



Comparações múltiplas

Universidade Estadual de Santa Cruz

Ivan Bezerra Allaman

Introdução

- Quando há **mais de dois tratamentos** e o teste de F rejeita H_0 apontando que pelo menos uma das médias difere das demais, então um teste de comparação múltipla se faz necessário para que possamos ranquear os tratamentos.
- Um erro comum é os usuários utilizarem testes de comparações múltiplas quando há apenas dois tratamentos. Perceba que quando há dois tratamentos e o teste de F rejeita H_0 , basta observamos as médias e ranquear os tratamentos.
- Outro erro comum é utilizar testes de comparações múltiplas quando os tratamentos são **quantitativos** em que a quantidade de níveis estudados permite o uso de algum tipo de regressão.
- Os métodos abaixo serão abordados considerando apenas **dados balanceados**.

Contrastes ortogonais

- É utilizado quando o pesquisador sabe especificamente quais tratamentos ele deseja comparar.
- Neste caso, não há interesse em comparar todos os tratamentos entre si.
- No quadro da análise de variância, desdobramos a soma de quadrados nos contrastes que queremos comparar e testamos por meio do teste de F.
- O conjunto de contrastes que irão ser testados no quadro da anova devem ser ortogonais, ou seja, a soma dos coeficientes do contraste e entre os contrastes devem ser iguais a zero.
- Logo, um contraste é uma combinação linear de totais de tratamentos ou de médias da forma:

$$\hat{Y} = c_1 y_{1.} + c_2 y_{2.} + c_k y_{k.}$$

em que $\sum_{i=1}^k c_i = 0$.

- Se há mais de um contraste de interesse, ou seja,

$$\hat{Y}_1 = c_1 y_1. + c_2 y_2. + \cdots + c_k y_k.$$

$$\hat{Y}_2 = d_1 y_1. + d_2 y_2. + \cdots + d_k y_k.$$

então $\sum_{i=1}^k c_i d_i = 0$.

- Há um detalhe muito importante que não podemos esquecer. O número de contrastes possíveis de analisarmos será sempre o número de tratamentos menos 1, ou seja, o mesmo graus de liberdade dos tratamentos.
- Se em um experimento há 4 tratamentos, então tem-se 3 graus de liberdade. Logo, podemos formar apenas 3 contrastes para serem testados na anova por vez.

- Uma vez que testamos os contrastes por meio da análise de variância, precisamos saber como se calcula a soma de quadrados. Utilizando médias tem-se:

$$SQ_{\hat{Y}} = \frac{r(\sum_{i=1}^k c_i \bar{y}_{i.})^2}{\sum_{i=1}^k c_i^2} = \frac{r\hat{Y}_i^2}{\sum_{i=1}^k c_i^2}$$

em que r é o número de repetições de cada tratamento.

Aplicação

1. Em uma experiência para comparar a eficiência de diversas técnicas no tratamento da dor produzida por uma intervenção cirúrgica superficial, 28 pacientes foram agrupados, ao acaso, em quatro grupos de sete, tratando-se o primeiro com placebo e, os demais, com dois tipos de analgésicos (A e B) e acupuntura. Os dados estão disponíveis no seguinte link:

http://nbcgib.uesc.br/lec/download/R/dados/DíazEx3_11.txt

O pesquisador está interessado em comparar o placebo versus os demais tratamentos, o analgésico A versus o B e os analgésicos A e B versus a acupuntura. Logo, temos três contrastes (\hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3) que podem ser visualizados no quadro abaixo:

	Acupuntura	AnalgA	AnalgB	Placebo
\hat{Y}_1	-1,00	-1,00	-1,00	3,00
\hat{Y}_2	0,00	-1,00	1,00	0,00
\hat{Y}_3	2,00	-1,00	-1,00	0,00

Vamos verificar se o contraste é ortogonal. Segue,

$$(-1 \cdot 0 \cdot 2) + (-1 \cdot -1 \cdot -1) + (-1 \cdot 1 \cdot -1) + (3 \cdot 0 \cdot 0) = 0$$

Uma vez constatado a ortogonalidade, vamos calcular a soma de quadrados de cada contraste e fazermos a anova.

