

Introdução aos modelos de regressão múltipla e análise de dados: parte 2

Prof. Caio Azevedo

Exemplo 2: Estudo da eficácia de escovas de dentes

- Considere o seguinte estudo na área de Odontopediatria.
- O objetivo é comparar duas escovas de dente (convencional e experimental, chamada de “hugger”) com respeito à redução de um índice de placa bacteriana (IPB) em crianças de ambos os sexos em idade pré-escolar.
- Os valores obtidos correspondem aos IPB's medidos em alguns dentes antes e depois da escovação dental de 14 crianças do sexo feminino e 12 do sexo masculino. Cada criança utilizou cada um dos tipos de escova sendo sempre a experimental, a primeira. O tipo de escova tende a ser melhor quanto maior for sua “capacidade de remoção” da placa bacteriana.

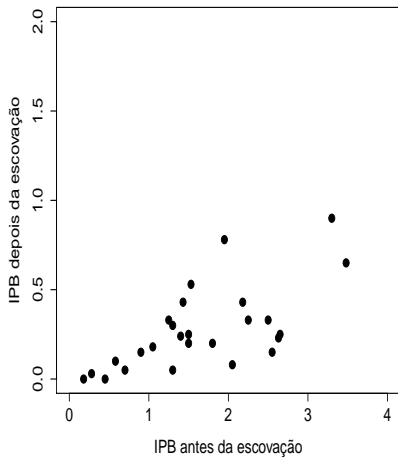
Dados

Criança	Tipo de escova				
	Sexo	Hugger		Convencional	
		Antes	Depois	Antes	Depois
1	F	2,18	0,43	1,2	0,75
2	F	2,05	0,08	1,43	0,55
⋮	⋮	⋮	⋮	⋮	⋮
25	M	1,3	0,05	2,73	0,85
26	M	2,65	0,25	3,43	0,88

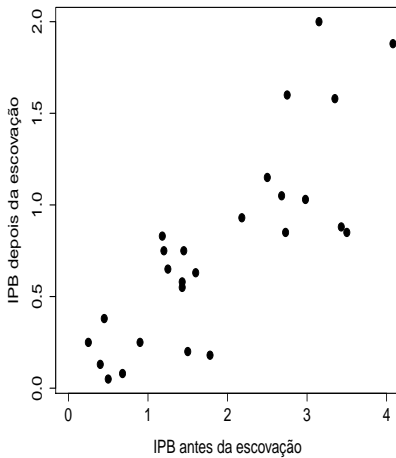
Voltando ao exemplo 2: Estudo da eficácia de escovas de dentes

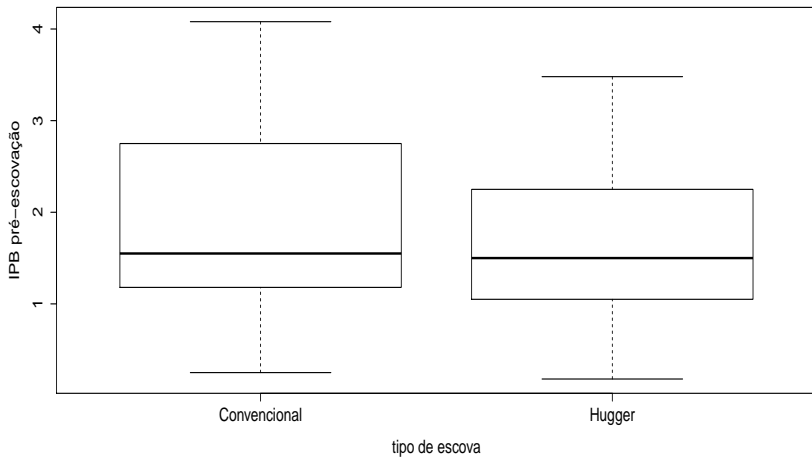
- Como utilizar os IPB's antes e depois ?
- Deve-se considerar a variável sexo?
- O fato de sempre se utilizar o tipo de escova experimental primeiramente pode ter influenciado os resultados?
- Medidas repetidas: cada criança é avaliada duas vezes. Possível dependência entre as observações.

Escova - Huger



Escova - Convencional





Comentário sobre os dados

- O modelo deve permitir, de forma simples, comparar o desempenhos dos dois tipos de escova, identificando e quantificando tal diferença, se ela existir.
- A variável resposta é positiva.
- O IPB pós-escovação deve ser menor do que o IPB pré-escovação. Em particular, se o IPB pré for igual a zero o IPB pós-escovação também deve ser igual a zero.
- Inicialmente, vamos desconsiderar a variável sexo.

