4- Tópicos Especiais - Capítulo 9
Magalhães e Lima

Comparação de Duas Médias - Introdução

Estamos interessados em comparar duas populações com relação às suas médias.

É muito comum a situação em que as observações são tomadas em uma mesma unidade amostral antes e depois de alguma intervenção. Neste caso, é razoável supor que a tomada de medidas em um mesmo elemento gere dependência entre as duas observações.

Os dois exemplos a seguir apresentam um caso com independência e outro com dependência entre as medidas

Ex 9.1 Magalhães e Lima pág 310

Uma distribuidora de combustíveis deseja verificar se um novo tipo de gasolina é eficaz na revitalização de motores velhos. Para isso, seleciona 12 automóveis de um mesmo modelo com mais de 8 anos de uso e após a regulagem de seus motores, verifica o consumo de combustível. Em seguida, o carro é abastecido com o novo combustível durante 15 semanas e uma nova medida do consumo é feita. Defina as variáveis aleatórias X_i e Y_i como o rendimento do automóvel i respectivamente antes e após as 15 semanas.

X_i e Y_i foram medidas em uma mesma unidade amostral

$\Rightarrow$ razoável supor dependência entre elas

Exemplo 9.2

Um estudo envolve a avaliação de um
novo sistema operacional de computador,
desenvolvido para crianças com idade entre
8 e 12 anos. Afirma-se que o novo sistema
é mais rápido que o atual, líder do mercado.
Para testar esta afirmação, foram seleciona-
dos em uma mesma escola dois grupos com
15 crianças cada. As crianças, sem conheci-
mento prévio relacionado ao uso de computa-
dores, utilizaram máquinas de mesma
configuração para realizar uma certa ta-
refa, que teve seu tempo anotado.
O 1º grupo trabalhou com o sistema operacio-
nal convencional e o 2º desenvolveu ativi-
dades no novo sistema. Ao final do expe-
rimento todas as 30 crianças haviam rea-
lizado a tarefa.

Grupos constituídos de 15 crianças diferentes
$\Rightarrow$ assume-se que os dois grupos são amostras
independentes

Em ambos os casos temos duas variáveis aleatórias X e Y e deseja-se testar

$H_0: \mu_X = \mu_Y \qquad \mu_X = E(X) \qquad \mu_Y = E(Y)$

1º Caso: Amostras dependentes (ou dados emparelhados)

2º Caso: Amostras independentes

---

a) Comparação de Médias com base em amostras dependentes (teste t-pareado)

Este caso surge quando as observações são tomadas na mesma unidade amostral antes e depois de algum tratamento, ou são observações associadas a situações em que houve emparelhamento prévio.

Dispomos de uma amostra de n observações pareadas: $(X_1, Y_1)\ (X_2, Y_2) \ldots (X_n, Y_n)$

X_i - medida da variável de interesse no i-ésimo elemento amostral, antes da aplicação do tratamento (ou na condição experimental A)

Y_i - " " " , após a aplicação do tratamento (ou na condição experimental B)

$i = 1, 2 \ldots n$

$X \sim N(\mu_X, \sigma_X^2)$ $\quad$ $Y \sim N(\mu_Y, \sigma_Y^2)$

Deseja-se testar:

$H_0 : \mu_X = \mu_Y$

Constrói-se a variável $D = Y - X$, de modo que

$D_i = Y_i - X_i$, $i = 1, 2 \ldots n$

$D_i \sim N(\mu_D, \sigma_D^2)$ com $\mu_D = E(Y_i - X_i) = \mu_Y - \mu_X$

$H_0 : \mu_X = \mu_Y \iff H_0 : \mu_D = 0$

∴ Teste bicaudal

$H_0: \mu_D = 0$
$H_a: \mu_D \neq 0$
} Teste para a média da distribuição Normal com variância desconhecida

Calcula-se

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} \quad \text{e} \quad S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$$

media amostral da variável D

variância amostral

Estatística de Teste:

$$T = \frac{\bar{D}}{S_D/\sqrt{n}}$$

Sob H_0, $T \sim t_{n-1}$

RC: $T \leq -t_c$ ou $T \geq t_c$

$P(T \geq t_c) = \alpha/2 \qquad T \sim t_{n-1}$

Teste unicaudal à direita

$H_0: \mu_X = \mu_Y \iff \mu_D = 0$ $\qquad D = Y - X$

$H_a: \mu_Y > \mu_X \iff \mu_D > 0$

$RC: T \geqslant t_c \qquad P(T \geqslant t_c) = \alpha \,, \; T \sim t_{n-1}$

Exemplo 9.1

Amostra de 12 carros com mais de 8 anos

Anotou-se o consumo (em Km/l) com o combustível usual e posteriormente com o novo combustível

Testar ao nível de significância de 0,02 a hipótese de igualdade dos consumos médios contra a alternativa que o consumo médio com o novo combustível é superior.