As médias são:

$$y_{acup\cdot} = 114,57; y_{analgA\cdot} = 85; y_{analgB\cdot} = 94,86; y_{plac\cdot} = 31,57.$$

Calculando os contrastes tem-se:

$$\hat{Y}_1 = -1 \cdot 114,57 - 1 \cdot 85 - 1 \cdot 94,86 + 3 \cdot 31,57 = -199,7143$$

$$\hat{Y}_2 = 0 \cdot 114,57 - 1 \cdot 85 + 1 \cdot 94,86 + 0 \cdot 31,57 = 9,8571$$

$$\hat{Y}_3 = 2 \cdot 114,57 - 1 \cdot 85 - 1 \cdot 94,86 + 0 \cdot 31,57 = 49,2857$$

Calculando a soma de quadrados levando em consideração 7 repetições por tratamento (r=7) tem-se:

$$SQ\hat{Y}_1 = \frac{7 \cdot (-199,7143)^2}{12} = 23266,71$$

$$SQ\hat{Y}_2 = \frac{7 \cdot (9,8571)^2}{2} = 340,07$$

$$SQ\hat{Y}_3 = \frac{7 \cdot (49,2857)^2}{6} = 2833,93$$

Portanto, segue o quadro da análise de variância:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trat	3	26440,71	8813,57	12,64	0,0000
trat: \hat{Y}_1	1	23266,71	23266,71	33,37	0,0000
trat: \hat{Y}_2	1	340,07	340,07	0,49	0,4916
trat: \hat{Y}_3	1	2833,93	2833,93	4,06	0,0551
Residuals	24	16732,29	697,18		

Logo, podemos concluir que considerando um nível de 5% de significância, apenas o contraste 1 foi significativo, ou seja, em média os tratamentos estudados em comparação ao placebo prolongam a duração do paciente até sentir dor novamente.

Teste de Dunnett

- É utilizado quando o único interesse é comparar um tratamento controle (testemunha) versus os demais tratamentos.
- O objetivo maior é apontar os tratamentos que são melhores que um padrão ou controle.
- A estatística de teste é uma modificação da estatística t de student e logo iremos definir por:

$$t_D = \frac{\bar{y}_{i.} - \bar{y}_c}{\sqrt{\frac{2}{r} \cdot QME}}$$

em que:

$\bar{y}_{i.}$ é a média do tratamento que queremos comparar com o controle, \bar{y}_c é a média do tratamento controle, r é o número de repetições, QME é o quadrado médio do resíduo,

- Uma vez calculado a estatística de teste é necessário saber qual é o p-valor.
- Neste caso iremos utilizar a função **pNCDun** do pacote **nCDunnett** do **r**. Para utilizar tal função, segue a explicação de seus parâmetros:

$$pNCDun(t_D, glres, rho, delta)$$

em que:

t_D = é o quantil da distribuição de Dunnett (já foi mostrado o cálculo no slide anterior).

glres = é o graus de liberdade do erro da anova.

rho = é um vetor de correlação. Utilizaremos o valor de 0,5. Se na pesquisa o graus de liberdade de tratamentos foi 3, então devemos criar um vetor de tamanho 3 com o valor de 0,5.

delta = é um vetor com o parâmetro de não-centralidade. Em nossos casos iremos utilizar o valor 0. Se na pesquisa o graus de liberdade de tratamentos foi 3, então devemos criar um vetor de tamanho 3 com o valor de 0.

Aplicação

2. Para avaliar a influência do tipo de acidose nos níveis de glicemia no cordão umbilical de um recém-nascido, obtiveram-se os dados que estão disponíveis no seguinte link: http://nbcgib.uesc.br/lec/download/R/dados/DíazEx1_11.txt

Após realizado a análise de variância e constatado diferenças significativas entre os tratamentos, tem-se o seguinte quadrado médio do erro $QME = 40,87$. O número de repetições é igual a $r = 10$. O graus de liberdade do erro é 36.

As médias dos tratamentos são:

	Médias
Acid. metabólica	79,40
Acid. mista	80,20
Acid. respiratória	70,20
Controle	62,80

Vamos calcular a estatística de teste para cada comparação. Será apresentado em detalhes apenas o cálculo para comparação entre acidose metabólica e o controle. Para as demais comparações o cálculo é semelhante.

$$t_{D\text{acidmeta vs controle}} = \frac{79,4 - 62,80}{\sqrt{2/10 \cdot 40,87}} \\ = 5,806$$

Para as demais comparações tem-se as estatísticas de teste: $t_{D\text{acidmist vs controle}} = 6,086$ e $t_{D\text{acidresp vs controle}} = 2,588$.

