Serviço Geológico do Brasil – CPRM

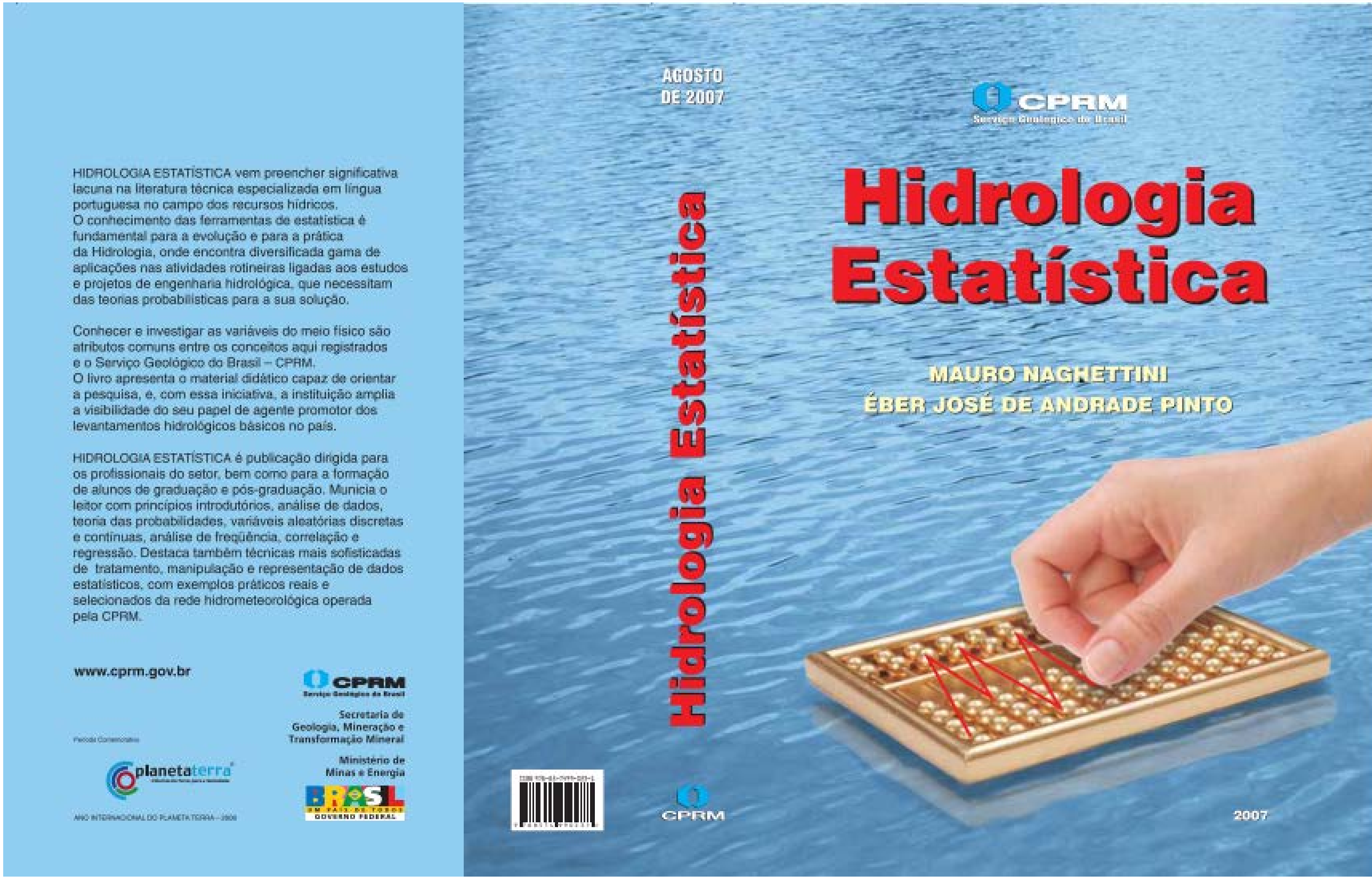
Correlação e Regressão Linear

Aula 06 : RLM

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

Belo Horizonte, 22 de outubro de 2020

Livro Texto



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

O procedimento para análise da RLM:

Etapa 1 Selecione as variáveis preditoras (Xi) que estão relacionadas à variável a ser prevista (Y) por alguma relação física.

