Mais sobre testes de comparação múltipla

Prof. Caio Azevedo

- Já vimos como realizar comparações de interesse (em termos de igualdade de médias, existência de interação etc), através dos testes para a comparações $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.
- O teste visto para testar a hipótese acima, pode ser facilmente adaptado para testar as hipóteses

$$H_0 : \mathbf{C}_{(r \times p)}\boldsymbol{\beta}_{(p \times 1)} = \mathbf{M}_{(r \times 1)} \ vs \ H_1 : \mathbf{C}_{(r \times p)}\boldsymbol{\beta}_{(p \times 1)} \neq \mathbf{M}_{(r \times 1)}$$

- Basta utilizar a seguinte estatística

$$Q = \frac{1}{r\widehat{\sigma}^2}\left(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{M}\right)'\left(\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right)^{-1}\left(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{M}\right)$$

e proceder da mesma forma anterior ($\mathbf{M} = \mathbf{0}$).

- Entretanto, existem outros testes que servem para realizar comparações específicas (não tão gerais quanto as comparações de tipo $\mathbf{C}\beta = \mathbf{M}$).
- Veremos alguns desses testes.
- Uma preocupação (dado que podem existir muitas comparações de interesse) é controlar o nível de significância geral (considerando-se todos os testes).

- No caso de comparações do tipo $\mathbf{C}\boldsymbol{\beta}$ é aconselhável usar, em cada teste, um $\alpha^* = \alpha/m$, em que α é o nível de significância usado para os testes da tabela ANOVA e m é o número total de comparações de interesse.
- O processo acima é chamado de controle de Bonferroni.
- Os testes que veremos controlam, cada um à sua maneira, o nível de significância global.

- Vamos nos concentrar no PCA com um único fator (embora os desenvolvimentos possam ser estendidos para outras planejamentos).
- Primeiramente, lembremos o conceito de contraste.
- Um vetor $\mathbf{C}_{(1 \times p)} = [c_1 \; c_2 \ldots c_p]$ é dito ser um contraste se $\sum_{i=1}^{k} n_i c_i = 0$. No caso de experimento balanceados, basta que $\sum_{i=1}^{k} c_i = 0$.
- Uma matriz $\mathbf{C}_{(q \times p)}$ é dita ser uma matriz de contrastes se suas linhas forem contrastes.

- Lembrando: temos $\mu_1, \mu_2, ..., \mu_k$ médias e supomos que o teste F relativo à ANOVA rejeitou a igualdade simultânea das médias.
- Nosso interesse então é testar hipóteses do tipo

$$H_0 : \mathbf{C}_{(1\times k)}\boldsymbol{\mu}_{(k\times 1)} = 0 \;\; vs \;\; H_1 : \mathbf{C}_{(1\times k)}\boldsymbol{\mu}_{(k\times 1)} \neq 0$$

em que $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_k)$

- Defina, para um dado **C**, o parâmetro $\gamma = \mathbf{C}\boldsymbol{\mu} = \sum_{i=1}^{k} c_i\mu_i$.
- Um estimador natural para γ é $\widehat{\gamma} = \sum_{i=1}^{k} c_i\overline{Y}_i$,em que $\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ (estimador de mínimos quadrados do modelo completo).

- Portanto, tem-se que $\mathcal{V}(\widehat{\gamma}) = \sigma^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}$.
- Se o experimento for balanceado, temos que $\mathcal{V}(\widehat{\gamma}) = \frac{\sigma^2}{n} \sum_{i=1}^{k} c_i^2$, onde n é o número de unidades experimentais em cada tratamento.
- Um estimador para variância de $\widehat{\gamma}$ é dado por $\widehat{\mathcal{V}(\widehat{\gamma})} = \widehat{\sigma}^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}$
- Dessa forma, temos que $\dfrac{\widehat{\gamma} - \gamma}{\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}} \sim t_{(n-k)}$.
- Assim, um $$IC[\gamma; 1-\alpha] = \left[\widehat{\gamma} - t_{(\alpha/2, n-k)}\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}; \widehat{\gamma} + t_{(\alpha/2, n-k)}\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}\right].$$
- Portanto, podemos construir intervalos de confiança para constrastes de interesse e utilizá-los para avaliar a veracidade das hipóteses em questão.

- Seja $\mathbf{D} = (d_1, d_2, ..., d_k)$ um outro contraste.
- Dizemos que $\mathbf{C}$ e $\mathbf{D}$ são contrastes ortogonais se $\sum_{i=1}^{k} n_i c_i d_i = 0$. No caso de um experimento balanceado, basta que $\sum_{i=1}^{k} c_i d_i = 0$.
- Em geral, para um conjunto de k tratamentos, podemos definir diversos contrastes (ortogonais) entre si, que representem hipóteses de interesse.
- Retomemos o exemplo 2 (dados de absorbância).

