Testes de hipóteses para a média de duas populações

terça-feira, 27 de junho de 2023 13:52

Nessa aula consideraremos apenas os casos em que as variâncias populacionais são desconhecidas --> t-Student

Ainda assim, temos dois casos a tratar: **Dados pareados e Dados não-pareados**

DADOS PAREADOS

Entende-se por dados pareados os dados de duas amostras dependentes ente si, ou seja, que são medidos segundo algum critério que introduz uma influência marcante entre os pares de valores (antes e depois de algum tratamento).

Exemplo: Vamos submeter 6 animais de uma espécie a uma nova dieta e avaliar o efeito desta (dieta) no ganho de peso dos animais.

Peso Animais antes da dieta
Peso Animais depois da dieta

Neste caso, as observações deverão ser feitas nos mesmos animais, medindo uma característica antes e depois de serem submetidos ao tratamento.

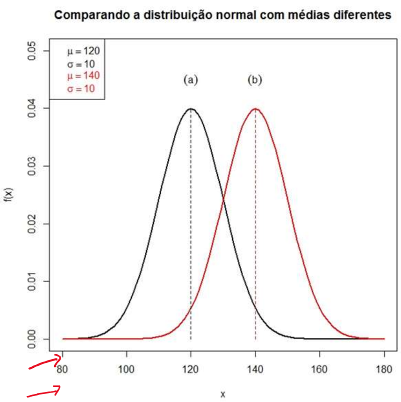
DADOS NÃO PAREADOS

Neste caso, os dados das duas amostras não estão relacionados entre si, são indivíduos diferentes.

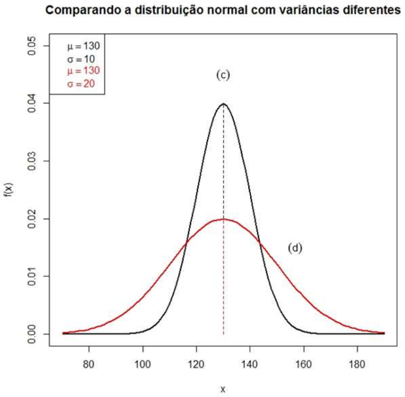
Exemplos: Produtividade de um mesmo tipo de cultura em duas áreas diferentes; tempo de execução de serviço de duas máquinas diferentes; Peso de animais (diferentes) submetidos a dois tipos de dieta; etc.

Para o caso de **amostras independentes (dados não pareados)** podemos ainda distinguir duas situações, quais sejam:

- Variâncias homogêneas: quando as variâncias, mesmo que desconhecidas, podem ser consideradas iguais (com base nos dados amostrais e confirmada por um teste);
- Variâncias heterogêneas: quando as variâncias, mesmo que desconhecidas, podem ser consideradas diferentes (dados amostras e teste).



Variâncias homogêneas



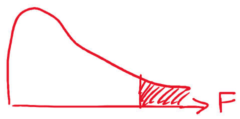
Variâncias heterogêneas

TESTE DE HIPÓTESE (BASEADO NA DISTRIBUIÇÃO F) PARA HOMOGENEIDADE DE VARIÂNCIAS (DUAS POPULAÇÕES).

Como em qualquer outro teste, começaremos pela **formulação das hipóteses: nula e alternativa**

$H0$: $\sigma_1^2 = \sigma_2^2$ → $Ho: \frac{\sigma_1^2}{\sigma_2^2} = 1$

Ha: $\sigma_1^2 > \sigma_2^2$ → $Ha: \frac{\sigma_1^2}{\sigma_2^2} > 1$



em que: σ_1^2 e σ_2^2 são as variâncias das populações 1 e 2, respectivamente.

Na hipótese nula ($H0$) estamos afirmando que a variância das duas populações são iguais (homogêneas) e na hipótese alternativa (Ha) está contempla a situação em que a variância da população 1 é maior do que a variância da população 2 (variâncias heterogêneas). Portanto, vamos sempre realizar um teste unilateral à direita (dado o que está posto na hipótese alternativa).

