Estatística

12 – Comparação entre várias populações

1) Comparação de Médias (Análise de Variância)

Página da FEG: www.feg.unesp.br/~marcela

Comparação de Médias de Várias Populações

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \\ H_1 \ : \text{PELO MENOS uma das médias é diferente} \end{cases}$$

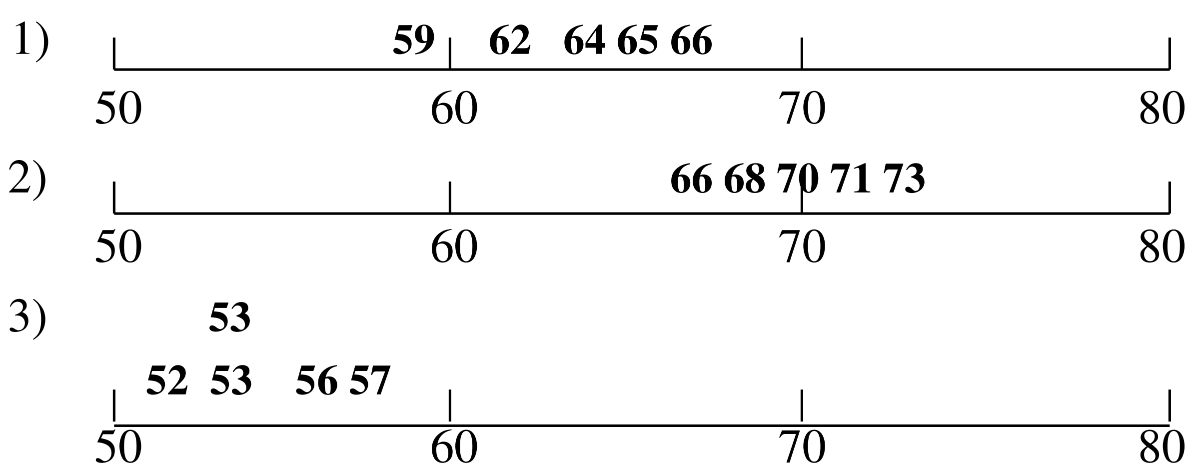
Amostras:

1) 64 66 59 65 62 $\quad \boxed{\overline{x}_1 = 63{,}2}$

2) 71 73 66 70 68 $\quad \boxed{\overline{x}_2 = 69{,}6}$

3) 52 57 53 56 53 $\quad \boxed{\overline{x}_3 = 54{,}2}$

Assim:



A partir de uma análise "visual" é razoável supor que H_0 deve ser **REJEITADA.**

Análise de Variância - ANOVA

Tabela 1: Medida da resistência de três tipos de concretos

						Média
Concreto A	69	74	77	70	71	72,2
Concreto B	69	65	69	66	68	67,4
Concreto C	74	77	76	80	69	75,2

Tabela 2: Valores da Tabela (1) subtraídos de 71

						Soma
Concreto A	-2	3	6	-1	0	6
Concreto B	-2	-6	-2	-5	-3	-18
Concreto C	3	6	5	9	-2	21
					Total	**9**

Tabela 3: Quadro de análise de variância

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	$F_{2,12,1\%}$
Fator	**154,8****	2	77,4	7,56	6,93
Erro	122,8	12	10,233		
Total	**277,6***	14			

* $SQ_{total} = (4+9+36+1+\ldots+4) - 81/15$

** $SQ_{fator} = (36 + 324 + 441) / 5 - 81 / 15$

Análise de Variância - ANOVA

Hipótese Ho	O que fazer?
Rejeitada	Calcular o **ZETA** para verificar qual $\mu_i \neq \mu_j$

Método de Tukey:
os concretos que apresentarem uma diferença da resistência média nas amostras superior a Zeta, isto é,

$$\left|\overline{X}_i - \overline{X}_j\right| > \text{Zeta, para i, j} \in \{A, B, C\} \text{ e } i \neq j$$

serão considerados como tendo diferentes resistências médias.

