Serviço Geológico do Brasil – CPRM

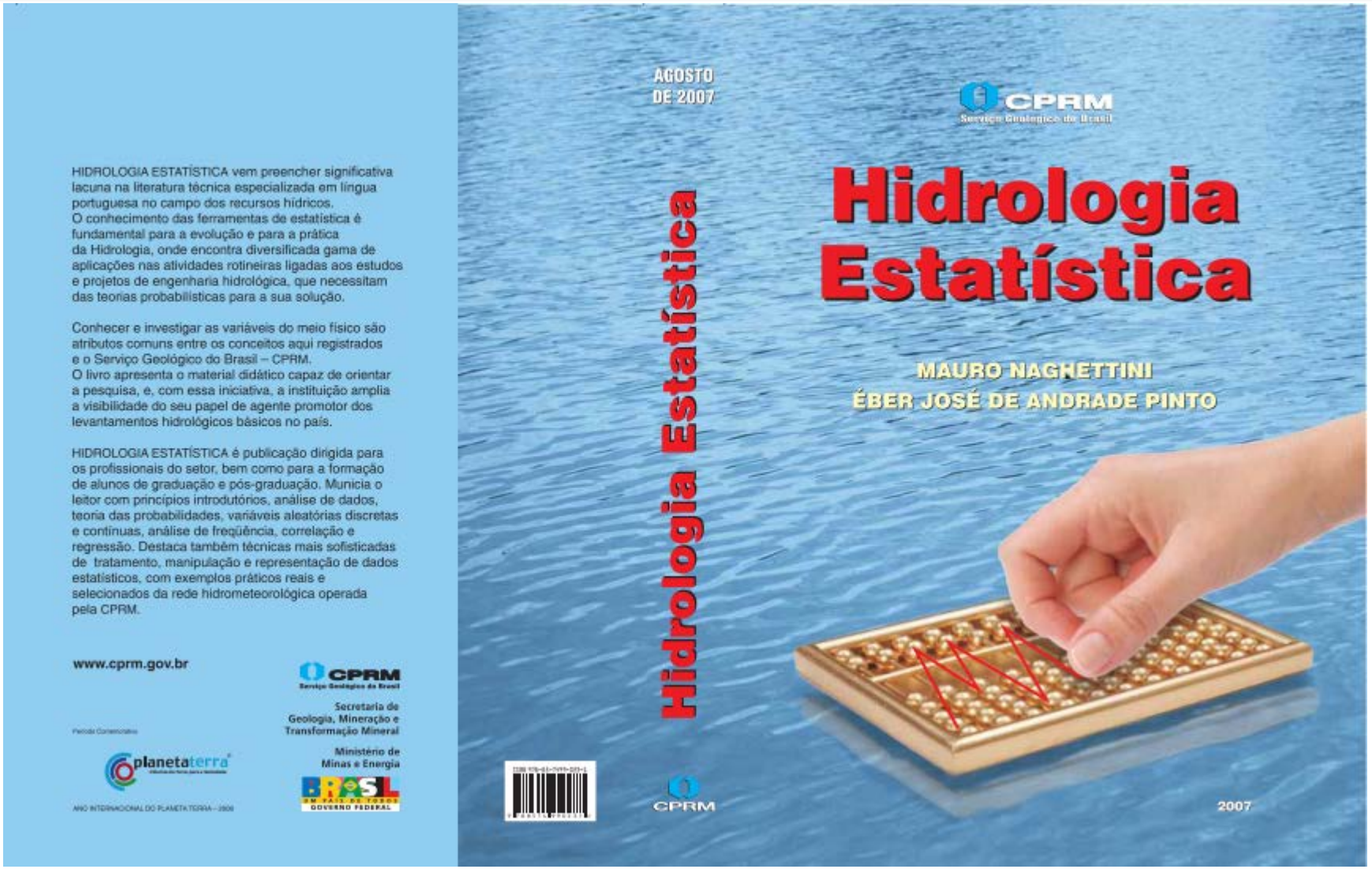
Correlação e Regressão Linear

Aula 03 : RLS

Eber José de Andrade Pinto
Coordenador Executivo do DEHID
eber.andrade@cprm.gov.br
www.cprm.gov.br

Belo Horizonte, 21 de outubro de 2020

Livro Texto



http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Livro-%22Hidrologia-Estatistica%22-981.html

Intervalos de Confiança para os Coeficientes da RLS

Variabilidade amostral → a reta de regressão estimada é uma das muitas retas possíveis.