Descrição do Exemplo 2

- Quanto maior a absorbância, melhor o solvente.
- Unidade experimental: 10 gramas de polpa do fruto de baguaçú.
- Casualização: a partir de 1 kg de polpa, foram sendo retiradas amostras de 10 gramas, onde foram aplicados os tratamentos, numa ordem aleatória.
- Experimento balanceado : mesmo número de observações (unidades experimentais) por nível do fator.
- Lembrando: tratamentos 1,2,3,4 e 5, representam respectivamente os tipos de solvente E50, EAW, MAW, E70, M1M.

Descrição do Exemplo 2

- Hipóteses de interesse:

$$
\begin{aligned}
H_0 : 2\mu_1 + 2\mu_2 + 2\mu_4 &= 3\mu_3 + 3\mu_5 \\
H_0 : \mu_1 + \mu_2 &= 2\mu_4 \\
H_0 : \mu_1 &= \mu_2 \\
H_0 : \mu_3 &= \mu_5
\end{aligned}
$$

- Implicam nos seguintes constrastes

$$
\begin{aligned}
\mathbf{C}_1 &= \begin{bmatrix} 2 & 2 - 3 & 2 & -3 \end{bmatrix} \\
\mathbf{C}_2 &= \begin{bmatrix} 1 & 1 & 0 & -2 & 0 \end{bmatrix} \\
\mathbf{C}_3 &= \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \end{bmatrix} \\
\mathbf{C}_4 &= \begin{bmatrix} 0 & 0 & 1 & 0 & -1 \end{bmatrix}
\end{aligned}
$$

Método de Scheffé para comparação de contrastes

- Considere um conjunto de m constrastes de interesse dados por

  $\gamma_u = c_{1u}\mu_1 + c_{2u}\mu_2 + ... + c_{ku}\mu_k, u = 1, ..., m$

- Os respectivos estimadores são dados por:

  $\widehat{\gamma}_u = \delta_u = c_{1u}\overline{Y}_1 + c_{2u}\overline{Y}_2 + ... + c_{ku}\overline{Y}_k, u = 1, ..., m$

- O erro-padrão associado ao u-ésimo estimador, é dado por

  $S_{\delta_u} = \sqrt{QMR \sum_{i=1}^{k} \frac{c_{iu}^2}{n_i}}$

  lembrando que QMR = Quadrado médio residual = $\widehat{\sigma}^2$

Método de Scheffé para comparação de contrastes (cont.)

- Schefé estabeleceu um valor crítico para o teste, da seguinte forma: Rejeita-se H_0 se $|\delta_u| > S_{\alpha,u} = S_{\delta_u}\sqrt{(k-1)F_{\alpha,k-1,n-k}}$ em que α é o nível de significância apropriado e $P(F > F_{\alpha,k-1,n-k}) = \alpha, F \sim F_{(k-1,n-k)}$
- Scheffé provou que a probabilidade do erro do tipo I para cada um dos testes não ultrapassa α.

Aplicando no exemplo

- As estimativas dos constrastes 1 e 2 são dadas por:

  $\delta_1 = 1,488$ e $\delta_2 = -0,109$

  e os respectivos erros-padrão, são dados por $S_{\delta_1} = 0,061$ e

  $S_{\delta_2} = -0,110$

- Assim, dado que para $\alpha = 0,05$, temos que $F_{(0,05,4,20)} = 2,866$, os valores críticos para cada teste, são dados por:

  $S_{0,05,1} = 0,209$; $S_{0,05,2} = 0,094$

- Portanto, $|\delta_1| > 0,209$ e $|\delta_2| > 0,094$. Assim, rejeita-se ambas as hipóteses.

Utilizando o R

- O teste de Scheffé também permite comparar médias par a par, dado que todas essas comparações estão relacionadas à contrastes.
- O procedimento é similar ao anterior.
- Existe uma pacote no R chamado *agricolae* que permite fazer comparações desse tipo, usando o método de Scheffé e outros que veremos.
- Vamos utilizá-lo em nosso exemplo.

Resultado da aplicação do teste de Scheffe

	tratamento	Média	Grupo	n	erro-padrão
1	E70	0,61	a	5	0,01
2	EAW	0,57	ab	5	0,01
3	E50	0,54	b	5	0,01
4	MAW	0,45	c	5	0,02
5	M1M	0,20	d	5	0,01

Testes específicos para comparação de pares de médias

- Apesar do teste de Scheffé também permiter comparação de médias duas a duas, ele tende a ser muito conservativo (rejeita igualdades entre as médias menos do que deveria).
- Veremos outros testes: Tukey, LSD de Fisher, Duncan e Dunnet.
- As hipóteses são

  $H_0 : \mu_i = \mu_j$ vs $H_1 : \mu_i \neq \mu_j, \forall i, j.$

Teste de Tukey

- O teste de Tukey faz uso de percentis da distribuição da seguinte estatística

$$Q = \frac{\overline{Y}_{max} - \overline{Y}_{min}}{\sqrt{QMR/n}} \tag{1}$$

em que n é o tamanho amostral para cada tratamento. Se o experimento for desbalanceado, pode-ser usar uma média aritmética dos tamanhos amostrais. Além disso $\overline{Y}_{max}$ é a maior média amostral e $\overline{Y}_{min}$ é a menor média amostral.

