

ESTATÍSTICA PARA ANÁLISE DE DADOS COM PYTHON

Prof. Luciano Galdino

Testes de hipóteses com duas amostras

Amostras independentes: Quando uma amostra selecionada de uma população não tiver relação com uma amostra selecionada em outra população.

Amostras dependentes: Quando uma amostra selecionada de uma população estiver relacionada com uma amostra selecionada em outra população. Também são chamadas de pareadas ou relacionadas.

Testes de hipóteses com duas amostras

Amostras Independentes

Hipótese nula (H_0) = hipótese estatística que geralmente diz que não há diferença entre os parâmetros de duas populações.

Hipótese alternativa (H_a) = É verdadeira quando a hipótese nula é falsa.

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

$$H_0: \mu_1 \leq \mu_2$$

$$H_a: \mu_1 > \mu_2$$

$$H_0: \mu_1 \geq \mu_2$$

$$H_a: \mu_1 < \mu_2$$

Teste z de duas amostras para a diferença entre médias

Condições:

- 1) Amostras aleatórias.
- 2) Amostras Independentes.
- 3) Tamanho das amostras com no mínimo 30 em cada população com distribuição normal e desvio padrão conhecido.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Nível de confiança	Zc
90%	1,645
95%	1,96
99%	2,575

Sem conhecimento de μ_1 e μ_2 , presume-se que eles são iguais.

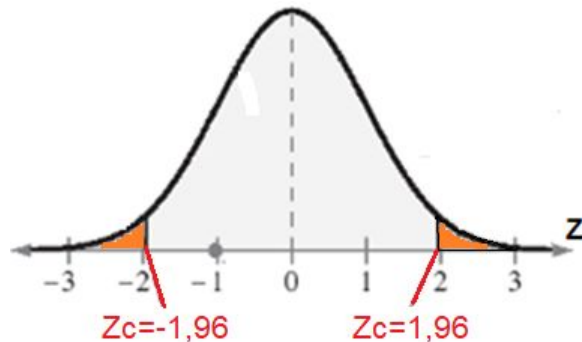
Exemplo 1: Um analista de crédito afirma que não existe diferença entre a média da dívida do cartão de crédito de homens para mulheres. Os resultados de uma pesquisa, com 400 pessoas (200 de cada grupo), apresentados por ele estão na tabela a seguir. Os resultados apoiam a sua afirmação? Considere nível de significância de 0,05.

PARÂMETROS	DÍVIDA MULHERES	DÍVIDA HOMENS
MÉDIA	R\$2290,00	R\$2370,00
DESVIO PADRÃO	R\$750,00	R\$800,00

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Nível de confiança	Zc
90%	1,645
95%	1,96
99%	2,575



$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{750^2}{200} + \frac{800^2}{200}}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = 77,54$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$z = \frac{(2290 - 2370) - 0}{77,54}$$

$$z = -1,03$$

Aceita H_0

Teste t duas amostras para diferença entre médias (Teste-t de Student)

Condições:

- 1) Amostras aleatórias.
- 2) Amostras Independentes.
- 3) Tamanho das amostras menor que 30 e distribuição normal.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

Variâncias iguais

$$g.l = n_1 + n_2 - 2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Variâncias diferentes

Menor entre:

$$g.l = n_1 - 1 \quad g.l = n_2 - 1$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Nível de confiança, c		0,50	0,80	0,90	0,95	0,98	0,99
Unicaudal, α		0,25	0,10	0,05	0,025	0,01	0,005
g.l.	Bicaudal, α	0,50	0,20	0,10	0,05	0,02	0,01
1		1,000	3,078	6,314	12,706	31,821	63,657
2		0,816	1,886	2,920	4,303	6,965	9,925
3		0,765	1,638	2,353	3,182	4,541	5,841
4		0,741	1,533	2,132	2,776	3,747	4,604
5		0,727	1,476	2,015	2,571	3,365	4,032
6		0,718	1,440	1,943	2,447	3,143	3,707
7		0,711	1,415	1,895	2,365	2,998	3,499
8		0,706	1,397	1,860	2,306	2,896	3,355
9		0,703	1,383	1,833	2,262	2,821	3,250
10		0,700	1,372	1,812	2,228	2,764	3,169
11		0,697	1,363	1,796	2,201	2,718	3,106
12		0,695	1,356	1,782	2,179	2,681	3,055
13		0,694	1,350	1,771	2,160	2,650	3,012
14		0,692	1,345	1,761	2,145	2,624	2,977
15		0,691	1,341	1,753	2,131	2,602	2,947
16		0,690	1,337	1,746	2,120	2,583	2,921
17		0,689	1,333	1,740	2,110	2,567	2,898
18		0,688	1,330	1,734	2,101	2,552	2,878
19		0,688	1,328	1,729	2,093	2,539	2,861
20		0,687	1,325	1,725	2,086	2,528	2,845
21		0,686	1,323	1,721	2,080	2,518	2,831
22		0,686	1,321	1,717	2,074	2,508	2,819
23		0,685	1,319	1,714	2,069	2,500	2,807
24		0,685	1,318	1,711	2,064	2,492	2,797
25		0,684	1,316	1,708	2,060	2,485	2,787

Escolha do teste adequado

TIPO DE AMOSTRA	TESTE PARAMÉTRICO (Normalidade e média como referência)	TESTE NÃO PARAMÉTRICO (mediana como referência)
UMA AMOSTRA	Teste Z (Desvio padrão populacional conhecido ou $n \geq 30$). Teste t (Desvio padrão populacional desconhecido e $n < 30$).	
DUAS AMOSTRAS DEPENDENTES		
DUAS AMOSTRAS INDEPENDENTES	Teste Z ($n \geq 30$ e desvio padrão conhecido). Teste t de Student ($n < 30$).	
TRÊS OU MAIS AMOSTRAS		
AMOSTRAS CORRELACIONADAS		