Mais sobre testes de comparação múltipla / Seleção e comparação de modelos

Notas de aula do Prof. Caio Azevedo modificadas pela Profa. Larissa Avila Matos

Revisão

- Já vimos como realizar comparações de interesse (em termos de igualdade de médias, existência de interação etc), através dos testes para a comparações $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}$.

- Como também já visto, o teste para testar a hipótese acima, pode ser facilmente adaptado para testar as hipóteses

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

- Vimos que, basta utilizar a seguinte estatística

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

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

- Entretanto, existem outros testes que servem para realizar comparações específicas (não tão gerais quanto as comparações de tipo $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{M}$).

- 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 $\boldsymbol{C\beta}$ 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 outros planejamentos).

- Primeiramente, lembremos o conceito de contraste.

- Um vetor $\boldsymbol{C}_{(1 \times p)} = [c_1 \, c_2 \ldots c_p]$ é dito ser um contraste se $\sum_{i=1}^{k} n_i c_i = 0$. No caso de experimentos balanceados, basta que $\sum_{i=1}^{k} c_i = 0$.

- Uma matriz $\boldsymbol{C}_{(q \times p)}$ é dita ser uma matriz de contrastes se suas linhas forem contrastes.

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

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

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

- Defina, para um dado $\boldsymbol{C}$, o parâmetro $\gamma = \boldsymbol{C}\boldsymbol{\mu} = \sum_{i=1}^{k} c_i\mu_i$.

- 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 $\dfrac{\widehat{\gamma} - \gamma}{\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}} \sim t_{(n-k)}$, e temos uma estatística do teste definida.

- Sob H_0 temos,

$$t_0 = \frac{\sum_{i=1}^{k} c_i \overline{Y}_i}{\sqrt{\widehat{\sigma}^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}}} \sim t_{(n-k)},$$

onde $p - valor = P(|t_{n-k}| > t_0)$.

- Assim, um intervalo de confiança para o contraste é dado por

$$IC[\gamma; 1-\alpha] = \left[\widehat{\gamma} - t_{(\alpha/2, n-k)}\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}; \widehat{\gamma} + t_{(\alpha/2, n-k)}\sqrt{\widehat{\mathcal{V}}(\widehat{\gamma})}\right].$$

- Portanto, podemos construir intervalos de confiança para constrastes de interesse e utilizá-los para avaliar a veracidade das hipóteses em questão.

- Ou seja, a relação do IC com o teste de hipótese é dada por: Se o IC contém zero, então não temos evidência para rejeitar H_0.

- Podemos testar contraste também utilizando a estatística F.

- Tomando-se o quadrado da estatística t_0 anterior, temos a estatística F_0 (sob H_0),

$$F_0 = t_0^2 = \frac{\left(\sum_{i=1}^{k} c_i \overline{Y}_i\right)^2}{\widehat{\sigma}^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}} \sim F_{(1,n-k)},$$

onde $p - valor = P(F_{(1,n-k)} > F_0)$.

- Então, rejeitamos H_0 se p-valor $< \alpha$ ou, de forma equivalente, se $F_0 > F_{(1,n-k)}$.

Contrastes Ortogonais

- Seja $\boldsymbol{D} = (d_1, d_2, ..., d_k)$ um outro contraste.
- Dizemos que $\boldsymbol{C}$ e $\boldsymbol{D}$ 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.
- Quando temos k tratamentos, sempre existe um conjunto de $k-1$ contrastes ortogonais que particiona SQF em componentes com 1 gl.

Método de Scheffé para comparação de contrastes

- Considere um conjunto de m constrastes de interesse dados por

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

- Os respectivos estimadores são dados por:

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

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

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

onde $\widehat{\sigma}^2 = QMR = \frac{SQR}{n-k}$.

- Scheffé estabeleceu um valor crítico para o teste, da seguinte forma:

$$\text{Rejeita-se } H_0 \; se \; |\delta_u| > S_{\alpha,u} = S_{\delta_u}\sqrt{(k-1)F_{\alpha,k-1,n-k}},$$

  em que α é o nível de significância apropriado e

$$P(F > F_{\alpha,k-1,n-k}) = \alpha, \quad F \sim F_{(k-1,n-k)}.$$

- Scheffé provou que a probabilidade do erro do tipo I para cada um dos testes não ultrapassa α.

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, E70, EAW, M1M, MAW.