Exemplo 2: desconsiderando o sexo

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij}, i = 1(\text{convencional}), 2(\text{hugger})(\text{tipo de escova}), \\ j = 1, \dots, 26(\text{criança}).$$

- $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_{ij} : IPB pré-escovação da criança j utilizando a escova do tipo i .
- Y_{ij} : IPB pós-escovação da criança j utilizando a escova do tipo i .
- β_{0i} : IPB pós-escovação esperado quando se utiliza a escova do tipo i para um IPB pré-escovação igual a 0.
- β_{1i} : incremento (positivo ou negativo) no IPB pós-escovação esperado quando se utiliza a escova do tipo i , para o aumento em uma unidade no IPB pré-escovação.

Exemplo 2: desconsiderando o sexo

- O que devemos esperar em relação às estimativas de $\beta_{0i}, i = 1, 2$?
- Como utilizar as estimativas de $\beta_{1i}, i = 1, 2$ para comparar o desempenho das escovas?

Estimativas dos parâmetros do modelo

Parâmetro	Est.	EP	IC(95%)	Estat.t	p-valor
β_{01} (convenc.)	0,013	0,101	[-0,190 ; 0,216]	0,128	0,8985
β_{02} (hugger)	-0,001	0,108	[-0,226 ; 0,207]	-0,089	0,9294
β_{11} (convenc.)	0,400	0,046	[0,307 ;0,493]	8,654	<0,0001
β_{12} (hugger)	0,174	0,058	[0,057 ; 0,291]	2,993	0,0044

Os dois interceptos são nulos e os dois coeficientes angulares parecem ser diferentes entre si. As estimativas dos coeficientes angulares sugerem uma superioridade da escova do tipo Hugger. Devemos ajustar um modelo reduzido sem interceptos.

Exemplo 2 (modelo reduzido): desconsiderando o sexo

$$Y_{ij} = \beta_{1i}x_{ij}, i = 1(\text{convencional}), 2(\text{hugger})(\text{tipo de escova});$$
$$j = 1, \dots, 26(\text{criança}).$$

- $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_{ij} : IPB pré-escovação da criança j utilizando a escova do tipo i .
- Y_{ij} : IPB pós-escovação da criança j utilizando a escova do tipo i .
- β_{1i} : diminuição (se $\beta_{1i} \in (0, 1)$) ou aumento (se $\beta_{1i} > 1$), no IPB quando se usa a escova do tipo i .

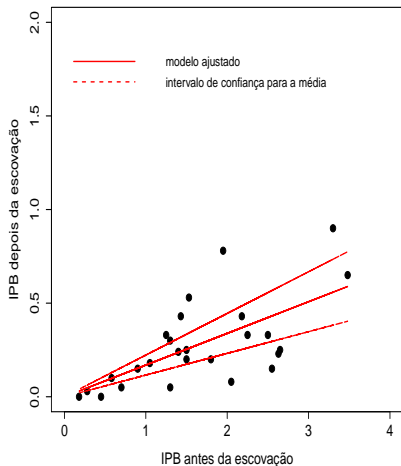
Estimativas dos parâmetros do modelo reduzido

Parâmetro	Est.	EP	IC(95%)	Estat.t	p-valor
β_{11} (convenc.)	0,405	0,023	[0,360 ; 0,450]	17,972	<0,0001
β_{12} (hugger)	0,169	0,027	[0,116 ; 0,223]	6,373	<0,0001

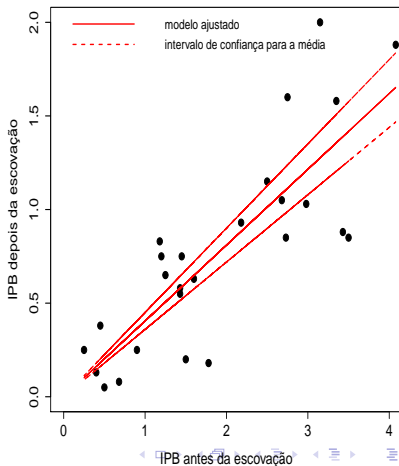
Os dois coeficientes angulares parecem ser diferentes. De fato, o teste do tipo Wald (veremos mais adiante) para testar igualdade vs diferença forneceu os seguintes resultados: 45,82 ($< 0,0001$). Os dois tipos de escova, de fato, reduzem o IPB, com evidente superioridade da escova do tipo Hugger, sendo a redução obtida para este tipo de escova da ordem de 16,9% [11,6%;22,3%].