Tabela 1: Medida da resistência de três tipos de concretos

						Média
Concreto A	69	74	77	70	71	72,2
Concreto B	69	65	69	66	68	67,4
Concreto C	74	77	76	80	69	75,2

Tabela 3: Quadro de análise de variância

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	$F_{2,12,1\%}$
Fator	**154,8****	2	77,4	7,56	6,93
Erro	122,8	**12**	**10,233**		
Total	**277,6***	14			

$$\text{Zeta} = q_{k,\nu,\alpha}\sqrt{\frac{S_R^2}{n}} = q_{3,12,1\%}\sqrt{\frac{S_R^2}{n}} =$$

$$= 5{,}04\sqrt{\frac{10{,}233}{5}} = 7{,}21$$

k=número de níveis=3;

n=tamanho da amostra=5,

$\nu = \mathbf{12}$

$q_{k,\nu,\alpha}$ é a Amplitude Studantizada.

Análise de Variância - ANOVA

$$\text{Zeta} = 7{,}21$$

$$\left|\overline{X}_A - \overline{X}_B\right| = |72{,}2 - 67{,}4| = 4{,}8$$

$$\left|\overline{X}_A - \overline{X}_C\right| = |72{,}2 - 75{,}2| = 3$$

$$\left|\overline{X}_B - \overline{X}_C\right| = |67{,}4 - 75{,}2| = 7{,}8$$

Conclusão:
Só podemos concluir que o concreto C é mais resistente que o concreto B.

Comparação de Médias de Várias Populações

2º Experimento:

Experimento com dois fatores, sem repetição

Tabela 4: Experimento - dois fatores, sem repetição

Fertilizante	A	B	C	D	E	F
Variedade 1	5,4	3,2	3,8	4,6	5,0	4,4
Variedade 2	5,7	4,0	4,2	4,5	5,3	5,0

Temos dois conjuntos de hipóteses:

1) Quanto ao fertilizante

$$H_0 : \mu_A = \mu_B = \mu_C = \ldots = \mu_F$$

$$H_1 : \text{Ao menos um } \mu_i \neq \mu_j \text{, para i, j} \in \{A, B, C, \ldots, F\} \text{ e i} \neq \text{j}$$

2) Quanto a variedade

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Comparação de Médias de Várias Populações

Tabela 5: Dados da Tabela (4)

Fertilizante	A	B	C	D	E	F	**Soma**
Variedade 1	5,4	3,2	3,8	4,6	5,0	4,4	**26,4**
Variedade 2	5,7	4,0	4,2	4,5	5,3	5,0	**28,7**
Soma	**11,1**	**7,2**	**8**	**9,1**	**10,3**	**9,4**	**55,1**

Tabela 6: Quadro de análise de variância

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	$F_{Tabelado}$
Fertilizante	5,154**	5	1,0308	22,03	$F_{5;5;1\%}$ = 10,97
Variedade	0,441***	1	0,4410	9,42	$F_{1;5;1\%}$ = 16,26
Erro	0,234	5	0,0468		
Total	5,829*	11			

* $SQ_{total} = [(5{,}4)^2 + (3{,}2)^2 + \ldots + (5)^2] - (55{,}1)^2 / 12$

** $SQ_{fertilizante} = [(11{,}1)^2 + (7{,}2)^2 + \ldots + (9{,}4)^2)] / 2 - (55{,}1)^2 / 12$

*** $SQ_{variedade} = [(26{,}4)^2 + (28{,}7)^2] / 6 - (55{,}1)^2 / 12$

Fertilizante	A	B	C	D	E	F	**Soma**
Variedade 1	5,4	3,2	3,8	4,6	5,0	4,4	**26,4**
Variedade 2	5,7	4,0	4,2	4,5	5,3	5,0	**28,7**
Soma	**11,1**	**7,2**	**8**	**9,1**	**10,3**	**9,4**	**55,1**

Fertilizante	A	B	C	D	E	F	**Soma**
Variedade 1	11	-11	-5	3	7	1	
Variedade 2	14	-3	-1	2	10	7	
Soma							

Tabela 6: Quadro de análise de variância

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	$F_{Tabelado}$
Fertilizante	515,4**	5	103,08	22,03	$F_{5;5;1\%}$ = 10,97
Variedade	44,1***	1	44,10	9,42	$F_{1;5;1\%}$ = 16,26
Erro	23,4	5	4,68		
Total	582,9*	11			

Comparação de Médias de Várias Populações

Tabela 6: Cálculo das médias das linhas e das colunas

Fertilizante	A	B	C	D	E	F	**Média**
Variedade 1	5,4	3,2	3,8	4,6	5,0	4,4	**4,4**
Variedade 2	5,7	4,0	4,2	4,5	5,3	5,0	**4,783**
Média	**5,55**	**3,6**	**4**	**4,55**	**5,15**	**4,7**	

1) Variedade:

$$\text{Zeta} = q_{k,\nu,\alpha}\sqrt{\frac{S_R^2}{n}} = q_{2,5,1\%}\sqrt{\frac{S_R^2}{n}} = 5{,}70\sqrt{\frac{0{,}0468}{6}} = 0{,}50$$

k=número de linhas=2; n=número de colunas=6,

ν =5, e $q_{k,\nu,\alpha}$ é a Amplitude Studantizada.