- Hipóteses de interesse:

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

- Implicam nos seguintes constrastes

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

- 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

$$S_{\delta_1} = 0,061 \;\; e \;\; S_{\delta_2} = 0,028.$$

- 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:

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

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

- As estimativas dos constrastes 3 e 4 são dadas por:

$$\delta_3 = -0,027 \quad e \quad \delta_4 = -0,253,$$

e os respectivos erros-padrão, são dados por

$$S_{\delta_3} = 0,016 \quad e \quad S_{\delta_4} = 0,016.$$

- 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:

$$S_{0,05,3} = 0,054 \quad e \quad S_{0,05,4} = 0,054.$$

- Portanto, $|\delta_3| < 0,054$ e $|\delta_4| > 0,054$. Assim, rejeita-se a hipótese para o contraste $\boldsymbol{C}_4$ e não rejeita-se a hipótese para o contraste $\boldsymbol{C}_3$.

Comandos no R

```
> library(DescTools)
> C1 <- cbind(2,2,2,-3,-3); C2 <- cbind(1,-2,1,0,0)
> C3 <- cbind(1,0,-1,0,0); C4 <- cbind(0,0,0,1,-1)
> ScheffeTest(aov(mabsor~solvfac), contrasts=t(rbind(C1,C2,C3,C4)))
  Posthoc multiple comparisons of means :
    Scheffe Test 95% family-wise confidence level
$solvfac
                            diff       lwr.ci       upr.ci    pval
E50,E70,EAW-M1M,MAW      1.48896   1.27982134   1.69809866 1.8e-14 ***
E50,EAW-E70             -0.10956  -0.20308965  -0.01603035  0.0163 *
E50-EAW                 -0.02752  -0.08151937   0.02647937  0.5731
M1M-MAW                 -0.25288  -0.30687937  -0.19888063 4.8e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

Comandos no R

```
> library("agricolae")
> rsch<-scheffe.test(ranova,"solvfac",group=TRUE)
> rsch$groups

   trt    means  M
1  E70  0.60788  a
2  EAW  0.56686 ab
3  E50  0.53934  b
4  MAW  0.44964  c
5  M1M  0.19676  d
```

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
$$Q = \frac{max\{T_i\} - min\{T_i\}}{\sqrt{QMR/n}} = \frac{\overline{Y}_{max} - \overline{Y}_{min}}{\sqrt{QMR/n}} \qquad (1)$$
em que $T_i = \overline{Y}_i - (\mu + \alpha_i)$, 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.

- Rejeita-se H_0, para um dado α, se

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

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

$$T_\alpha = w_T \sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

com

$$w_T = \frac{q_\alpha(k, f)}{\sqrt{2}},$$

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

- **Obs.:** Os valores de $q_\alpha(k, f)$ são tabelados.

- Assim, um conjunto de intervalos de confiança simultâneos de $100(1-\alpha)\%$ para todas as diferenças de pares $\mu_i - \mu_j$, $i \neq j$, é dado por

$$IC[\mu_i - \mu_j; 1-\alpha] = \left[(\overline{Y}_i - \overline{Y}_j) - w_T \sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}; \right.$$
$$\left. (\overline{Y}_i - \overline{Y}_j) + w_T \sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \right].$$

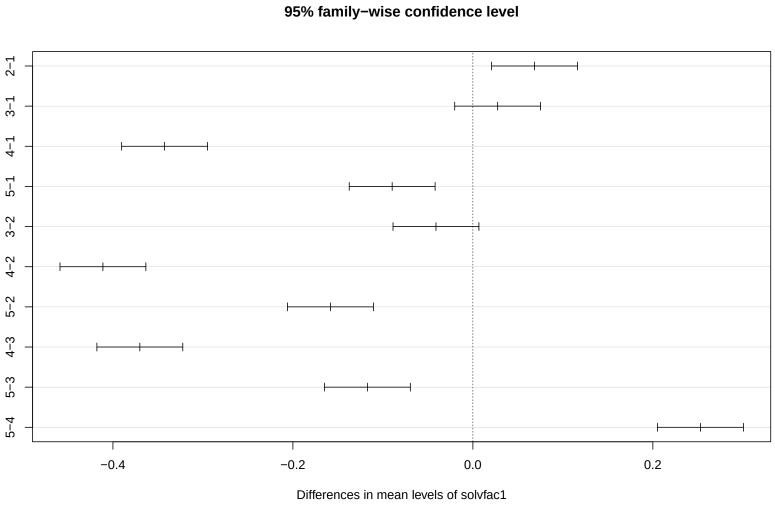
Resultado da aplicação do teste de Tukey

	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

Comandos no R