Parâmetros *a* e *b* → estimadores pontuais dos parâmetros populacionais α e β.

$$a - t_{1-\frac{\alpha}{2},n-2}.s_a \leq \alpha \leq a + t_{1-\frac{\alpha}{2},n-2}.s_a \qquad s_a = \sqrt{s_e^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

$$b - t_{1-\frac{\alpha}{2},n-2}.s_b \leq \beta \leq b + t_{1-\frac{\alpha}{2},n-2}.s_b \qquad s_b = \sqrt{\frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

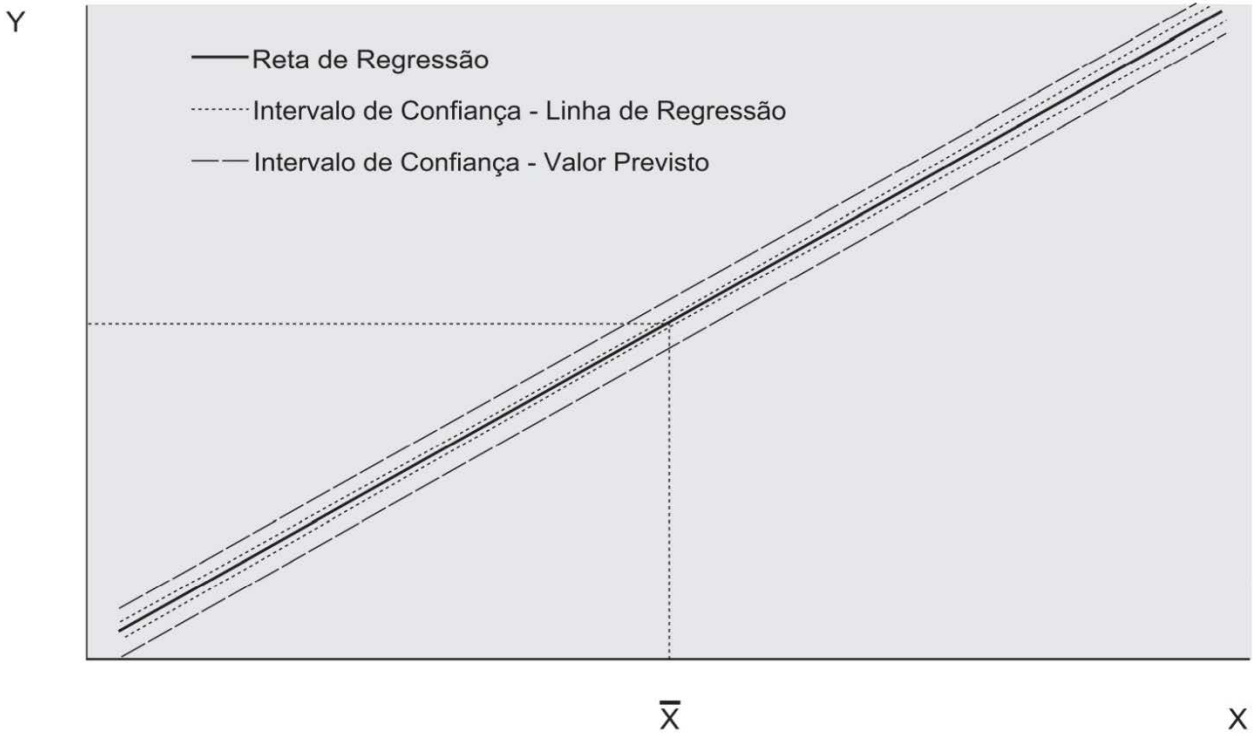
$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

$t_{1-\frac{\alpha}{2},n-2}$ é valor da distribuição *t* de Student, para um nível de significância α e (n – 2) graus de liberdade