Retas ajustadas e intervalos de confiança para as médias

Escova - Huger



Escova - Convencional



Estimativa da superioridade da escova do tipo Hugger

- Defina $\theta = \beta_{12}/\beta_{11}$: o quanto a escova do tipo Hugger reduz o IPB em relação à escova do tipo convencional.
- Estimador $\hat{\theta} = \hat{\beta}_{12}/\hat{\beta}_{11}$.
- Novamente, utilizaremos o método delta para fazer inferência sobre θ com base em $\hat{\theta}$.
- Método Delta: sob os resultados obtidos e assumindo válidas as condições de regularidade, temos que $\hat{\theta} \approx N(\theta, \sigma_{\theta}^2)$, em que

$$\sigma_{\theta}^2 = \sigma^2 (\mathbf{\Delta})' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{\Delta})$$

Estimativas da superioridade da escova do tipo Huger

- $\mathbf{\Delta} = \begin{bmatrix} -\frac{\beta_{12}}{\beta_{11}^2} & \frac{1}{\beta_{11}} \end{bmatrix}'$.
- Logo $IC(\theta, \gamma) \approx \left[\hat{\theta} - z_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_\theta^2}; \hat{\theta} + z_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_\theta^2} \right]$, em que $P(Z \leq z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $Z \approx N(0, 1)$, $\hat{\sigma}_\theta^2 = \hat{\sigma}^2 \left(\hat{\mathbf{\Delta}} \right)' (\mathbf{X}'\mathbf{X})^{-1} \left(\hat{\mathbf{\Delta}} \right)$ e $\hat{\mathbf{\Delta}} = \begin{bmatrix} -\frac{\hat{\beta}_{12}}{2\hat{\beta}_{11}^2} & \frac{1}{\hat{\beta}_{11}} \end{bmatrix}'$.

Estimativas da superioridade da escova do tipo Hugger

- Em nosso exemplo $\tilde{\theta} = 0,417(0,070)$, $IC(\theta, 95\%) = [0,282; 0,554]$.
Conclusão, espera-se que o IPB pós escovação quando se usa a escova do tipo Hugger seja da ordem de 41,7%[28,2%;55,4%] do IPB em relação a se usar a escova do tipo Convencional.
- Exercício: obter as estimativas pontual e intervalar de θ , utilizando $\hat{\theta}$ via bootstrap não-paramétrico. Obter também uma aproximação empírica da distribuição de $\hat{\theta}$.

ANOVA para modelos de regressão com intercepto nulo

- Suponha o seguinte modelo:

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \xi_i, \xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$$

- Logo $Y_i \stackrel{ind.}{\sim} N(\sum_{j=1}^{p-1} \beta_j x_{ji}, \sigma^2)$.
- O modelo acima define uma média (condicional aos valores de $x_{ji}, j = 1, \dots, p - 1; i = 1, \dots, n$) para cada observação Y_i .
- Defina $\hat{Y}_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_p x_{pi}$ (valor predito pelo modelo).
- O resíduo é definido por $R_i = \hat{\xi}_i = Y_i - \hat{Y}_i$.

- Nosso objetivo é considerar um modelo que explique adequadamente a variabilidade dos dados, ou seja, um modelo para o qual os resíduos sejam “pequenos”.
- Pode-se provar que a soma de quadrados total $SQT = \sum_{i=1}^n Y_i^2$, pode ser decomposta como:

$$SQT = \underbrace{\sum_{i=1}^n \hat{Y}_i^2}_{SQM} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{SQR}$$

- Assim, quanto maior for o valor de SQM em relação à SQR, maior será a contribuição da parte sistemática para explicar a variabilidade dos dados. Portanto, mais “provável” que exista (pelo menos um) $\beta_j \neq 0, j = 1, \dots, p - 1$.

- Lembrando da forma matricial do modelo: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$, pode-se demonstrar que:
 - $\text{SQT} = \mathbf{Y}'\mathbf{I}\mathbf{Y} = \mathbf{Y}'\mathbf{Y}$.
 - $\text{SQM} = \mathbf{Y}'\mathbf{H}\mathbf{Y}$, em que $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (chamada de matriz de projeção ou matriz “hat”).
 - $\text{SQR} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$.
- Pode-se provar que as matrizes \mathbf{H} e $\mathbf{B} = \mathbf{I} - \mathbf{H}$ são ortogonais, ou seja, $\mathbf{H}\mathbf{B} = \mathbf{0}$.
- Dizemos que $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$ e $\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ projetam \mathbf{Y} em dois subespaços ortogonais, pois $\mathbf{H}(\mathbf{I} - \mathbf{H}) = \mathbf{0}$.
- Podemos utilizar raciocínio análogo ao que foi usado para o modelo com intercepto (Exercício).