```
> rtuk1
Tukey multiple comparisons of means 95% family-wise confidence level
Fit: aov(formula = mabsor ~ solvfac1)
$solvfac1
        diff         lwr          upr     p adj
2-1  0.06854  0.02081671  0.116263292 0.0028819
3-1  0.02752 -0.02020329  0.075243292 0.4416646
4-1 -0.34258 -0.39030329 -0.294856708 0.0000000
5-1 -0.08970 -0.13742329 -0.041976708 0.0001465
3-2 -0.04102 -0.08874329  0.006703292 0.1141101
4-2 -0.41112 -0.45884329 -0.363396708 0.0000000
5-2 -0.15824 -0.20596329 -0.110516708 0.0000000
4-3 -0.37010 -0.41782329 -0.322376708 0.0000000
5-3 -0.11722 -0.16494329 -0.069496708 0.0000038
5-4  0.25288  0.20515671  0.300603292 0.0000000
```

```
> plot(rtuk1)
```



95% family−wise confidence level
2−1
3−1
4−1
5−1
3−2
4−2
5−2
4−3
5−3
5−4
−0.4
−0.2
0.0
0.2
Differences in mean levels of solvfac1

Comparação entre Bonferroni, Scheffé e Tukey

- Suponha que $k = 5$, $n = 35$ e $\alpha = 0.05$, e que apenas as 10 comparações entre os pares de médias ($\mu_i - \mu_j$, $i \neq j$) são de interesse para o pesquisador e estas foram especificamente selecionadas antes da pesquisa (ou seja, foram pré-planejadas). Se compararmos os coeficientes críticos para os três métodos, obtemos

$$\begin{array}{ll} \text{Bonferroni:} & w_B = t_{30;0,025/10} = 3,02; \\ \text{Scheffé:} & w_S = \sqrt{4\ F_{4;30;0.05}} = 3,28; \\ \text{Tukey:} & w_T = \frac{1}{\sqrt{2}}\ q_{5;30;0.05} = 2,91. \end{array}$$

- Uma vez que w_T é inferior a w_B, que é menor do que w_S para este exemplo, os intervalos de Tukey serão mais curtos que os intervalos de Bonferroni, que serão mais curtos que os intervalos de Scheffé.

Teste LSD de Fisher

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

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

- Rejeita-se H_0 se

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

em que $LSD = t_{(\alpha/2,n-k)}\sqrt{QMR\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$ e temos que $P(T > t_{(\alpha/2,n-k)}) = \alpha/2, T \sim t_{(n-k)}$.

- Se o teste F da ANOVA é significante a um nível α, cada par de média é então testado por um teste t de nível α.

- **Desvantagem:** controla apenas o nível α de cada teste, mas não controla o nível de significância geral.

- **Regra de decisão:** duas médias μ_i e μ_j são significativamente diferentes se $|\overline{Y}_i - \overline{Y}_j| > LSD$.

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 $r_\alpha(p, f),\ p = 2, 3, \cdots, k$, α é o nível de significância e f são os graus de liberdade do resíduo.

- Calcula-se $R_p = r_\alpha(p, f) S_{\overline{Y}_i},\ p = 2, 3, \cdots, k$.

- Compara-se, então a maior média com a menor (tal diferença é comparada com R_k).

- 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.

- 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.

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

- Neste caso, pode-se usar o pacote **multcomp**.

