

Prova 1

César A. Galvão - 19/0011572

2022-07-11

Contents

1	Questao 1	3
1.1	3
1.2	3
1.3	4
1.4	4
1.5	5
1.6	6
2	Questão 2	7
2.1	7
2.2	7

1 Questao 1

rep	Emb1	Emb2	lac2022	lacSP
Rep_1	517	504	499	485
Rep_2	521	515	493	477
Rep_3	516	502	497	487
Rep_4	522	506	487	496

1.1

A comparação das médias dos grupos, neste caso as empresas, será realizada mediante análise de variância. O modelo escolhido para tal é o modelo de efeitos, expresso na equação a seguir

$$y_{ij} = \mu + \tau_i + e_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, n \quad (1)$$

em que μ é a média geral, τ_i é a média ou efeito dos grupos e e_{ij} é o desvio do elemento. Os grupos são indexados por i e os indivíduos de cada grupo indexados por j .

As hipóteses do teste são as seguintes:

$$\begin{cases} H_0 : \tau_1 = \dots = \tau_a = 0, & (\text{O efeito de tratamento é nulo}) \\ H_1 : \exists \tau_i \neq 0 \end{cases} \quad (2)$$

que equivale dizer

$$\begin{cases} H_0 : \mu_1 = \dots = \mu_a \\ H_1 : \exists \mu_i \neq \mu_j, i \neq j. \end{cases} \quad (3)$$

Realiza-se inicialmente a análise de variância, cujos resultados são expostos na tabela a seguir:

term	df	sumsq	meansq	statistic	p.value
empresa	3	2490.5	830.167	25.446	0
Residuals	12	391.5	32.625	NA	NA

Pelo p-valor ínfimo (arredondado para 0 considerando 3 dígitos decimais) constante na tabela de análise de variância, pode-se rejeitar a hipótese de que não há diferença entre as médias dos grupos. Isto quer dizer que há pelo menos uma média diferente das demais a $\alpha = 0,05$.

1.2

Para confirmar esses resultados, realiza-se testes diagnósticos de normalidade e homocedasticidade sobre os resíduos, para verificar os pressupostos necessários para a análise de variância.

Na tabela a seguir, considerando que a hipótese nula do teste de Shapiro-Wilk é a normalidade dos dados – neste caso dos resíduos – não se pode rejeitar normalidade dos dados com o nível de confiança desejado.

statistic	p.value	method
0.9816996	0.9756089	Shapiro-Wilk normality test

Pode-se dizer pelo p-valor da tabela a seguir, em que constam os resultados do teste de Levene para homocedasticidade cuja hipótese nula é a igualdade de variâncias, que as variâncias entre os grupos são iguais.

teste	F statistic	p.value	df	df.residual
Teste Levene de Homogeneidade	0.4	0.7555447	3	12

Conclui-se portanto que os pressupostos para a realização da ANOVA estão cumpridos e, conforme a tabela desta análise, existe pelo menos uma média de grupo diferente das demais.

1.3

Estima-se, considerando \bar{x} o estimador natural para μ e $QMRES = \hat{\sigma}^2$:

\bar{x}	τ_1	τ_2	τ_3	τ_4	$\hat{\sigma}^2$
501.5	519	506.75	494	486.25	32.62

1.4

Compararemos dois subgrupos, formados a partir do conjunto inicial de tratamentos, para realizar o teste de comparação de médias utilizando contrastes. A saber, compararemos a média das medidas de IAC2022 com a média dos demais. Construímos os seguintes contrastes:

$$\Gamma_1 = \sum_{i=1}^a c_i \mu_i \quad \text{em que} \quad \sum_{i=1}^4 c_i = 0; \text{ e } c_i = \left\{ -\frac{1}{3}, -\frac{1}{3}, 1, -\frac{1}{3} \right\} \quad (4)$$

Para a construção dos demais contrastes ortogonais, fazemos

$$\Gamma_2 = \sum_{i=1}^3 c_i \mu_i \quad \longrightarrow c_i = \left\{ 1, -\frac{1}{2}, 0, -\frac{1}{2} \right\}$$

$$\Gamma_3 = \sum_{i=1}^2 c_i \mu_i \quad \longrightarrow c_i = \{0, 1, 0, -1\}$$

de modo que todos os $c_i, i = 1, 2, 3$ são ortogonais entre si. Dessa forma, as hipóteses testadas são as seguintes:

$$\begin{aligned} \text{Contraste 1: } & \begin{cases} H_0 : \mu_3 = \frac{\mu_1 + \mu_2 + \mu_4}{3} \\ H_1 : \mu_3 \neq \frac{\mu_1 + \mu_2 + \mu_4}{3} \end{cases} \\ \text{Contraste 2: } & \begin{cases} H_0 : \mu_1 = \frac{\mu_2 + \mu_4}{2} \\ H_1 : \mu_1 \neq \frac{\mu_2 + \mu_4}{2} \end{cases} \text{ e} \\ \text{Contraste 3: } & \begin{cases} H_0 : \mu_2 = \mu_4 \\ H_1 : \mu_2 \neq \mu_4 \end{cases} \end{aligned}$$

A estatística de teste para a realização dos contrastes é definida conforme a expressão a seguir, em que QMRES é a soma de quadrados dos resíduos da ANOVA exposta anteriormente:

$$\frac{\left(\sum_{i=1}^n c_i \bar{y}_{i.} \right)^2}{\sum_{i=1}^n c_i^2 \frac{\text{QMRES}}{n}} \sim F(1, an - a = 12) \quad (5)$$

Além disso, considera-se

$$\frac{\left(\sum_{i=1}^n c_i \bar{y}_{i.} \right)^2}{\sum_{i=1}^n c_i^2 \frac{1}{n}} = \frac{\text{SQContraste}_i}{1(g.l.)} = \text{QMContraste}_i \quad (6)$$

tal que, se os contrastes forem calculados da forma correta, a soma dos quadrados médios dos contrastes deve ser igual ao quadrado médio dos tratamentos.