Quando desejamos comparar as variâncias populacionais devemos utilizar a estatística Fc, dada pelo quociente (divisão) entre as duas estimativas das variâncias populacionais, ou seja,

$$Fc = \frac{s_1^2}{s_2^2}$$

Amostra 1: tamanho n_1; média amostral $\bar{X}_1$; variância amostral s_1^2.
Amostra 2: tamanho n_2; média amostral $\bar{X}_2$; variância amostral s_2^2.

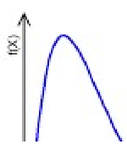
Com os dados das amostras vamos calcular a Estatística Fc e comparar seu valor com o quantil tabelado da Distribuição F, considerando ν_1 e ν_2 graus de liberdade associados ao numerador e denominador, respectivamente.

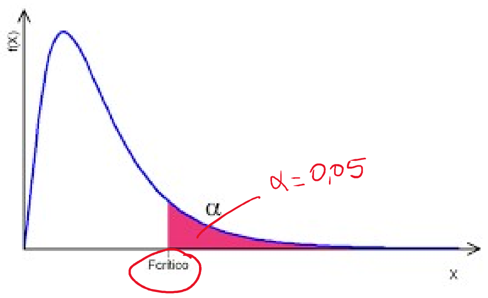
$Fc = \frac{s_1^2}{s_2^2}$ — s_1^2 → Amostra → $\nu_1 = n_1 - 1$; s_2^2 → Amostra → $\nu_2 = n_2 - 1$

Numerador de Fc: $\nu_1 = n_1 - 1$ Sempre
usaremos a tabela considerando $\alpha = 0{,}05$
Denominador de Fc: $\nu_2 = n_2 - 1$

ν_1

ν_2 → $F_{(\nu_1, \nu_2, 0{,}05)}$





$v_2 \longrightarrow F_{(v_1, v_2, 0,05)}$

Exemplo: Com o objetivo de comparar as produções médias, em toneladas/hectare, de duas variedades de milho (A e B), foram observadas 5 unidades experimentais para cada variedade e os resultados obtidos foram os seguintes.

Variedade A	1,3	1,4	1,1	1,4	1,5
Variedade B	1,8	1,6	1,9	1,9	1,8

$\mu_A = \mu_B$ antes $\rightarrow \dfrac{S_A^2}{S_B^2}$

Dos dados amostrais obtemos,

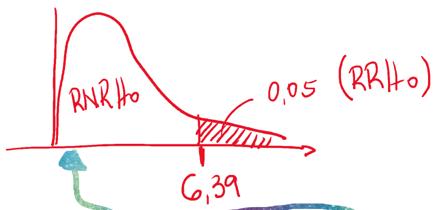
Variedade A: tamanho $n_A = 5$; média amostral $\bar{X}_A = 1{,}34$; variância amostral $s_A^2 = 0{,}0231$.
Variedade B: tamanho $n_B = 5$; média amostral $\bar{X}_B = 1{,}80$; variância amostral $s_B^2 = 0{,}0150$.

Vamos executar o teste F para verificar se as variâncias podem ser consideradas homogêneas ou não ($\alpha = 0{,}05$)...e só depois faremos um teste para comparação das médias (em função do resultado obtido)

1) Hipóteses

$Ho: \dfrac{\sigma_A^2}{\sigma_B^2} = 1$ (variâncias Homog.) vs $Ha: \dfrac{\sigma_A^2}{\sigma_B^2} > 1$

2) Distribuição de Referência (Distribuição F)



$\nu_1 = 4$

$\nu_2 = 4 \rightarrow 6{,}32$

3) Estatística do Teste (Fc)

$F_c = \dfrac{s_A^2}{s_B^2} = \dfrac{0{,}0231}{0{,}0150} = 1{,}54$

4) Conclusão

Como $F_c = 1{,}54$ está na região de não rejeição da hipótese Ho, concluímos que as variâncias populacionais são homogêneas ($\alpha = 0{,}05$)

Como o resultado do teste F indicou que as variâncias populacionais são HOMOGÊNEAS, vamos agora proceder o TESTE DE HIPÓTESE PARA COMPARAÇÃO DAS MÉDIAS POPULACIONAIS (CONSIDERANDO VARIÂNCIAS HOMOGÊNEAS).