Intervalos de Confiança para a Reta de RLS

$$y' = \alpha + \beta . x'$$

$$(a+bx') - t_{1-\frac{\alpha}{2},n-2} . s_e . \sqrt{\frac{1}{n} + \frac{(x'-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \leq y' \leq (a+bx') + t_{1-\frac{\alpha}{2},n-2} . s_e . \sqrt{\frac{1}{n} + \frac{(x'-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$



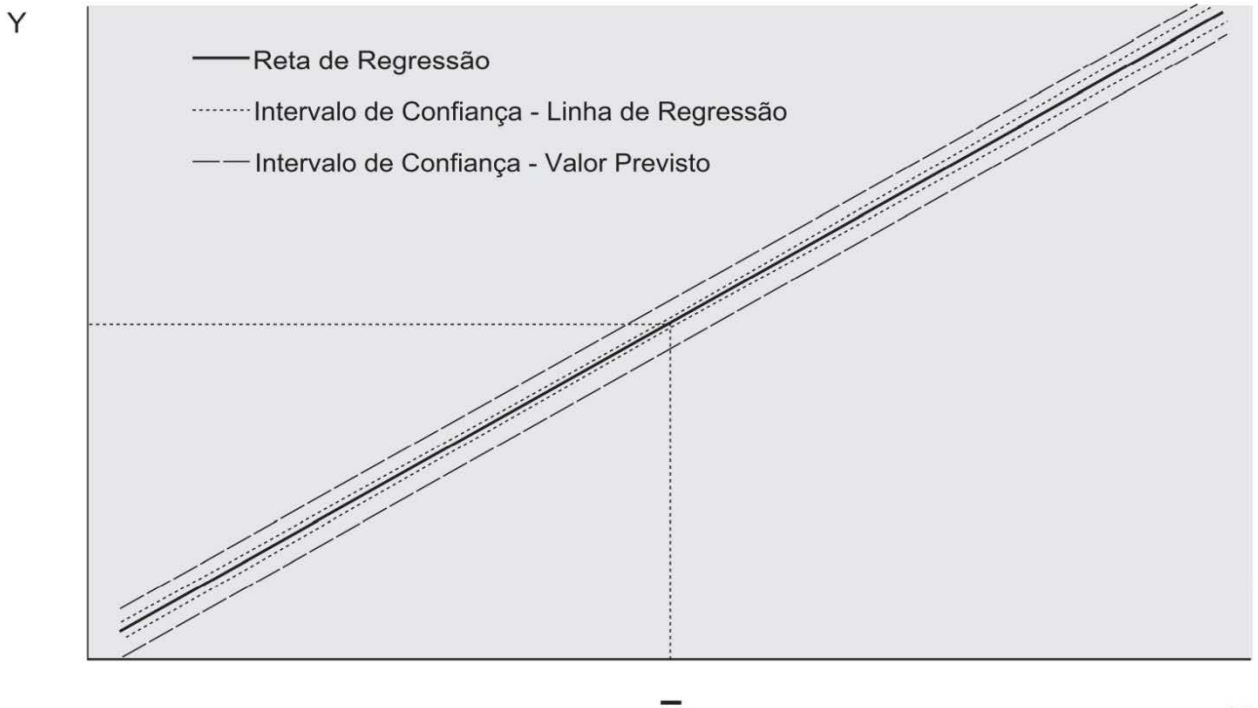
$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

$t_{1-\frac{\alpha}{2},n-2}$ é valor da distribuição *t* de Student, para um nível de significância α e (*n* – 2) graus de liberdade

Intervalos de Confiança para um Valor Previsto pela RLS

→ **acrescentar + 1 erro padrão da estimativa** ←

$$(a+bx') - t_{1-\frac{\alpha}{2},n-2}.s_e.\sqrt{1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \le \hat{y}' \le (a+bx') + t_{1-\frac{\alpha}{2},n-2}.s_e.\sqrt{1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$



$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

$t_{1-\frac{\alpha}{2},n-2}$ é valor da distribuição *t* de Student, para um nível de significância α e $(n-2)$ graus de liberdade

Avaliação da RLS

Linearidade → gráfico de dispersão e TH sobre o coeficiente angular β

Hipótese Nula: $H_0 : \beta = 0$ (não há relação linear)

