, y_2),\ldots,(x_N,y_N)$
- Diferença entre os valores de cada par
  - $D_1 = x_1 - y_1,\ldots, D_n = x_n - y_n$

Teste T para amostras pareadas

- Amostra das diferenças de cada par
  - d_1, d_2,...,d_n

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

- Média das diferenças

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

- Desvio padrão das diferenças

Teste T para amostras pareadas

- A distância entre a média das diferenças e zero, medida em desvios padrão – score padronizado

$$T_p = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- Rejeitar H_0 se a distância entre a média das diferenças e zero, expressa por T_p, for grande.

Teste T para amostras pareadas

- Assim, rejeitamos H_0 se

$$|T_p| \geq t_{n-1,1-\alpha/2}$$

  - Rejeitar H_0 – Há diferença significativa antes e depois do tratamento
  - Aceitar H_0 – Não há diferença significativa antes e depois do tratamento

Exemplo 1

- Estudo com o objetivo de avaliar a efetividade de uma dieta combinada com um programa de exercícios físicos na redução do nível de colesterol
- Sendo μ_A e μ_D as médias dos níveis de colesterol antes e depois do programa
- Vamos testar as seguintes hipóteses:
  - H_0: $\mu_A = \mu_D$, ou seja, o programa não tem efeito sobre os níveis de colesterol
  - H_1: $\mu_A \neq \mu_D$, ou seja, o programa tem efeito sobre os níveis de colesterol

Exemplo 1- roteiro para resolução

- Sabendo que $n_A = n_B = 12$
  - Calcule as **médias** de cada grupo(antes e depois)
  - Calcule as **diferenças** entre os grupos
  - Calcule a **média das diferenças**
  - Calcule a **diferença subtraída da média da diferença** (facilita o cálculo da variância)
  - Calcule as **variâncias das diferenças**
  - Calcule o **somatório das variâncias**

Exemplo 1

Antes (x_1)	Depois (x_2)
201	*200*
231	*236*
221	*216*
260	*233*
228	*224*
237	*216*
326	*296*
235	*195*
240	*207*
267	*247*
284	*210*
201	*209*

$\bar{x}_1 = 244{,}25$

$\bar{x}_2 = 224{,}08$

Exemplo 1

Antes (x_1)	Depois (x_2)	Diferenças ($d = x_1 - x_2$)
201	*200*	*1*
231	*236*	*-5*
221	*216*	*+5*
260	*233*	*+27*
228	*224*	*+4*
237	*216*	*+21*
326	*296*	*+30*
235	*195*	*+40*
240	*207*	*+33*
267	*247*	*+20*
284	*210*	*+74*
201	*209*	*-8*

$\bar{x}_1 = 244{,}25$ $\bar{x}_2 = 224{,}08$ $\bar{d} = 20{,}17$

Exemplo 1

Antes (x_1)	Depois (x_2)	Diferenças	Desvio $(d - \bar{d})$
201	*200*	*1*	*-19,16*
231	*236*	*-5*	*-25,16*
221	*216*	*+5*	*-15,16*
260	*233*	*+27*	*6,83*
228	*224*	*+4*	*-16,16*
237	*216*	*+21*	*0,83*
326	*296*	*+30*	*9,83*
235	*195*	*+40*	*19,83*
240	*207*	*+33*	*12,83*
267	*247*	*+20*	*-0,16*
284	*210*	*+74*	*53,83*
201	*209*	*-8*	*-28,16*

$\bar{x}_1 = 244{,}25$ $\bar{x}_2 = 224{,}08$ $\bar{d} = 20{,}17$

Exemplo 1

Antes (x_1)	Depois (x_2)	Diferenças ($d = x_1 - x_2$)	Desvio $(d - \bar{d})$	Variância $(d - \bar{d})^2$
201	200	1	-19,16	367,36
231	236	-5	-25,16	633,36
221	216	+5	-15,16	230,03
260	233	+27	6,83	46,69
228	224	+4	-16,16	261,36
237	216	+21	0,83	0,69
326	296	+30	9,83	96,69
235	195	+40	19,83	393,36
240	207	+33	12,83	164,69
267	247	+20	-0,16	0,03
284	210	+74	53,83	2898,03
201	209	-8	-28,16	793,36
$\bar{x}_1 = 244,25$	$\bar{x}_2 = 224,08$	$\bar{d} = 20,17$	------------	$\Sigma = 5885,7$