Etapa 2 Plote as variáveis preditoras (Xi) em relação à variável a ser prevista (Y)

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

Etapa 4 Calcule os coeficientes de correlação entre as variáveis. Matriz de correlação.

Etapa 5 Calcule os coeficientes de regressão.

No EXCEL: Função PROJ.LIN(), PROJ.LOG() e a 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 r^2 parcial.

O procedimento para análise da RLM:

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.
- Os testes F parciais e totais são usados para avaliar cada preditor e a significância total da equação.
- O sinal de cada coeficiente de regressão deve ser comparado com o coeficiente de correlação para o critério de predição apropriado. Os sinais devem ser os mesmos.
- Examine os resíduos para identificar deficiências na equação de regressão e verifique as suposições do modelo.

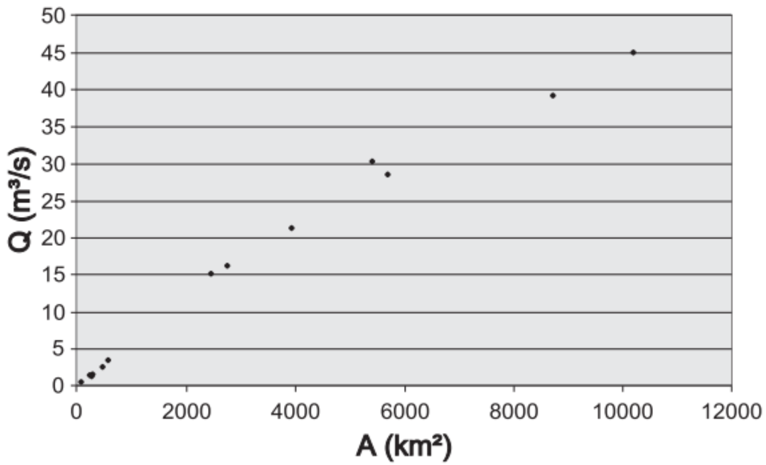
Etapa 8 Se a precisão da equação de regressão não for aceitável, reformule a equação de regressão ou transforme as variáveis. Uma solução satisfatória nem sempre é possível a partir dos dados disponíveis.