X_i – consumo de combustível usando o combustível padrão, i-ésimo carro

Y_i – " combustível novo

$i = 1, 2, \dots 12$

automóvel	Consumo Com. Novo (Y)	Consumo comb. Padrão (X)	$D = Y - X$
1	11,6	8,1	3,5
2	8,8	7,9	0,9
3	9,9	6,8	3,1
4	9,5	7,8	1,7
5	11,6	7,6	4,0
6	9,1	7,9	1,2
7	10,6	5,7	4,9
8	10,8	8,4	2,4
9	13,4	8,0	5,4
10	10,6	9,5	1,1
11	10,5	8,0	2,5
12	11,4	6,8	4,6
			35,3

$H_0: \mu_X = \mu_Y \quad (\mu_D = 0)$

$H_a: \mu_Y > \mu_X \quad (\mu_D > 0)$

Construir Y de modo que a hipótese de interesse seja $\mu_Y > \mu_X$.

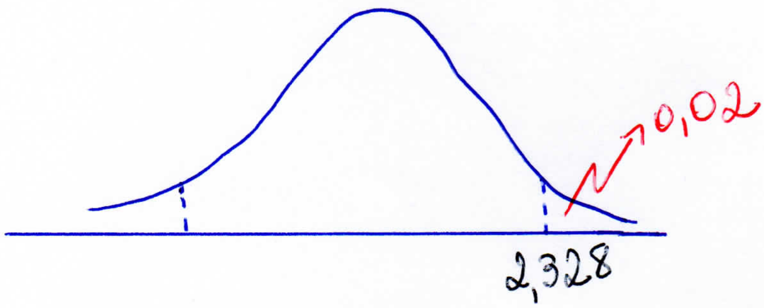
$$\bar{D} = \frac{35,3}{12} = 2,94$$

Após cálculos $S_D^2 = 2,4$

$$T = \frac{\bar{D} - 0}{S_D/\sqrt{n}} = \frac{2,94}{1,55/\sqrt{12}} = 6,48$$

$\alpha = 0{,}02$ $\quad$ 11 g l

Dist t_{11}



Procurar na tabela a soma das caudas

4%

11 2,328

RC : $T \geq 2{,}328$ $\quad$ $T_{obs} = 6{,}48 \in RC$

Rejeita-se H_0, os dados sugerem que o consumo médio com o novo combustível é maior que o combustível usual

Exemplo

Em um determinado colégio, o diretor deseja testar se existe diferença entre as notas médias de Português e Matemática dos alunos do 1º ano do ensino médio. Para uma amostra de 10 alunos do 1º ano do ensino médio observa-se

aluno	1	2	3	4	5	6	7	8	9	10
Nota P	10	2	4	9	7	3	1	5	6	7
nota M	4	3	10	5	9	6	7	2	8	1
$D=Y-X$	6	-1	-6	4	-2	-3	-6	3	-2	6