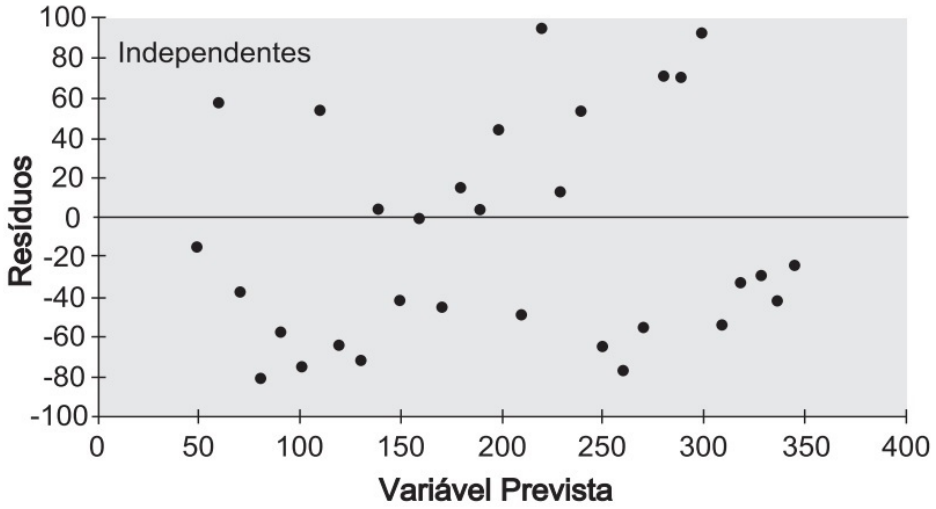
Hipótese Alternativa: $H_1 : \beta \neq 0$ (há relação linear)

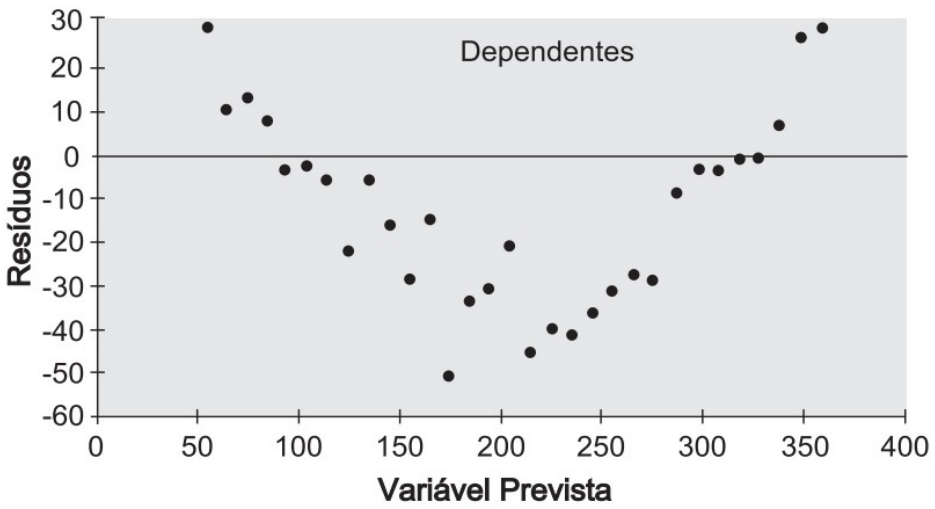
Estatística de Teste: $t = (b - \beta)/s_b$ ou, sob H_0, $t = b/s_b \sim t$ Student com $n - 2$ gl

Decisão: rejeitar H_0 se $|t| > t_{1-\alpha/2, n-2}$

Avaliação da RLS

Independência dos Resíduos (+ prescrição incorreta do modelo)