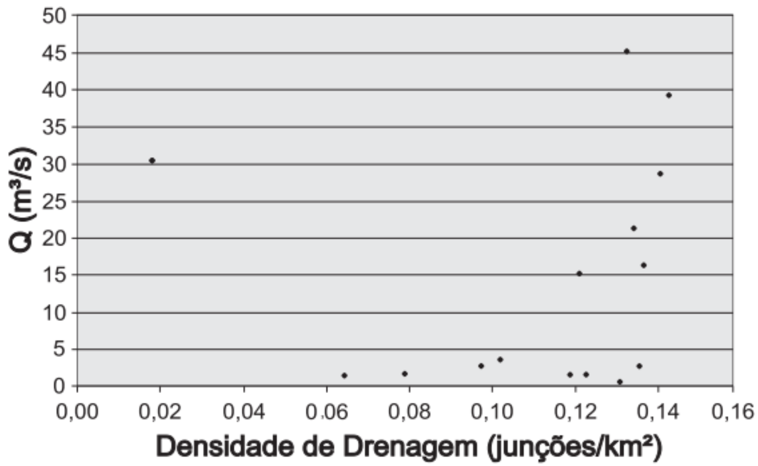
EXEMPLO 9.2

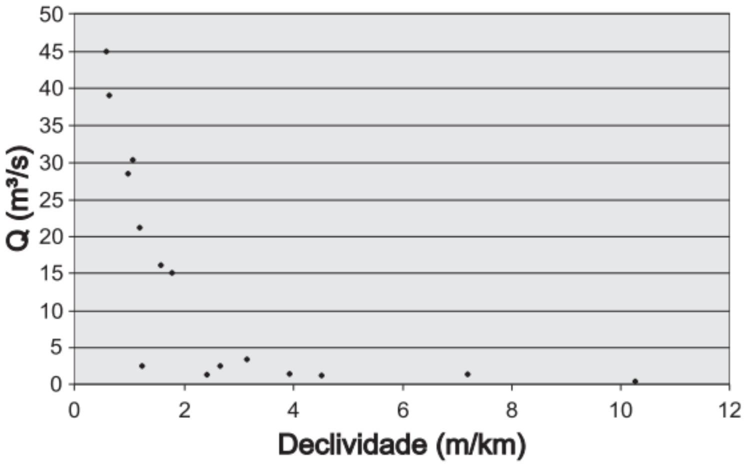
Em um estudo de regionalização de vazões mínimas com 7 dias de duração na bacia do rio Paraopeba, no qual foi aplicado o método "*index-flood*", definiu-se uma região homogênea com 15 estações fluviométricas. Nesse estudo, as médias das vazões mínimas anuais com 7 dias de duração foram utilizadas como fator de adimensionalização das séries. Defina um modelo de regressão que permita a estimativa do fator "*index-flood*" em locais que não possuam estações fluviométricas utilizando as variáveis explicativas da tabela abaixo.

Estação	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Qmin méd (m^3/s)	2,6	1,49	1,43	3,44	1,37	2,53	15,12	16,21
Área (Km^2)	461	291	244	579	293	486	2465	2760
I equiv (m/km)	2,69	3,94	7,20	3,18	2,44	1,25	1,81	1,59
DD (Junções/Km^2)	0,098	0,079	0,119	0,102	0,123	0,136	0,121	0,137
Estação	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Qmin méd (m^3/s)	21,16	30,26	28,53	1,33	0,43	39,12	45	
Área (Km^2)	3939	5414	5680	273	84	8734	10192	
I equiv (m/km)	1,21	1,08	1,00	4,52	10,27	0,66	0,60	
DD (Junções/Km^2)	0,134	0,018	0,141	0,064	0,131	0,143	0,133	

Verificação de Linearidades com *Y*







Matriz dos coeficientes de correlação simples

	Qmin méd (m^3/s)	Área (Km^2)	I equiv (m/km)	DD (Junções/Km^2)
Qmin méd (m^3/s)	1			
Área (Km^2)	0,992	1		
I equiv (m/km)	-0,625	-0,594	1	
DD (Junções/Km^2)	0,141	0,186	-0,049	1

Multicolinearidade? Aparentemente não!

MODELO

$$Q = \beta_0 . A^{\beta_1} . X_2^{\beta_2} . X_3^{\beta_3}$$

$$ln\,Q = ln\,\beta_0 + \beta_1\, ln\,A + \beta_2\, ln\,X_2 + \beta_3\, ln\,X_3$$

SELEÇÃO DO MODELO

Modelo 1 $Q = \beta_0 . A^{\beta_1}$ $\ln Q = \ln \beta_0 + \beta_1 \ln A$

	gl	SQ	MQ	F
Regressão	1	33,04321	33,04321	2915,798
Resíduo	13	0,147322	0,011332	
Total	14	33,19053		

$$(F = 2916) > [F(0{,}05;1;13) = 4{,}67]$$

modelo é significativo para α=5%

Modelo 2 $Q = \beta_0 . A^{\beta_1} . I^{\beta_2}$ $\ln Q = \ln \beta_0 + \beta_1 \ln A + \beta_2 \ln I$

F_{total} $\longrightarrow$ modelo é significativo

$F_{parcial}$ $\longrightarrow$ $(F_P = 3{,}04) < [F(0{,}05;1;12) = 4{,}75]$ $\longrightarrow$ a inclusão de I não é significativa

Modelo 3 $Q = \beta_0 . A^{\beta_1} . DD^{\beta_2}$ $\ln Q = \ln \beta_0 + \beta_1 \ln A + \beta_2 \ln DD$

F_{total} $\longrightarrow$ modelo é significativo

$F_{parcial}$ $\longrightarrow$ $(F_P = 0{,}40) < [F(0{,}05;1;12) = 4{,}75]$ $\longrightarrow$ a inclusão de DD não é significativa