- O nível de significância global (considerando todos os testes) é exatamente igual à α.

Teste de Tukey (cont.)

- Rejeita-se H_0, para um dado α, se

$$|\overline{Y}_i - \overline{Y}_j| > T_\alpha,$$

em que $\overline{Y}_i$ média amostral do i-ésimo tratamento e

$$T_\alpha = \frac{q_\alpha(k,f)}{\sqrt{2}} \sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

f é o número de graus de liberdade do resíduo e $q_\alpha(k,f)$ é o quantil de ordem α da distribuição da estatística (1)

Resultado da aplicação do teste de Tukey

	tratamento	Média	Grupo	n	erro-padrão
1	E70	0,61	a	5	0,01
2	EAW	0,57	ab	5	0,01
3	E50	0,54	b	5	0,01
4	MAW	0,45	c	5	0,02
5	M1M	0,20	d	5	0,01

Teste LSD de Fisher

- O teste LSD ("Least significance difference") de Fisher, baseia-se na seguinte estatística

$$T = \frac{\overline{Y}_i - \overline{Y}_i}{\sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

- Rejeita-se H_0 se

$$|\overline{Y}_i - \overline{Y}_j| > t_{(\alpha/2,n-k)}\sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

em que $P(T > t_{(\alpha/2,n-k)}) = \alpha/2$, $T \sim t_{(n-k)}$

Resultado da aplicação do teste LSD

	Tratamento	média	grupo	n	erro-padrão	LIIC	LSIC
1	E70	0,61	a	5	0,01	0,59	0,62
2	EAW	0,57	b	5	0,01	0,55	0,58
3	E50	0,54	b	5	0,01	0,51	0,56
4	MAW	0,45	c	5	0,02	0,41	0,48
5	M1M	0,20	d	5	0,01	0,17	0,22

Teste de Duncan

- No teste de Duncan as médias amostrais são dispostas de modo crescente e para cada uma delas é calculado o erro-padrão: $S_{\overline{Y}_i} = \sqrt{QMR/n_h}, n_h = \frac{k}{\sum_{i=1}^{k} n_i^{-1}}$
- Obtem-se quantis tabelados por Duncan, denotados por $r_\alpha(p,f), p = 2,3,...,k$, α é o nível de significância e f são os graus de liberdade do resíduo.
- Calcula-se $R_p = r_\alpha(p,f)S_{\overline{Y}_i}, p = 2,3,..,k$
- Compara-se, então a maior média com a menor (tal diferença é comparada com R_k).

Teste de Duncan (cont.)

- Depois, compara-se a maior média com a segunda menor (tal diferença é comparada com R_{k-1}).
- Continua-se o processo acima até que todas as médias tenham sido comparadas com a maior.
- Depois, compara-se a segunda maior com a menor (tal diferença é comparada com R_{k-1}).
- Repete-se o processo até que todas as $\frac{k(k-1)}{2}$ diferenças tenham sido consideradas.

Teste de Duncan (cont.)

- Se a diferença observada for maior que $R_.$ rejeita-se H_0.
- Para evitar-se contradições, duas médias não serão consideradas diferentes, se elas estiverem entre duas outras médias que não foram consideradas diferentes.

Resultado da aplicação do teste de Duncan

	tratamento	Média	Grupo	n	erro-padrão
1	E70	0,61	a	5	0,01
2	EAW	0,57	b	5	0,01
3	E50	0,54	b	5	0,01
4	MAW	0,45	c	5	0,02
5	M1M	0,20	d	5	0,01

Teste de Dunnett (comparação com um tratamento controle)

- Seja $\mu_r, r \in \{1, 2, ..., k\}$ a média correspondente ao tratamento controle.
- As hipóteses de interesse são:

  $H_0 : \mu_i = \mu_r$ *vs* $\mu_i \neq \mu_r, \forall i \neq r$
- Rejeita-se H_0 se

  $|\overline{Y}_i - \overline{Y}_j| > d_\alpha(k-1, f)\sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_r}\right)}$

  em que a constante $d_\alpha(k-1, f)$ corresponde à valores tabelados por Dunnet, para um dado nível de significância α e graus de liberdade para o resíduo f.

Teste de Dunnett (cont.)

- Rejeita-se H_0 se

$$|\overline{Y}_i - \overline{Y}_j| > d_\alpha(k-1, f)\sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_r}\right)}$$

em que a constante $d_\alpha(k-1, f)$ corresponde à valores tabelados por Dunnet, para um dado nível de significância α e graus de liberdade para o resíduo f.

Resultado da aplicação do teste de Dunnet

Hipótese	Estimativa	Erro-padrão	Estatística	pvalor
E70 - E50 $= 0$	0,07	0,02	4,30	0,001
EAW - E50 $= 0$	0,03	0,02	1,73	0,279
M1M - E50 $= 0$	-0,34	0,02	-21,48	$<$ 0,001
MAW - E50 $= 0$	-0,09	0,02	-5,62	$<$ 0,001