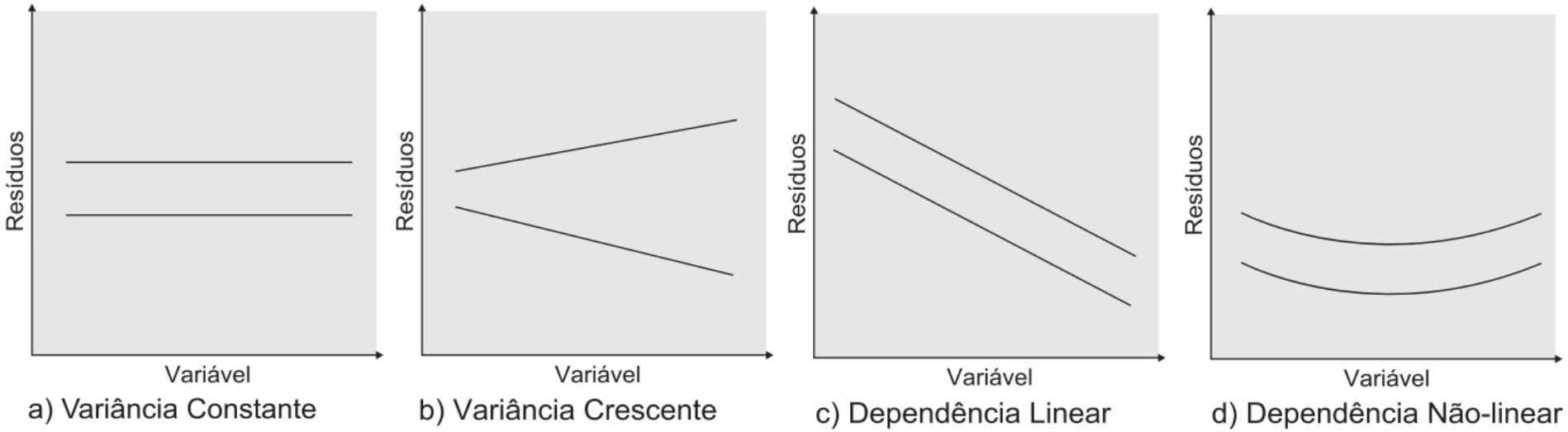


Avaliação da RLS

Normalidade dos Resíduos → testes de aderência, papel de probabilidade

Resíduos com Média Nula → OK pela estimação pelo MMQ

Resíduos Homocedásticos (Variância Constante) →

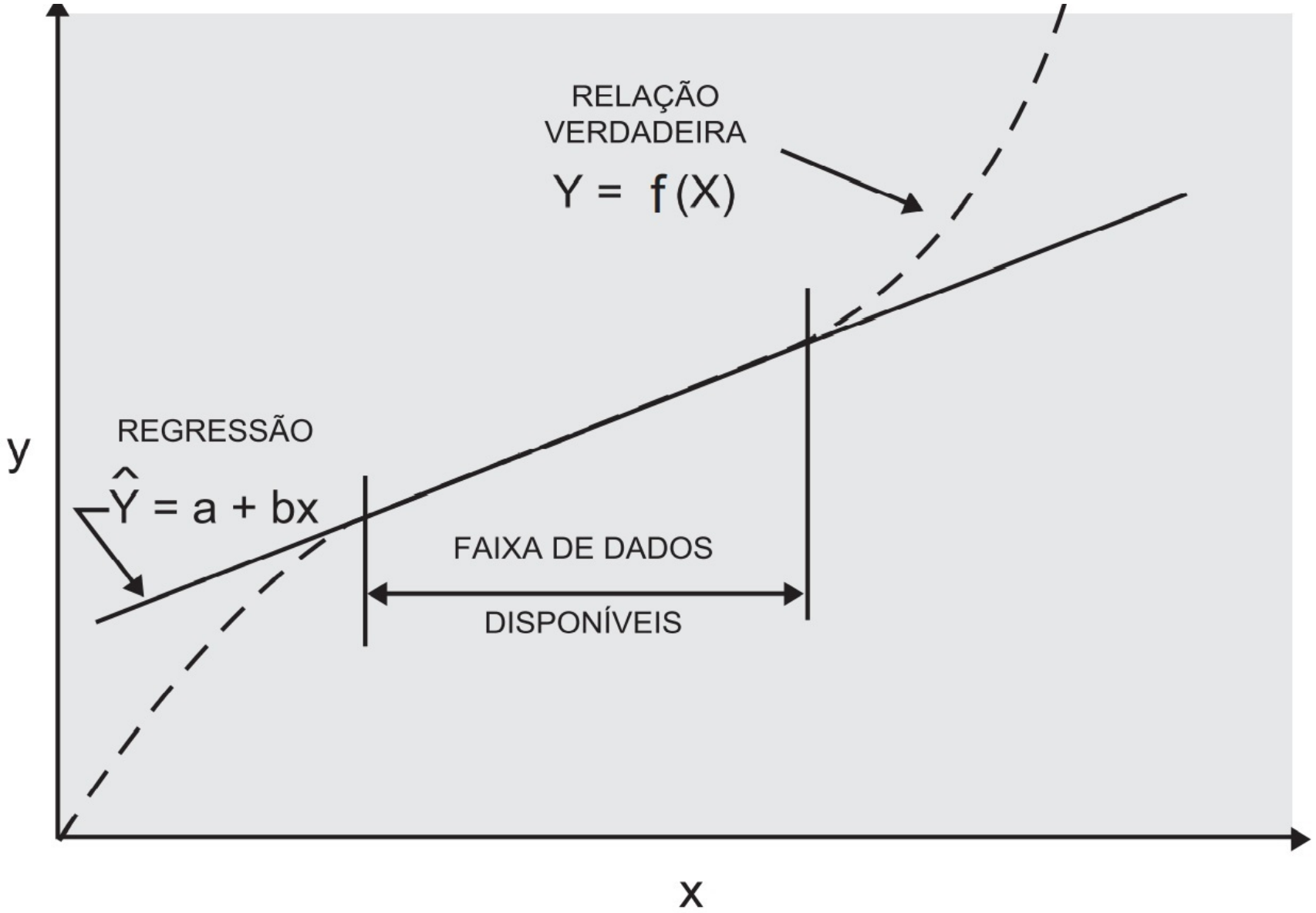


Hipóteses a serem atendidas para alcançar os objetivos da RLS por MMQ

Hipótese	Objetivos			
	Prever y dado x	Prever y e uma variância para a previsão	Obter o melhor estimador linear não enviesado de y	Testar hipóteses, estimar intervalos de confiança ou previsão
A forma do modelo está correta: y está linearmente relacionado a x	X	X	X	X
Os dados usados para ajustar o modelo são representativos dos dados de interesse	X	X	X	X
A variância dos resíduos é constante (homocedástico). Não depende de x ou de qualquer outra variável, como o tempo.		X	X	X
Os resíduos são independentes de x.			X	X
Os resíduos são normalmente distribuídos.				X

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical Methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.] Página 228

Cuidado com a Extrapolação!



Regressão Não-Linear Simples

Modelos Polinomiais (e.g: parabólico → $Y = a + bX + cX^2$ **)**