Modelo 4 $Q = \beta_0 . A^{\beta_1} . I^{\beta_2} DD^{\beta_3}$ $\ln Q = \ln \beta_0 + \beta_1 \ln A + \beta_2 \ln I + \beta_3 \ln DD$

F_{total} $\longrightarrow$ modelo é significativo

testes anteriores de $F_{parcial}$ $\longrightarrow$ modelo tem excesso de variáveis explicativas

COEFICIENTES DE REGRESSÃO

$$Q = \beta_0 . A^{\beta_1} . X_2^{\beta_2} . X_3^{\beta_3}$$

$$\ln Q = \ln \beta_0 + \beta_1 \ln A + \beta_2 \ln X_2 + \beta_3 \ln X_3$$

Modelo	$ln(\beta_0)$	β_1	β_3	β_2	r^2	Erro Padrão
QA	-5,1696	0,9889			**0,9956**	**0,1065**
QAI	-5,7309	1,0551	0,1344		**0,9965**	**0,0990**
QADD	-5,24512	0,9884	-0,0348		**0,9957**	**0,1090**
QAIDD	-5,7579	1,05224	0,12930	-0,0223	**0,9965**	**0,1025**

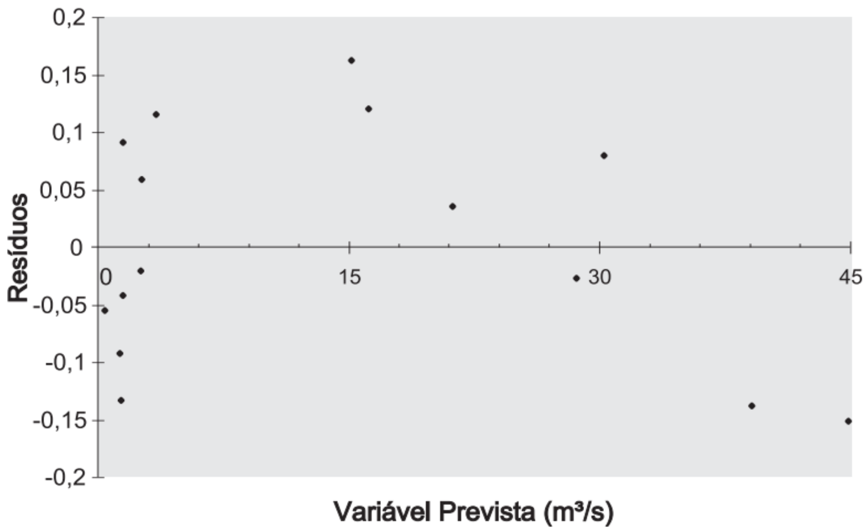
MODELO SELECIONADO

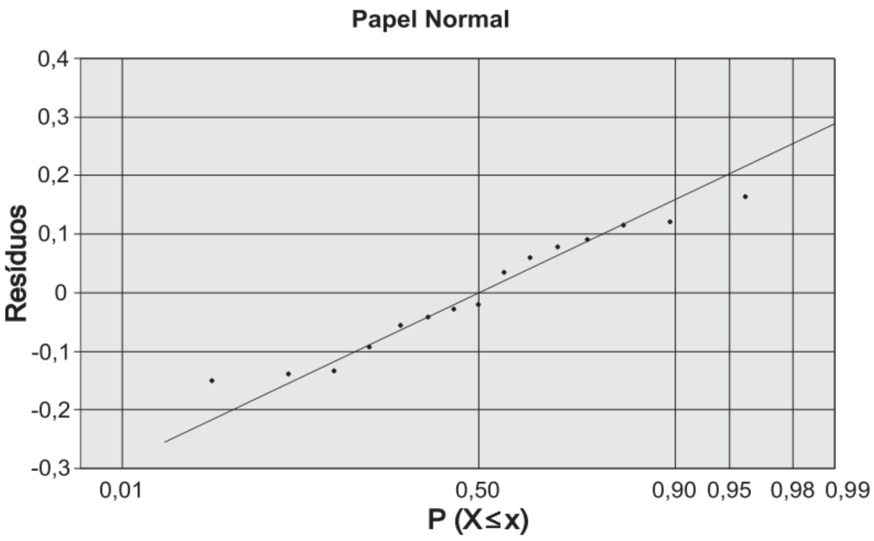
$$\ln Q = \ln \beta_0 + \beta_1 \ln A + \beta_2 \ln X_2 + \beta_3 \ln X_3$$