Exemplo 1

- Uma vez seguido roteiro e realizado todos os cálculos:
  - Descobrimos que houve redução nos níveis de colesterol depois do programa:
    - Médias antes= 244,25
    - Média depois=224,08
    - Diferença média = 20,17, no caso, representa uma redução … o padrão da redução é

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

$$s_d = \sqrt{\frac{5885{,}7}{12-1}} = 23{,}13$$

Exemplo 1

- Aplicando o teste T,temos:

$$T_p = \frac{\bar{d}}{s_d/\sqrt{n}}$$

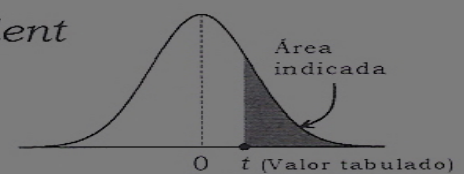
$$T_p = \frac{20{,}17}{\frac{23{,}13}{\sqrt{12}}}$$

$$\boxed{T_p = 3{,}02}$$

Exemplo 1

- Na tabela t de student, sabendo que o grau de liberdade é 11 (n-1)

Tabela 5 Distribuição *t* de *Student*



gl	Área na cauda superior								
	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015

Exemplo 1- Conclusão

Considerando que o teste é bicaudal, então:

$P = 0{,}012 \times 2$

$$P = 0{,}024$$

Com $p<0{,}05$, há evidência de que,em média, o programa altera o nível de colesterol

Exemplo 2

- A seguir encontram-se as médias obtidas pelos gêmeos de uma amostra de 20 pares de gêmeos monozigóticos submetidos a duas técnicas de ensino (N = nova e A = antiga).
  - Sorteou-se em cada par de MZ qual será submetido à técnica A e qual será submetido à B.
  - Ao final do experimento fez-se uma série de provas, cujas médias serão comparadas.
  - Pergunta-se: técnica nova é melhor que a antiga?

Exemplo 2

Gêmeos	Técnica N (x_1)	Técnica A (x_2)	$d = x_1 - x_2$	d^2
1	54	55	-1	1
2	57	54	3	9
3	60	57	3	9
4	70	60	10	100
5	55	50	5	25
6	62	62	0	0
7	53	48	5	25
8	70	61	9	81
9	62	55	7	49
10	56	55	1	1
11	61	52	9	81
12	65	70	-5	25
13	54	57	-3	9
14	59	51	8	64
15	60	58	2	4
16	68	61	7	49
17	65	60	5	25
18	70	66	4	16
19	63	54	9	81
20	58	56	2	4
Total	1222	1142	80	658

Exemplo 2

- É importante notar que a pergunta refere-se a *média melhor*. Obviamente refere-se a maior.
- H. Nula: as duas técnicas tem o mesmo efeito, ou seja, a média dos desvios não difere de zero

$$H_0 = \overline{\mu_d} = 0$$

- H. Alternativa: a técnica nova N é melhor que a antiga A, ou seja, a média dos desvios é maior que zero

$$H_a = \overline{\mu}_d > 0$$

Exemplo 2

- $\bar{d} = \dfrac{\sum d}{n} = 4$

- $s_d = \sqrt{\dfrac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} = 4{,}2177$

- $T19 = \dfrac{4}{\frac{4{,}22}{\sqrt{20}}} = 4{,}255319$ com **P < 0,001**

- Procura-se o valor do **t** crítico na fileira correspondente a n-1 graus de liberdade, e na coluna com o *dobro do grau de significância* desejado (10%). Nesse caso, **t19** = 1,729.

Exemplo 2 - Conclusão

- Como o valor de **t19** (4,25) é maior que **tc** (1,729) rejeita-se a hipótese nula, e os resultados obtidos após a técnica nova diferem dos antigos.

- A técnica nova é melhor que a antiga já que ocasiona médias maiores.