Equações Normais de Regressão

$$\begin{cases} \sum_i Y_i = na + b\sum_i X_i + c\sum_i X_i^2 \\ \sum_i X_i Y_i = a\sum_i X_i + b\sum_i X_i^2 + c\sum_i X_i^3 \\ \sum_i X_i^2 Y_i = a\sum_i X_i^2 + b\sum_i X_i^3 + c\sum_i X_i^4 \end{cases}$$

Modelos Linearizáveis

$$y = ax^b \qquad ln\, y = ln\left(ax^b\right) \qquad ln\, y = ln\, a + ln\left(x^b\right)$$

$$ln\, y = ln\, a + b\, ln\, x$$

$$Z = ln\, y \qquad k = ln\, a \qquad V = ln\, x$$

$$Z = k + b.V$$

Ver Anexo 10

Importante: ver solução de exemplo completo de regressão Exemplo 9.1 pg. 375 HE

O procedimento para análise da RLS:

Etapa 1 Selecione a variável preditora (X) que está relacionada à variável a ser prevista (Y) por alguma relação física.

Etapa 2 Plote a variável preditora (X) em relação à variável a ser prevista (Y)

Etapa 3 Determine a forma da equação desejada; isto é, linear ou curvilíneo.

Etapa 4 Calcule o coeficiente de correlação entre as variáveis.

Etapa 5 Calcule os coeficientes de regressão.

No EXCEL: Função PROJ.LIN(), PROJ.LOG() e na ferramenta Análise de Dados/Regressão

Etapa 6 Calcule o erro padrão da estimativa, Se; desvio padrão da variável a ser prevista, Sy; e o coeficiente de determinação, r^2.