```
> rdunn<- glht(ranova, linfct=mcp(solvfac="Dunnett"))
> summary(rdunn)
 Simultaneous Tests for General Linear Hypotheses
 Multiple Comparisons of Means: Dunnett Contrasts
 Fit: aov(formula = mabsor ~ solvfac)
 Linear Hypotheses:
                   Estimate Std. Error t value  Pr(>|t|)
E70 - E50 == 0      0.06854    0.01595   4.298  0.00119 **
EAW - E50 == 0      0.02752    0.01595   1.726  0.27881
M1M - E50 == 0     -0.34258    0.01595 -21.481  < 0.001 ***
MAW - E50 == 0     -0.08970    0.01595  -5.624  < 0.001 ***
---
 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
 (Adjusted p values reported -- single-step method)
```

Seleção e comparação de modelos

Comparação dos modelos

- Vimos como verificar se um determinado modelo (normal-linear-homocedástico) se ajusta adequadamente aos dados.

- Uma outra questão de interesse surge quando se dispõe de diversos modelos (que se ajustam adequadamente aos dados) e respondem às perguntas de interesse, e queremos escolher um como o “mais apropriado”.

- Há diversas técnicas disponíveis para este fim.

- Veremos técnicas baseadas em testes de hipótese e comparação de estatísticas de ajuste.

Teste da razão de verossimilhanças

- Sejam M_1 e M_2 dois modelos, em que M_1 está encaixado em M_2, ou seja, o modelo M_1 é um caso particular de M_2.

- Por exemplo, M_1 é um modelo linear e M_2 é um modelo quadrático.

- Neste caso temos que

  H_0 : **o modelo M_1 é preferível ao modelo M_2.**

- Denote por $L_i(\widetilde{\boldsymbol{\theta}})$ e $l_i(\widetilde{\boldsymbol{\theta}})$ o máximo da verossimilhança e da log-verossimilhança do modelo i, respectivamente.

- $\widetilde{\boldsymbol{\theta}}$ corresponde à estimativa de máxima verossimilhança obtida sob o modelo i.

- A estatística do TRV é dada por $\Delta = \frac{L_1(\widehat{\boldsymbol{\theta}})}{L_2(\widehat{\boldsymbol{\theta}})}$.

- Rejeita-se H_0 se $\Delta < \delta_c$, em que δ_c é um valor critico adequado.

- Alternativamente, rejeitamos H_0 se

$$\Lambda = -2ln(\Delta) = -2\left(l_1(\widehat{\boldsymbol{\theta}}) - l_2(\widehat{\boldsymbol{\theta}})\right) > \lambda_c,$$

  em que $P(Q > \lambda_c) = \alpha$, $Q \sim \chi^2_{(\gamma)}$ e $\gamma =$ número de parâmetros do modelo M_2 - número de parâmetros do modelo M_2.

- $p - valor = P(Q > \lambda)$, em que λ é o valor observado da estatística Λ.