2) Fertilizante:

$$\text{Zeta} = q_{n,\nu,\alpha}\sqrt{\frac{S_R^2}{k}} = q_{6,5,1\%}\sqrt{\frac{S_R^2}{k}} = 8{,}91\sqrt{\frac{0{,}0468}{2}} = 1{,}36$$

k=número de linhas=2; n=número de colunas=6,

ν =5

Análise de Variância - ANOVA

2) Fertilizante:

$$\text{Zeta} = 1{,}36$$

$$\left|\overline{X}_A - \overline{X}_B\right| = 1{,}95$$

$$\left|\overline{X}_A - \overline{X}_C\right| = 1{,}55$$

$$\left|\overline{X}_B - \overline{X}_E\right| = 1{,}55$$

Conclusão:

Podemos concluir que há diferença entre os fertilizantes A e B, A e C e B e E.

Comparação de Médias de Várias Populações

3º Experimento:

Experimento com dois fatores, com repetição

		Operário			
		1	**2**	**3**	**4**
	I	54	46	55	51
	I	52	47	54	60
Método	**II**	59	61	59	56
	II	57	55	61	57
	III	59	63	63	59
	III	62	58	61	60

Temos três conjuntos de hipóteses:

1) Quanto ao operário

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \text{Ao menos um } \mu_i \neq \mu_j \text{, para i, j} \in \{1,2,3,4\} \text{ e i} \neq \text{j}$$

2) Quanto ao método

$$H_0 : \mu_I = \mu_{II} = \mu_{\mathrm{III}}$$

$$H_1 : \text{Ao menos um } \mu_i \neq \mu_j \text{, para i, j} \in \{\text{I, II, III}\} \text{ e i} \neq \text{j}$$

3) Quanto a interação método-operário

$$H_0 : \mu_{1I} = \mu_{2I} = \mu_{3\mathrm{I}} = \ldots = \mu_{4\mathrm{III}}$$

$$H_1 : \text{Ao menos um } \mu_i \neq \mu_j \text{, para i, j} \in \{\text{1I,2I,3I,...,4III}\} \text{ e i} \neq \text{j}$$

Tabela 7: Dados subtraídos de 59

Operário

		1		**2**		**3**		**4**		
	I	-5	***-12***	-13	***-25***	-4	***-9***	-8	***-7***	-53
	I	-7		-12		-5		1		
Método	**II**	0	***-2***	2	***-2***	0	***2***	-3	***-5***	-7
	II	-2		-4		2		-2		
	III	0	***3***	4	***3***	4	***6***	0	***1***	13
	III	3		-1		2		1		
		-11		-24		-1		-11		**-47**

Tabela 8: Quadro de Análise de Variância

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	$F_{Tabelado}$
Método	286,34**	2	143,17	20,08	6,93
Operário	44,46***	3	14,82	2,08	5,95
Método-Operário	72,66****	6 (2x3)	12,11	1,70	4,82
Erro	85,50	12	**7,13**		
Total	488,96*	23			

$$* \ SQ_{total} = [(-5)^2 + (-13)^2 + \ldots + (1)^2] - (-47)^2 / 24$$

$$** \ SQ_{metodo} = [(-53)^2 + (-7)^2 + (13)^2)] / 8 - (-47)^2 / 24$$

$$*** \ SQ_{operario} = [(-11)^2 + (-24)^2 + (-1)^2 + (-11)^2] / 6 - (-47)^2 / 24$$

$$**** \ SQ_{metodo-operario} = [(-12)^2 + (-25)^2 + \ldots + (1)^2] / 2 - (-47)^2 / 24 - SQ_{metodo} - SQ_{operario}$$

Conclusão:

De acordo com o Quadro de Análise de Variância, apenas o **método** influencia no tempo de execução da tarefa.