Utilizando o pacote já mencionado, tem-se os seguintes p-valores para as comparações **acidmeta vs controle**, **acidmist vs controle** e **acidresp vs controle** respectivamente:

```
1-pNCDun(tD = 5.806, glres = 36, rho = c(0.5,0.5,0.5), delta =  
c(0,0,0)) = 0,0000104
```

```
1-pNCDun(tD = 6.086, glres = 36, rho = c(0.5,0.5,0.5), delta =  
c(0,0,0)) = 0,00000837
```

```
1-pNCDun(tD = 2.588, glres = 36, rho = c(0.5,0.5,0.5), delta =  
c(0,0,0)) = 0,0366
```

Considerando um $\alpha = 0,05$ podemos concluir que há diferenças de todos os tipos de acidose em relação ao tratamento controle.

Teste de Tukey

- O teste de Tukey é indicado para Comparações 2 a 2, no qual o pesquisador está interessado em comparar todos contra todos.
- O teste de Tukey é considerado muito rigoroso, ou seja, controla bem a taxa de erro do tipo I. Logo, para que tal teste aponte diferenças entre os tratamentos estudados é necessário que a variância devido ao acaso (erro experimental) seja pequeno. Pode acontecer do teste F na análise de variância apontar diferenças entre os tratamentos e o teste de Tukey não conseguir discriminar nenhum tratamento devido ao erro experimental estar elevado.
- Muitas vezes os pesquisadores utilizam o teste de Tukey de maneira indiscriminada, pela comodidade de comparar todos contra todos. Em muitos casos é possível formar contrastes que irão responder diretamente a hipótese do pesquisador, não precisando deste modo de comparações inúteis ou desnecessárias.

- Para que possamos ranquear os tratamentos de acordo com o teste em questão, devemos calcular a **diferença mínima significativa (DMS)** que é dada pela seguinte expressão:

$$DMS = q_{\alpha}(t, \text{glres}) \cdot \sqrt{\frac{QME}{r}}$$

em que:

$q_{\alpha}(t, \text{glres})$ = valor da amplitude total estudentizada, num nível α de probabilidade, obtido da tabela de Tukey, em função do número de tratamentos (t) e do número de graus de liberdade do erro (glres).

QME = quadrado médio do erro.

r = número de repetições

- No nosso caso, o quantil $q_\alpha(t, glres)$ iremos obter utilizando a função `qtukey(p, nmeans, df)` em que:

p = é o nível de **confiança** adotado na pesquisa,

$nmeans$ = é o número de tratamentos,

df = é o graus de liberdade do erro experimental obtivo da anova.

- Após o cálculo da DMS, é necessário dispor as médias em ordem decrescente, da maior para a menor média e em seguida colocamos letras do alfabeto para ilustrar o ranqueamento.
- Supondo que em uma pesquisa para avaliar 4 tratamentos (A,B,C e D) as médias foram respectivamente: 30, 26, 28 e 20 e que a DMS foi de 3, segue uma figura mostrando o passo a passo de como se deve colocar as letras.

1ª passo: Ordenar

Tratamentos	Médias
A	30
C	28
B	22
D	20

2ª passo: Inicie com a letra **a** do alfabeto a partir da maior média. Vai comparando com as médias subsequentes e repetindo a mesma letra que foi iniciada (letra **a** no caso). Quando a diferença entre as médias for maior do que a DMS calculado pare o processo.

Tratamentos	Médias
A	30 a
C	28 a
B	26
D	20

3ª passo: Recomeçe o processo com outra letra do alfabeto (**b** neste caso) a partir do segundo tratamento e prossiga como no passo anterior. Termine quando o último tratamento receber uma letra.

Tratamentos	Médias
A	30 a
C	28 a b
B	26 b
D	20 c

Aplicação

3. Considere um experimento (fictício) de alimentação de porcos em que se usaram quatro rações (A, B, C, D), cada uma fornecida a cinco animais escolhidos ao acaso. Os aumentos de peso observados, em quilogramas, constam na tabela a seguir:

A	B	C	D
35,00	40,00	39,00	27,00
19,00	35,00	27,00	12,00
31,00	46,00	20,00	13,00
15,00	41,00	29,00	28,00
30,00	33,00	45,00	30,00

Após o teste F na análise de variância ter apontado diferenças significativas entre os tratamentos, devemos então utilizar o teste de Tukey para ranquearmos os tratamentos do melhor para o pior. Vamos supor que o nível de significância adotado antes da pesquisa foi de 5%. Utilizando a função `qtukey` tem-se:

$$qtukey(0.95, 4, glres = 16) = 4,0461$$

O quadrado médio do resíduo (QME) foi: 68,75. O número de repetições é igual a 5. Logo, calculando a DMS temos:

$$DMS = 4,0461 \cdot \sqrt{\frac{68,75}{5}} = 15,003$$

Logo, o teste de Tukey é:

	Médias	G1	G2
B	39,00	a	
C	32,00	a	b
A	26,00	a	b
D	22,00		b

Segue a interpretação: o tratamento **B** proporcionou aumento de peso médio semelhante aos tratamentos **C** e **A** e superior ao tratamento **D**. Os tratamentos **C**, **A** e **D** foram semelhantes entre si.