Y- Nota de Português $Y \sim N(\mu_Y, \sigma_Y^2)$

X- Nota de Matemática $X \sim N(\mu_X, \sigma_X^2)$

$H_0: \mu_X = \mu_Y \iff \mu_D = 0$

$H_a: \mu_X \neq \mu_Y \iff \mu_D \neq 0$

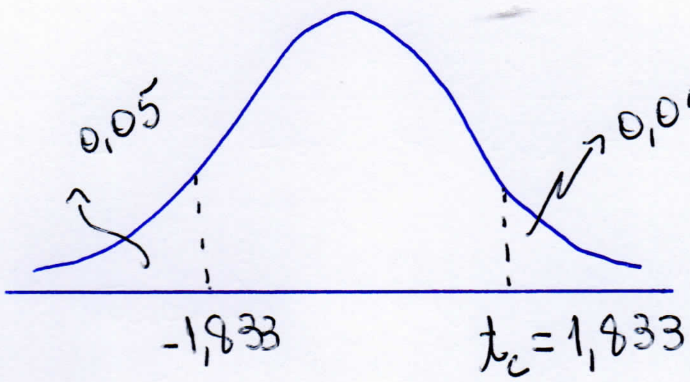
Cálculos

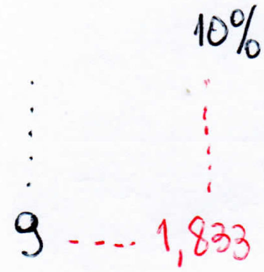
$$\sum_{i=1}^{10} D_i = -1 \qquad \bar{D} = -0{,}1 \qquad \sum_{i=1}^{10} (D_i - \bar{D})^2 = 186{,}9$$

$$n = 10 \qquad S_D^2 = \frac{186{,}9}{9} = 20{,}76$$

Fixando $\alpha = 0{,}10$

RC: $T \geq t_c$ ou $T \leq -t_c$ $\qquad T \sim t_9$





RC: $T \geq 1{,}833$ ou $T \leq -1{,}833$

$$T = \frac{-0{,}1}{\frac{4{,}5}{\sqrt{10}}} = -0{,}07 \notin RC$$

Não rejeitamos H_0. Os dados sugerem que não existe diferença entre as notas médias de Português e Matemática para os alunos do 1º ano do ensino médio desse colégio.

b) Comparação de Médias com base em amostras independentes

Neste caso, dispomos de duas amostras independentes (unidades amostrais distintas e não relacionadas)

$X_1, X_2 \dots X_{n_1}$ amostra de tamanho n_1 da pop 1

$X \sim N(\mu_X, \sigma_X^2)$

$Y_1, Y_2 \dots Y_{n_2}$ amostra de tamanho n_2 da pop 2

$Y \sim N(\mu_Y, \sigma_Y^2)$

Teste de Hipóteses Bicaudal

$H_0: \mu_X = \mu_Y$

$H_a: \mu_X \neq \mu_Y$

$$\left.\begin{array}{l} \bar{X} \sim N\left(\mu_X, \dfrac{\sigma_X^2}{n_1}\right) \\[2ex] \bar{Y} \sim N\left(\mu_Y, \dfrac{\sigma_Y^2}{n_2}\right) \end{array}\right\} \text{independentes}$$

$$E(\bar{Y}-\bar{X}) = \mu_X - \mu_Y$$

$$Var(\bar{Y}-\bar{X}) \underset{\text{indep.}}{=} Var(\bar{Y}) + Var(\bar{X}) = \frac{\sigma_Y^2}{n_2} + \frac{\sigma_X^2}{n_1}$$

$$\Rightarrow \bar{Y}-\bar{X} \sim N\left(\mu_Y - \mu_X, \frac{\sigma_Y^2}{n_2} + \frac{\sigma_X^2}{n_1}\right)$$

Sob H_0, $\mu_X = \mu_Y$ e se σ_X^2 e σ_Y^2 são conhecidos

$$Z = \frac{\bar{Y}-\bar{X}}{\sqrt{\dfrac{\sigma_Y^2}{n_2} + \dfrac{\sigma_X^2}{n_1}}} \sim N(0,1)$$

Rejeita-se H_0 se $Z \geq z_c$ ou $Z \leq -z_c$

z_c tal que $P(Z \geq z_c) = \dfrac{\alpha}{2}$ $\quad Z \sim N(0,1)$

Teste Unicaudal

$H_0: \mu_X = \mu_Y$

$H_a: \mu_Y > \mu_X$

RC: $Z \geq z_c$ $\quad P(Z \geq z_c) = \alpha \quad Z \sim N(0,1)$

Se σ_X^2 e σ_Y^2 são desconhecidos, sob $H_0: \mu_X = \mu_Y$, utiliza-se a estatística de teste:

$$T = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{S_Y^2}{n_2} + \frac{S_X^2}{n_1}}}$$

S_X^2 e S_Y^2 variâncias amostrais associadas às amostras das v. X e Y respectivamente

Sob H_0, $T \sim t_K$ onde

$$K = \frac{\left(\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}\right)^2}{\frac{\left(\frac{S_X^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_Y^2}{n_2}\right)^2}{n_2 - 1}}$$

As regiões críticas e a execução dos testes são similares ao caso anterior.