Estatísticas de comparação de modelos

- O TRV é apropriado na comparação de modelos encaixados.
- Além disso, ele não leva em consideração (diretamente) o número de parâmetros do modelo (somente na distribuição da estatística).
- Existem várias alternativas, em termos de estatística para comparar modelos, que "penalizam" a verossimilhança em relação ao número de parâmetros, tamanho da amostra entre outros fatores.
- Veremos o AIC e o BIC.

- O AIC e BIC são dados, respectivamente, por:

$$
\begin{aligned}
AIC &= -2l_i(\widetilde{\theta}) + 2k, \\
BIC &= -2l_i(\widetilde{\theta}) + 2k\ln(n),
\end{aligned}
$$

que $l_i(\widetilde{\theta})$ denota a log-verossimilhança do i-ésimo modelo avaliada em alguma estimativa, k é o número de parâmetros e n é o número de observações.

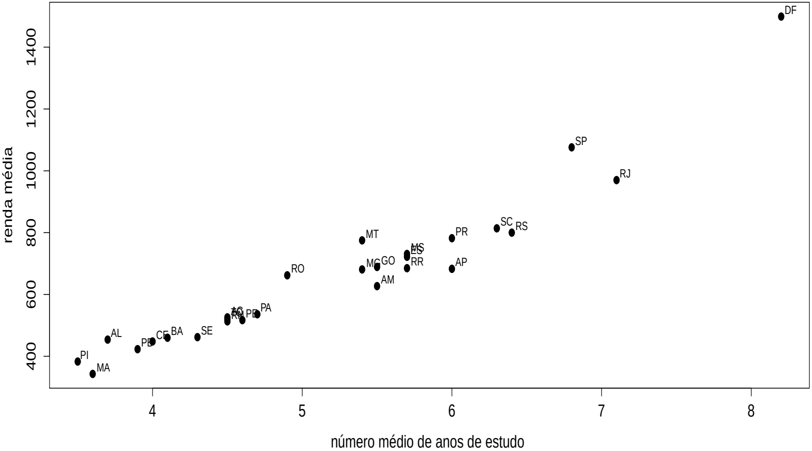
- Portanto, o modelo que apresentar os menores valores, será o modelo "melhor ajustado" aos dados.

Exemplo Censo

- O conjunto de dados em questão foi extraído do censo do IBGE de 2000 e apresenta para cada unidade da federação o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio.

- Um dos objetivos é estudar o relacionamento da renda média mensal em função do número médio de anos de estudo.

Dispersão entre anos de escolaridade e renda



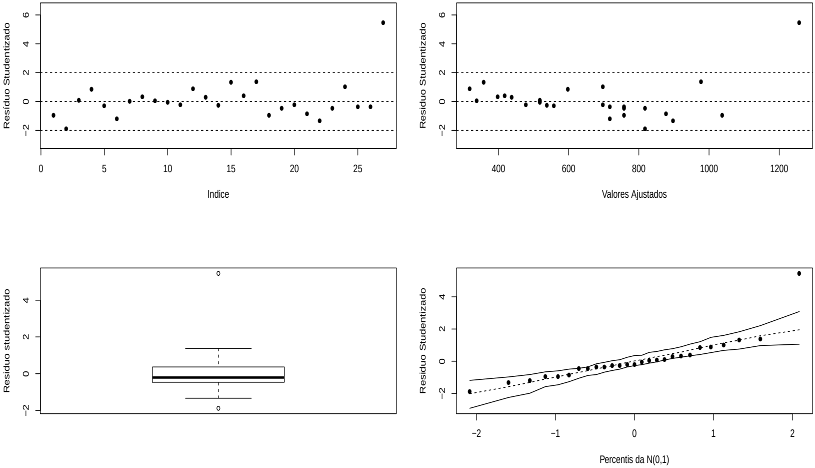
- Modelo 1: $Y_j = \beta_0 + \beta_1 x_j + \xi_j$

- Modelo 2: $Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \xi_j$

  em que

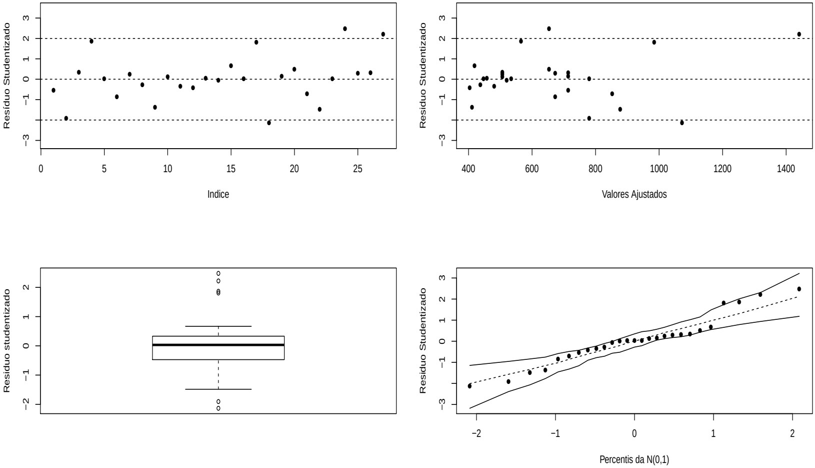
$$\xi_j \overset{ind.}{\sim} N(0, \sigma^2)$$

Modelo 1



Resíduo Studentizado
Indice
Resíduo Studentizado
Valores Ajustados
Resíduo studentizado
Resíduo Studentizado
Percentis da N(0,1)

Modelo 2



Resíduo Studentizado
Indice
Resíduo Studentizado
Valores Ajustados
Resíduo studentizado
Resíduo Studentizado
Percentis da N(0,1)

- Modelo 1

Parâmetro	Estimativa	EP	Estat. t	p-valor
β_0	-381,28	69,40	-5,49	$<$0,0001
β_1	199,83	13,03	15,34	$<$0,0001

- Modelo 2

Parâmetro	Estimativa	EP	Estat. t	p-valor
β_0	546,98	196,80	2,78	0,0104
β_1	-152,62	72,86	-2,09	0,0469
β_2	31,92	6,54	4,88	0,0001

- Estatísticas de comparação dos modelos

Estatística	Modelo 1	Modelo 2
AIC	315,26	298,66
BIC	319,15	303,85
log-verossim.	-154,63	-145,33

- TRV, estatísticas e pvalor entre parênteses: 18,80 ($< 0{,}0001$).

Modelos ajustados