Teste t de student (Least Significant
Difference - LSD)

- É o menos rigoroso entre os teste de comparações múltiplas, ou seja, tende a discriminar mais diferenças entre os tratamentos do que os demais métodos.
- Normalmente, quando o coeficiente de variação do experimento é moderadamente alto (entre 20 e 40%), o teste de Tukey pode não conseguir discriminar diferenças entre os tratamentos. Logo, o teste t pode ser utilizado nestes casos.
- A diferença mínima significativa é calculada como:

$$DMS = q_{(\alpha/2, glres)} \cdot \sqrt{QME \cdot \frac{2}{r}}$$

em que:

$q_{(\alpha/2, glres)}$ = é o quantil da distribuição t de student cujo valor pode ser obtido por meio da função `qt` do R. Vamos usar sempre o valor absoluto.

$glres$ = é o graus de liberdade do erro experimental.

QME e r são o quadrado médio do erro e o número de repetições respectivamente.

Aplicação

4. Foi realizado um estudo com 20 animais para comparação de quatro dietas. As dietas foram: A - dieta padrão; B - dieta padrão suplementada com amendoim; C - dieta padrão suplementada com girassol; D - dieta padrão suplementada com abóbora. Os ganhos de peso (g) avaliados considerando a variação absoluta entre o início e o final do experimento estão apresentados abaixo:

Dieta A	Dieta B	Dieta C	Dieta D
25,00	31,00	22,00	33,00
26,00	25,00	26,00	29,00
20,00	28,00	28,00	31,00
23,00	27,00	25,00	34,00
21,00	24,00	29,00	28,00

Após o teste F na análise de variância ter apontado diferenças significativas entre os tratamentos, devemos então utilizar o teste LSD para ranquearmos os tratamentos do melhor para o pior. Vamos supor que o nível de significância adotado antes da pesquisa foi de 5%. Utilizando a função qt tem-se:

$$|qt(0.05/2, glres = 16)| = 2,12$$

O quadrado médio do resíduo (QME) foi: 7. O número de repetições é igual a 5. Logo, calculando a DMS temos:

$$DMS = 2,12 \cdot \sqrt{7 \cdot \frac{2}{5}} = 3,547$$

Seguindo os mesmos passos para colocar as letras na tabela de comparações, segue o teste LSD.

	Médias	Groups
D	31,00	a
B	27,00	b
C	26,00	bc
A	23,00	c

Segue a interpretação: o tratamento **D** proporcionou ganho de peso médio superior a todos os tratamentos avaliados. O tratamento **B** foi semelhante ao tratamento **C** e superior ao tratamento **A**. O tratamento **C** foi tanto semelhante ao tratamento **B** quanto ao tratamento **A**.

Teste de Scott-Knott

- É um excelente teste de comparação múltipla, pois controla bem as taxas de erro do tipo I e II.
- Tem como vantagem aos demais testes o fato de não sobrepor médias, facilitando demasiadamente a interpretação dos resultados.
- Até o início de 2017 só era possível para dados desbalanceados, quando Conrado et al. 2017 estenderam o teste para dados desbalanceados.
- O particionamento dos grupos é um tanto oneroso e complexo a medida que aumenta o número de tratamentos, dependendo deste modo de um recurso computacional para aplicação do teste.

- O teste de Scott-Knott funciona da seguinte maneira:
 1. Ordenam-se as médias dos tratamentos para se obter todas as partições possíveis dos tratamentos em dois subgrupos. A expressão usual do número de partições é $k-1$, onde k é o número de tratamentos considerados.
 2. Determina-se todas as somas de quadrado (B_{0i}) entre dois subgrupos das $k-1$ partições possíveis pela expressão

$$B_{0i} = \frac{T_{1i}^2}{r_{1i}} + \frac{T_{2i}^2}{r_{2i}} - \frac{(T_{1i} + T_{2i})^2}{r_{1i} + r_{2i}}$$

em que:

B_{0i} é a soma de quadrado da *i*-ésima partição;

T_{1i} é a soma total dos valores das médias dos tratamentos pertencentes ao subgrupo 1 da *i*-ésima partição;

T_{2i} é a soma total dos valores das médias dos tratamentos pertencentes ao subgrupo 2 da *i*-ésima partição;

r_{1i} é o número de tratamentos do subgrupo 1 da *i*-ésima partição;

r_{2i} é o número de tratamentos do subgrupo 2 da *i*-ésima partição.