Modelo	$\ln(\beta_0)$	β_1	β_3	β_2	r^2	Erro Padrão
QA	-5,1696	0,9889			0,9956	0,1065

$$\beta_0 = exp[\ln(\beta_0)] = exp(-5{,}1696) = 0{,}00569$$

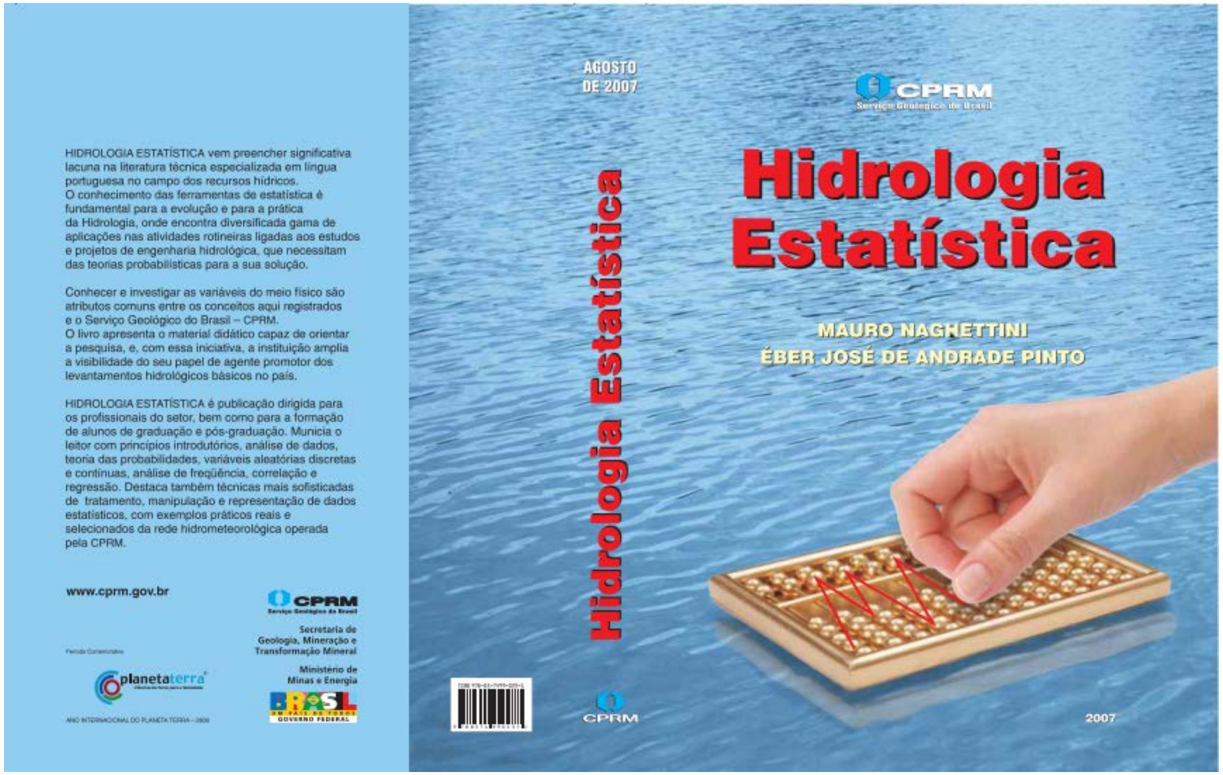
$$Q = 0{,}00596.A^{0{,}9889}$$





Recomendações

Para consolidar conhecimentos estudar no livro texto o iten 9.8



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, 22 de outubro de 2020