O **primeiro passo**, como sempre, é formular as hipóteses do teste (agora em relação às médias)

$H0$: $\mu_A = \mu_B$
Ha: $\mu_A \neq \mu_B$ (teste de hipótese bilateral)

O que implica em hipóteses do tipo:

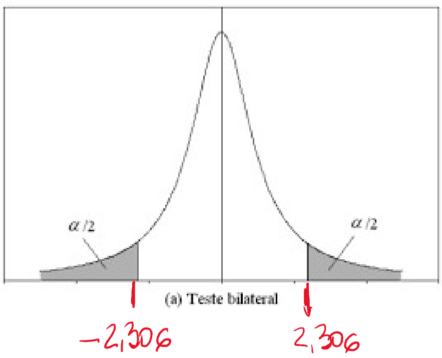
$H0$: $\mu_A - \mu_B = 0$
Ha: $\mu_A - \mu_B \neq 0$

Como **segundo passo**-> faremos o teste considerando um nível de significância α (se não especificado, utilizaremos 5%)

No **terceiro passo,** faremos o esquema da Distribuição de Referência (no início deste material mencionamos que iremos considerar sempre as variâncias populacionais desconhecidas) que é a t-Student.
Lembre-se: Concluímos, para nosso exemplo de milho, que as variâncias populacionais são homogêneas

Quando as variâncias populacionais são consideradas homogêneas precisamos compor um novo grau de liberdade ν
para consultar a tabela da Distribuição t-Student (linhas da tabela) da seguinte forma:

$\nu = n_1 + n_2 - 2$ $\Rightarrow (n_1 - 1) + (n_2 - 1) = n_1 - 1 + n_2 - 1 = n_1 + n_2 - 2$



(a) Teste bilateral

t-Stu. α 0,025
$\nu^* = 8 \rightarrow 2{,}306$

Variedade A: tamanho $n_A = 5$; média amostral $\bar{X}_A = 1{,}34$; variância amostral $s_A^2 = 0{,}0231$.
Variedade B: tamanho $n_B = 5$; média amostral $\bar{X}_B = 1{,}80$; variância amostral $s_B^2 = 0{,}0150$.

No **quarto passo** calculamos a Estatística do teste (Tc). Também neste passo precisamos considerar que as variâncias foram consideradas homogêneas (no teste F) para usar a estatística adequada...que é!

$Ho: \mu_A - \mu_B = 0$

$$Tc = \frac{(\bar{X}_A - \bar{X}_B) - \cancel{(\mu_A - \mu_B)}^{\,0}}{S_{comb}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

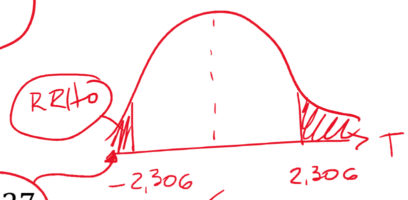
Em que: o desvio-padrão combinado (S_{comb}) é calculado por:

$$S_{comb} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}$$

Amostras:
Variedade A: tamanho $n_A = 5$; média amostral $\bar{X}_A = 1{,}34$; variância amostral $s_A^2 = 0{,}0231$.
Variedade B: tamanho $n_B = 5$; média amostral $\bar{X}_B = 1{,}80$; variância amostral $s_B^2 = 0{,}0150$

$$S_{comb} = \sqrt{\frac{(5-1) \times 0{,}0231 + (5-1) \times 0{,}0150}{5+5-2}} = 0{,}1380$$

Logo, a Estatística Tc é dada por:



$$Tc = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_{comb} \times \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{(1{,}34 - 1{,}80) - (0)}{0{,}1380\sqrt{\frac{1}{5} + \frac{1}{5}}} = -5{,}27$$

$H_0: \cancel{\mu_A = \mu_B}$

$H_a: \mu_A \neq \mu_B$

No **quinto passo** concluímos sobre o teste:
Como o valor de $Tc = -5{,}27$ é maior (em módulo) do que o valor tabelado ($t_{(0{,}025;8)} = 2{,}306$), concluímos pela rejeição da hipótese $H0$, ou seja, as produções média das duas variedade de milho são estatisticamente diferentes entre si (para um nível de significância de 5%). Portanto, a variedade B é mais produtiva que a variedade A