Seja B_0 a maior soma de quadrados entre dois subgrupos de médias, $B_0 = \max B_{0i}$,

3. Obtém-se o valor da estatística λ do teste pela seguinte expressão

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_0}{\hat{\sigma}_0^2}$$

em que π é a constante 3, 14...

O estimador de máxima verossimilhança $\hat{\sigma}_0^2$ é calculado como:

$$\hat{\sigma}_0^2 = \frac{1}{(k + v)} \cdot \left[\sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + v \frac{QME}{r} \right]$$

em que

v é o graus de liberdade do erro da anova;

$\bar{y}_{i.}$ é a média do **i-ésimo** tratamento, com $i = 1, \dots, k$;

$\bar{y}_{..}$ é a média geral;

QME é o quadrado médio do erro;

r é o número de repetições;

A estatística da razão de verossimilhança λ tem distribuição χ^2 com $v_0 = k/(\pi - 2)$ graus de liberdade, sob H_0 , e considera que todos os subgrupos da partição formada são iguais.

4. Se $\lambda > \chi_{\alpha, v_0}^2$ onde χ_{α, v_0}^2 corresponde ao quantil superior α de uma distribuição qui-quadrado com v_0 graus de liberdade, rejeita-se a hipótese de que os dois subgrupos são iguais em favor da hipótese alternativa de que os mesmos são diferentes entre si;
5. Repetem-se os passos de (1) a (4) até não se rejeitar a hipótese de formação de novas partições dentro dos subgrupos considerados.

Aplicação

5. Vamos aproveitar os dados da aplicação 4.

Passo 1:

Considerando o ordenamento do maior para o menor, tem-se as seguintes partições possíveis:

	G1	G2
partição 1	D	B, C, A
partição 2	D, B	C, A
partição 3	D, B, C	A

Passo 2:

Calculando a soma de quadrados das partições (B_0).

x

B_{01} 24,08

B_{02} 20,25

B_{03} 18,75

Logo, o valor máximo de $B_0 = 24,08$.

Passo 3:

Calculando a estatística λ . Antes, vamos calcular o valor de $\hat{\sigma}_0^2$.

Considerando

$k = 4, v = 16, \sum(\bar{y}_{i.} - \bar{y}_{..})^2 = 32,75, QME = 7, r = 5$ tem-se:

$$\hat{\sigma}_0^2 = \frac{1}{4 + 16} \cdot \left[32,75 + 16 \cdot \frac{7}{5} \right] \\ = 2,7575$$

Logo a estatística λ é:

$$\lambda = \frac{3,14}{2(3,14 - 2)} \cdot \frac{24,08}{2,7575} = 12,0173798$$

Considerando $v_0 = 4/(3,14 - 2) = 3,5038768$ tem-se o seguinte p-valor:

p-valor = $pchisq(12.01238, 3.50388, lower.tail = F) = 0.01153.$

Se adotarmos $\alpha = 5\%$, rejeitamos a hipótese de que os grupos são iguais. Logo, iremos repetir os passos com os tratamentos B, C e A.

Passo 1:

Tem-se as seguintes partições:

	G1	G2
partição 1	B	C, A
partição 2	B, C	A

Passo 2:

Calculando a soma de quadrados das partições (B_0).

x

B_{01} 4,17

B_{02} 8,17

Logo, o valor máximo de $B_0 = 8,17$.

Passo 3:

Calculando a estatística λ . Antes, vamos calcular o valor de $\hat{\sigma}_0^2$.

Considerando

$k = 3, v = 16, \sum(\bar{y}_{i.} - \bar{y}_{..})^2 = 8,67, QME = 7, r = 5$ tem-se:

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{1}{3 + 16} \cdot \left[8,6666667 + 16 \cdot \frac{7}{5} \right] \\ &= 1,635\end{aligned}$$

Logo a estatística λ é:

$$\lambda = \frac{3,14}{2(3,14 - 2)} \cdot \frac{8,17}{1,6350877} = 4,0750977$$

Considerando $v_0 = 3/(3,14 - 2) = 2,6279076$ tem-se o seguinte p-valor:

p-valor = `pchisq(4.075, 2.628, lower.tail = F)` = 0.2043.

Logo, considerando $\alpha = 5\%$ não rejeita-se a hipótese de que os grupos são iguais. Portanto, tem-se a seguinte apresentação dos resultados.

Levels	Means	SK(5%)
D	31,00	a
B	27,00	b
C	26,00	b
A	23,00	b