As estatísticas são expostas na tabela de análise de variância a seguir, decomposta em seus contrastes.

Fonte de variação	g.l.	SQ	MQ	Estatística F	p-valor
empresa	3	2490.5	830.1667	25.4457	0.0000
C1	1	300.0	300.0000	9.1954	0.0104
C2	1	1350.0	1350.0000	41.3793	0.0000
C3	1	840.5	840.5000	25.7625	0.0003
Residuals	12	391.5	32.6250	NA	NA

Dessa forma, pode-se dizer que há diferença entre as médias de todas as empresas listadas.

1.5

Calcula-se a DHS de Tukey da seguinte forma:

USAR qtukey sqres(qmres/n)

$$q_s = \frac{Y_A - Y_B}{\sqrt{\text{QMRES}/n}} \quad (7)$$

Tal que o numerador seja a diferença entre a maior e a menor das médias e o denominador seja o erro padrão. Encontra-se o valor da DHS de 11.467.

A tabela do teste de Tukey é exposta a seguir:

Contraste	D. estimada.	LI	LS	p-valor ajust.
Emb2-Emb1	-12.25	-24.241	-0.259	0.0448
lac2022-Emb1	-25.00	-36.991	-13.009	0.0002
lacSP-Emb1	-32.75	-44.741	-20.759	0.0000
lac2022-Emb2	-12.75	-24.741	-0.759	0.0361
lacSP-Emb2	-20.50	-32.491	-8.509	0.0013
lacSP-lac2022	-7.75	-19.741	4.241	0.2711

A um nível de significância de 5%, de fato parece não haver diferença significativa apenas entre lacSP e lac2022. Isso é evidenciado na tabela do teste de Tukey tanto pelo p-valor quanto pela distância estimada, maior que a DHS calculada.

1.6

Desejamos calcular $\beta(\tau_1 = 507, \tau_2 = 501, \tau_3 = 497, \tau_4 = 495)$. Para isso, utilizaremos $n = 4, \alpha = 0,05$ e $\sigma^2 = \frac{QMRES}{n}$. A probabilidade será calculada da seguinte forma:

$$P\left(F_{\text{obs}} < F_{\text{crit}} \mid \phi^2 = \frac{n}{\sigma^2} \sum_{i=1}^4 \tau_i^2\right), \quad (8)$$

considerando a variância para os resíduos. Portanto,

$$\hat{\phi}^2 = \frac{n}{QMRES} \sum_{i=1}^4 \tau_i^2 \quad (9)$$

é o parâmetro de não-centralidade (pnc ou, em inglês, *ncp*) da distribuição F e, sob H_0 , $\phi^2 = 0$.

O valor $F_{\text{crit}} = F(\gamma = 0,95; gl_1 = 3; gl_2 = 12, \phi^2 = 0)$ é de 3.49. Considerando $\phi^2 = 10.2989$, obtém-se

$$\begin{aligned} P(F_{\text{obs}} < F_{\text{crit}} \mid \phi^2 \text{ sob } H_1) &= P(F_{\text{obs}} < 3,49 \mid \hat{\phi}^2 = 10,299) \\ &= 0,38 \end{aligned}$$

2 Questão 2

2.1

SP	RJ
6062	5682
6116	5714
6070	5716
5942	5665
5990	5589
6034	5702
6004	5688
5969	5720
5950	5804
5941	5850

Deseja-se realizar o seguinte teste de hipóteses:

$$\begin{cases} H_0 : \mu_{SP} \leq \mu_{RJ}; \text{ou} & \mu_{SP} - \mu_{RJ} \leq 0 \\ H_A : \mu_{SP} > \mu_{RJ}; \text{ou} & \mu_{SP} - \mu_{RJ} > 0 \end{cases} \quad (10)$$

Trata-se de um teste de comparação de médias simples. Avalia-se inicialmente teste de normalidade e igualdade de variâncias sobre os dados. A tabela a seguir exhibe resultados de testes Shapiro-Wilk, de acordo com os quais não há evidências para se rejeitar normalidade.

UF	statistic	p.value	method
RJ	0.9262958	0.4124612	Shapiro-Wilk normality test
SP	0.9238969	0.3906054	Shapiro-Wilk normality test

```
## Multiple parameters; naming those columns num.df, den.df
```

Ao se realizar teste de igualdade de variâncias, obtém-se p-valor de 0.623. Dessa forma, pode-se considerar que ambas as amostras advêm de populações com variâncias iguais.

Finalmente, realiza-se o teste de comparação de médias. Será considerado um teste não pareado, com amostras de variâncias iguais.

Obtém-se estatística de teste $t = 9,9086$ com 18 graus de liberdade e p-valor significativo a nível de significância inferior a 0,001. Pode-se dizer portanto que de fato a média salarial de SP é superior à do RJ. Um intervalo de confiança para essa diferença compreende o intervalo $[243, 21; +\infty)$.

2.2

$$P(t_{\text{obs}} > t_{\text{crit}} | \mu_A = 100) \quad (11)$$

$$t_{\text{crit}} = \frac{294 - 100}{S_{\text{comb}} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}} \quad (12)$$

$$= \frac{294 - 100}{66,527 \cdot \sqrt{\frac{2}{10}}} \quad (13)$$

$$= 6,547 \quad (14)$$

Calcula-se portanto a probabilidade de se observar um valor $t_{\text{obs}} > 6,547$ usando `pt(tcrit, 18, lower.tail = FALSE)`. Obtem-se 0.0000019.