Exemplo 2. Foi realizado um experimento com o objetivo de comparar os tempos gastos na manobra com os arados Fuçador e Erechim. Ambos os arados são de tração animal. Os dados obtidos com 11 repetições para cada arado, foram os seguintes:

Fuçador	0,20	0,22	0,18	0,23	0,12	0,20	0,13	0,12	0,13	0,22	0,17
Erechim	0,36	0,48	0,33	0,43	0,40	0,43	0,33	0,36	0,35	0,40	0,35

Espera-se que o arado Fuçador produza melhores resultados (gaste menos tempo médio na manobra). Testar a hipótese de que a média de tempo gasto com o arado Fuçador (μ_F) é menor do que a média do tempo gasto com o arado Erechim (μ_E), ao nível de 5%. Vamos

1) Faça um teste para verificar se as variâncias são, ou não, homogêneas;
2) Faça um teste para comparação das médias populacionais.

TESTE PARA HOMOGENEIDADE DE VARIÂNCIAS

AMOSTRAS:
FUÇADOR: $n = 11$, $\bar{x}_F = 0{,}17$, $s_F^2 = 0{,}0018$
ERECHIM: $n = 11$, $\bar{x}_E = 0{,}38$, $s_E^2 = 0{,}0023$ ← maior

HIPÓTESES: $H_0: \frac{\sigma_E^2}{\sigma_F^2} = 1$ VERSUS $H_a: \frac{\sigma_E^2}{\sigma_F^2} > 1$

$\alpha = 0{,}05$

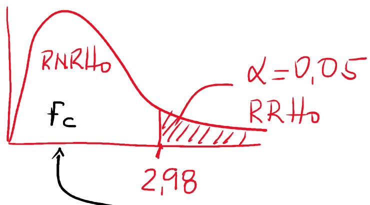


TABELA F

$\nu_1 = 11 - 1 = 10$

$\nu_2 = 10 \rightarrow 2{,}98$

ESTATÍSTICA DO TESTE (F_c)

$$F_c = \frac{s_E^2}{s_F^2} = \frac{0{,}0023}{0{,}0018} = 1{,}27$$

Como $F_c = 1{,}27$ está na RNRHo concluímos que as variâncias são homogêneas.

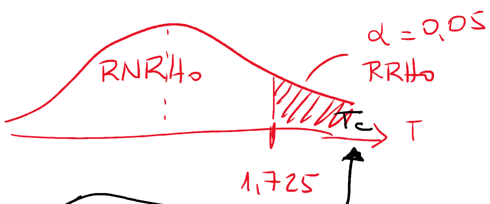
CONSIDERANDO QUE AS VARIÂNCIAS SÃO HOMOGÊNEAS

TESTE HIPÓTESE PARA COMPARAÇÃO DAS MÉDIAS DAS 2 POPULAÇÕES

$H_0: \mu_E = \mu_F \Rightarrow \mu_E - \mu_F = 0$

$H_a: \mu_E > \mu_F \Rightarrow \mu_E - \mu_F > 0$ (ENUNCIADO $\Rightarrow$ Fuçador melhor que o Erechim, portanto menor tempo de manobra)

$\alpha = 0{,}05$



$\nu = n_1 + n_2 - 2 = 11 + 11 - 2 = 20$

t-student 0,05

$20 \rightarrow 1{,}725$

ESTATÍSTICA DO TESTE (T_c)

$$T_c = \frac{(\bar{X}_E - \bar{X}_F) - (\mu_E - \mu_F)}{S_c\sqrt{\frac{1}{n_E} + \frac{1}{n_F}}}$$

na Ho!

, em que: $S_c = \sqrt{\frac{(n_E - 1)s_E^2 + (n_F - 1)s_F^2}{n_E + n_F - 2}}$

$$S_c = \sqrt{\frac{(10 \times 0{,}0023) + (10 \times 0{,}0018)}{11 + 11 - 2}} = 0{,}0453$$

$$T_c = \frac{(0{,}38 - 0{,}17) - (0)}{0{,}0453\sqrt{\frac{1}{11} + \frac{1}{11}}} = 10{,}87$$

Como Tc está na região de rejeição da Ho, concluímos (com um nível de significância de 5%), que o arado Fuçador tem um tempo de manobra inferior ao arado Erechim.

tem um tempo de [illegible]