Não há diferença entre operários e nem interação entre Método-Operário, ou seja, não existe um método específico com o qual um determinado operário se adapte melhor.

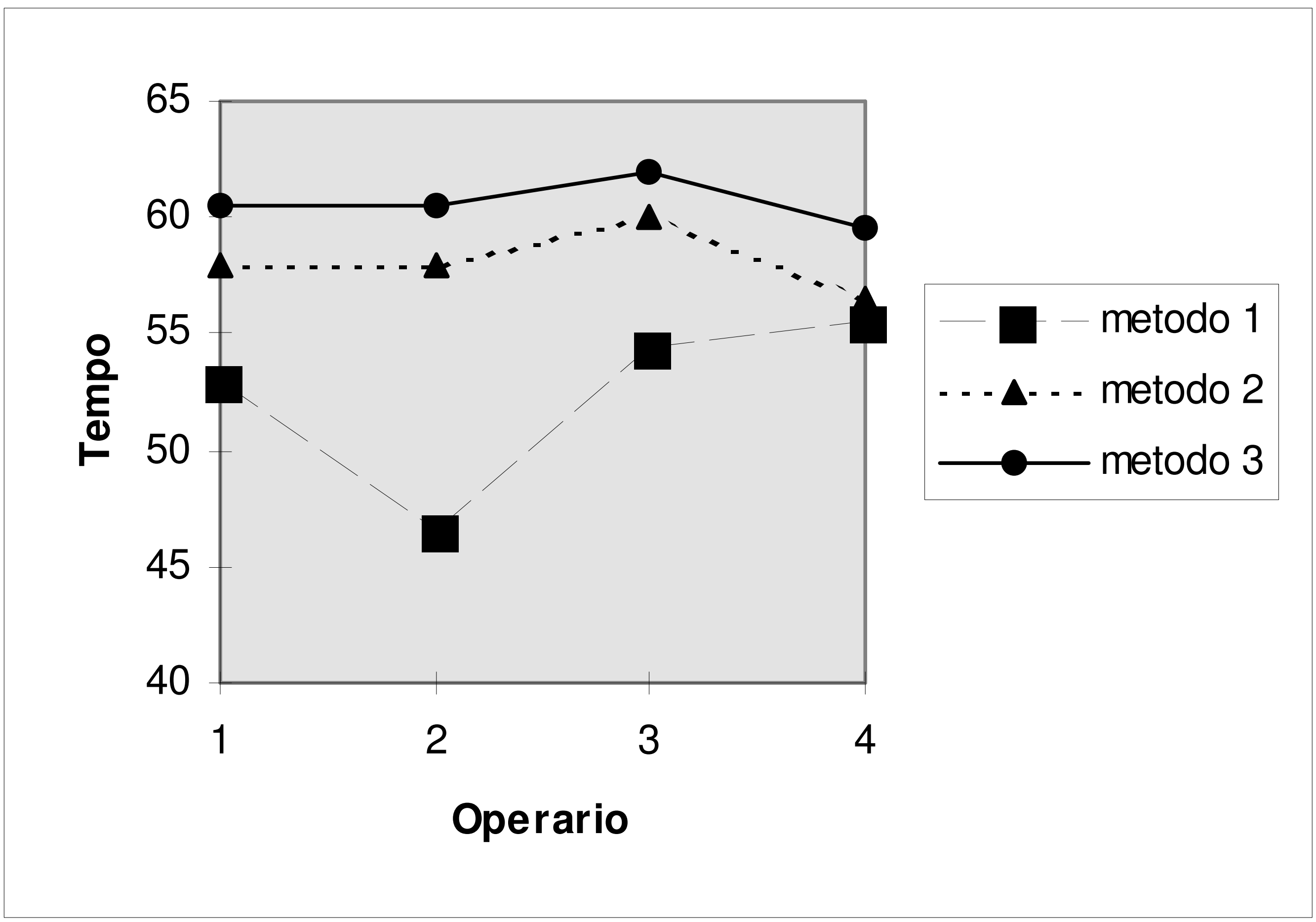


Figura 1: Interação Método - Operário

		Operário			
		1	**2**	**3**	**4**
	I	54	46	55	51
	I	52	47	54	60
Método	**II**	59	61	59	56
	II	57	55	61	57
	III	59	63	63	59
	III	62	58	61	60