Tabela de ANOVA (matricial)

- Para testar $H_0 : \beta_1 = \beta_2 = \dots = \beta_{(p-1)} = 0$ vs $H_1 : \text{Há pelo menos uma diferença.}$

FV	SQ	GL	QM	Estatística F	pvalor
Modelo	$SQM = \mathbf{Y}'\mathbf{H}\mathbf{Y}$	p	$QMM = \frac{SQM}{p}$	$F_t = \frac{QMM}{QMR}$	$P(X \geq f_t H_0)$
Resíduo	$SQR = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$	n-p	$QMR = \frac{SQR}{n-p}$		
Total	SQT	n			

FV: fonte de variação, SQ: soma de quadrados, Gl: graus de liberdade, QM: quadrado médio. Portanto, rejeita-se H_0 se $f_t \geq f_c$ ou, analogamente, se $p - \text{valor} = P(X \geq f_t | H_0) \leq \alpha$, em que f_t é o valor calculado da estatística F_t e $P(X \geq f_c | H_0) = \alpha$, $X \sim F_{(p, n-p)}$.

Anova para modelos sem intercepto: Exemplo 2 (sem intercepto)

FV	GL	SQ	QM	Estat. F	p-valor
Modelo	2	22,97	11,48	181,80	<0,0001
Resíduo	50	3,16	0,06		

Conclui-se que pelo menos um dos coeficientes β_{2i} , $i = 1, 2$ é diferente de zero. Ou seja, o IPB pré-escovação influencia o IPB pós-escovação para pelo menos um tipo de escova.

Previsão para uma única observação

- Já vimos como estimar pontual e intervalarmente a média $\mu_i = \mathcal{E}(Y_i) = \mathbf{X}'_i \boldsymbol{\beta}$ em que \mathbf{X}'_i é a i -ésima linha da matriz \mathbf{X} .
- Para isso usamos $\hat{\mu}_i = \mathbf{X}'_i \hat{\boldsymbol{\beta}}$. Note que $\mathcal{E}(\hat{\mu}_i) = \mathbf{X}'_i \mathcal{E}(\hat{\boldsymbol{\beta}}) = \mathbf{X}'_i \boldsymbol{\beta} = \mu_i$. Como medida de precisão adotamos o erro quadrático médio (EQM):

$$\begin{aligned}\mathcal{E}[(\hat{\mu}_i - \mu_i)^2] &= \mathcal{E}[(\hat{\mu}_i - \mathcal{E}(\hat{\mu}_i))^2] + [\mathcal{E}(\hat{\mu}_i) - \mu_i]^2 \\ &= \mathcal{V}(\hat{\mu}_i) + 0 = \mathcal{V}(\hat{\mu}_i)\end{aligned}$$

- Considere agora que temos o interesse em prever o valor de uma única observação, que possui um determinado perfil em termos de suas covariáveis \mathbf{X}'_h . Denotaremos-na por Y_h . Note que tal observação não foi utilizada para se obter $\hat{\boldsymbol{\beta}}$.

Previsão para uma única observação

- Note ainda que Y_h é uma variável aleatória e não, à rigor, um parâmetro (valor “fixo”).
- Como estimador pontual adotaremos, novamente, $\hat{\mu}_h = \hat{Y}_h = \mathbf{X}'_h \hat{\boldsymbol{\beta}}$. Note que $\mathcal{E}(\hat{Y}_h - Y_h) = \mathcal{E}(\hat{Y}_h) - \mathcal{E}(Y_h) = \mu_h - \mu_h = 0$, em que $\mu_h = \mathbf{X}'_h \boldsymbol{\beta}$. Logo $\mathcal{E}[(\hat{Y}_h - Y_h)^2] = \mathcal{V}(\hat{Y}_h - Y_h)$.
- Como medida de precisão adotamos:

$$\begin{aligned}\mathcal{E}[(\hat{Y}_h - Y_h)^2] &= \mathcal{V}(\hat{Y}_h - Y_h) = \mathcal{V}(\hat{Y}_h) + \mathcal{V}(Y_h) - 2 \underbrace{\text{Cov}(\hat{Y}_h, Y_h)}_{\text{ind.}} \\ &= \mathcal{V}(\mathbf{X}'_h \hat{\boldsymbol{\beta}}) + \sigma^2 = \mathbf{X}'_h \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{X}_h + \sigma^2 \\ &= \sigma^2 \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h + \sigma^2 = \sigma^2 \left[\mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h + 1 \right]\end{aligned}$$