O procedimento para análise da RLS:

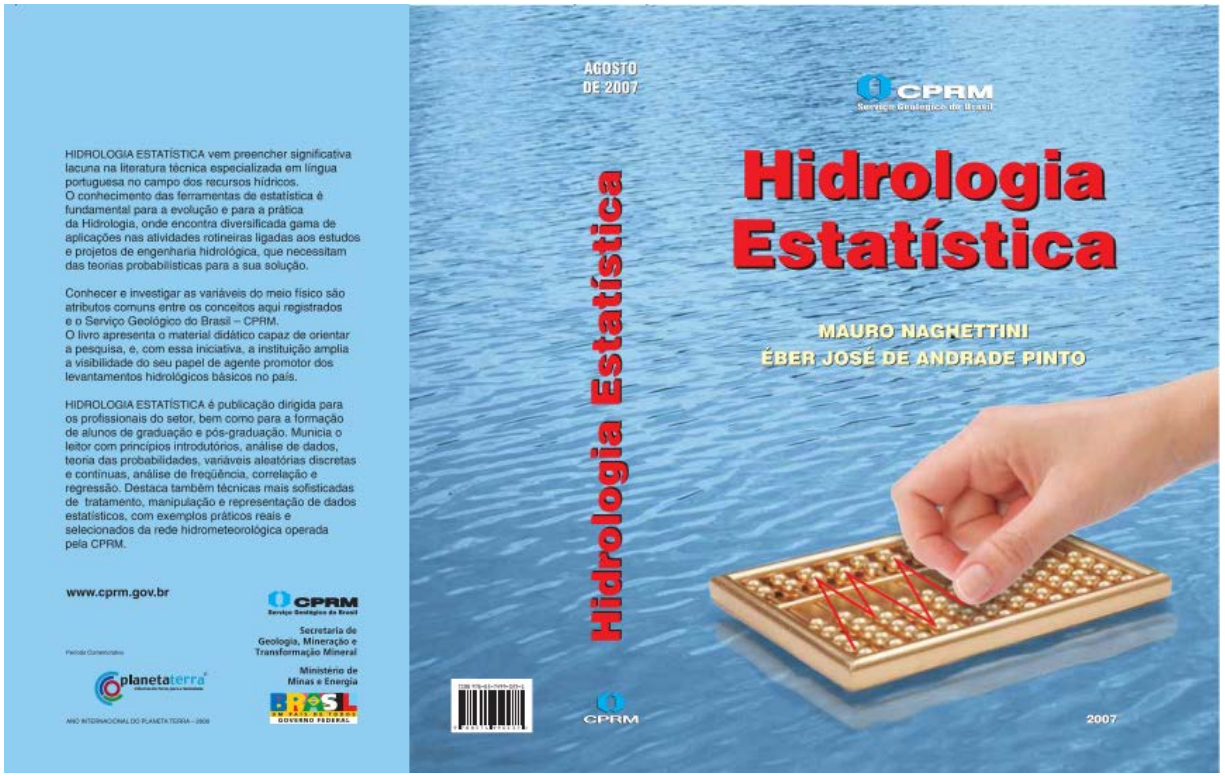
Etapa 7 Avalie a equação de regressão pelos seguintes métodos:

- O erro padrão da estimativa tem os limites $0 \leq Se \leq Sy$; se Se -> 0 maior parte da variância é explicada pela regressão.
- Coeficiente de determinação tem limites $0 \leq r^2 \leq 1$; quando r^2 -> 1, melhor será o “ajuste” da linha de regressão aos dados.
- Examine os resíduos para identificar deficiências na equação de regressão e verifique as suposições do modelo.

Etapa 8 Se a precisão da equação de regressão não for aceitável, reformule a equação de regressão ou transforme as variáveis. Uma solução satisfatória nem sempre é possível a partir dos dados disponíveis.

Recomendações

Para consolidar conhecimentos estudar no livro texto os itens 9.5, 9.6 e 9.7



http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Livro-%22Hidrologia-Estatistica%22-981.html

Serviço Geológico do Brasil – CPRM

Departamento de Hidrologia da CPRM

Eber José de Andrade Pinto
Coordenador Executivo do DEHID
eber.andrade@cprm.gov.br
www.cprm.gov.br

Belo Horizonte, 21 de outubro de 2020