Previsão para uma única observação

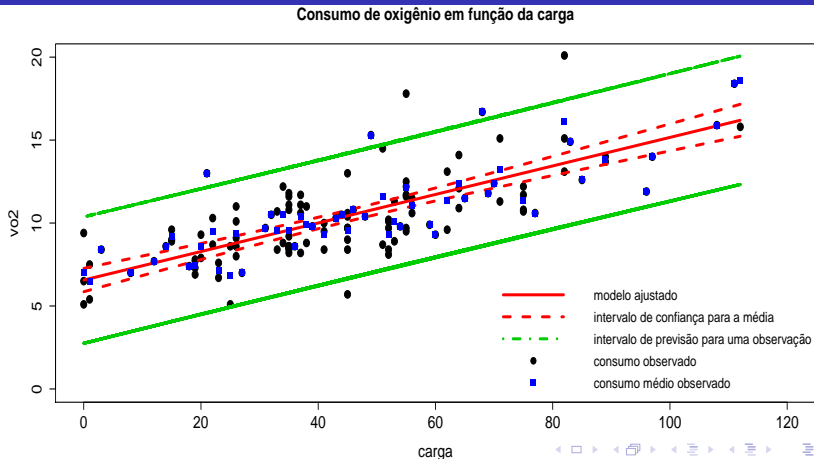
- Assim $\frac{\hat{Y}_h - Y_h}{\sqrt{\hat{\sigma}_{\hat{Y}_h}^2}} \sim N(0, 1)$.
- Pode-se provar que: $\frac{\hat{Y}_h - Y_h}{\sqrt{\hat{\sigma}_{\hat{Y}_h}^2}} \sim t_{(n-p)}$ em que $\hat{\sigma}_{\hat{Y}_h}^2 = \hat{\sigma}^2 [\mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h + 1]$ (exercício).
- Logo, $IP[Y_h; \gamma] = \left[\hat{Y}_h - t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_{\hat{Y}_h}^2}; \hat{Y}_h + t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_{\hat{Y}_h}^2} \right]$, em que $P(X \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $X \sim t_{(n-p)}$. (IP : Intervalo de Previsão).

Estimação da média e previsão para uma única observação

	Média (μ_i)	Observ. (Y_h)
Estimador	$\hat{\mu}_i = \mathbf{X}'_i \hat{\beta}$	$\hat{\mu}_h = \mathbf{X}'_h \hat{\beta}$
Distr. do estimador/preditor	$N(\mu_i, \sigma^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i)$	$N(\mu_h, \sigma^2 \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h)$
Medida de precisão	$\sigma^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i$	$\sigma^2 [1 + \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h]$
Quantidade pivotal	$\frac{\hat{\mu}_i - \mu_i}{\hat{\sigma} \sqrt{\mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i}}$	$\frac{\hat{\mu}_h - Y_h}{\hat{\sigma} \sqrt{1 + \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h}}$
$IC(\cdot, \gamma) / IP(\cdot, \gamma)$	$\hat{\mu}_i \pm t_{\frac{1+\gamma}{2}} \hat{\sigma} \sqrt{\mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i}$	$\hat{\mu}_h \pm t_{\frac{1+\gamma}{2}} \hat{\sigma} \sqrt{1 + \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h}$

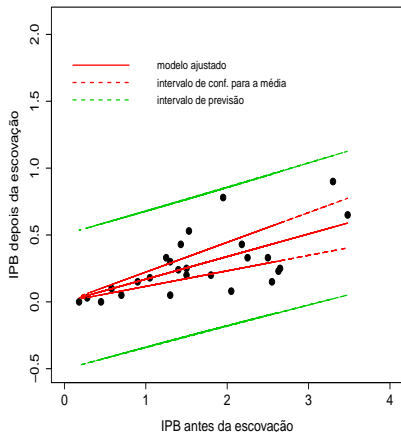
em que $P(X \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $X \sim t_{(n-p)}$

Exemplo 1: ajuste para o modelo final (sem o fator etiologia)

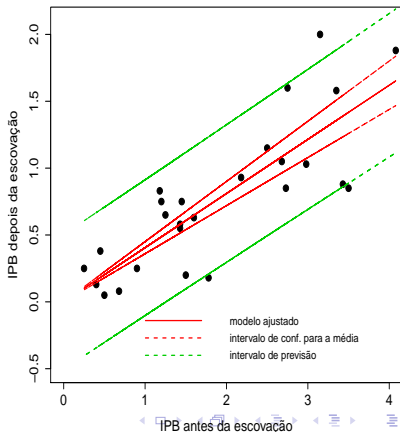


Exemplo 2: ajuste para o modelo final (sem considerar o fator “sexo”)

Escova - Huger



Escova - Covencional



Exemplo 3: ajuste para o modelo final (quadrático)

