

Computação Experimental - Lista 4

Antônio Henrique de Moura Rodrigues, Yan Victor dos Santos, Samuel

Andrade do Couto, Victor Neves Martorelli¹

Professor: Genáina Nunes Rodrigues²

26 de junho de 2018

Exercícios do Livro do Jain, capítulo 14: 14.2, 14.4, 14.5, 14.7 e 15.2

Questão 14.2

For the disk I/O and CPU data in Example 14.1, find a linear formula to predict the number of disk I/O's given the CPU time. Answer the following questions about this regression:

- A Which parameters are significant?
- B What percentage of variation is explained by the regression?
- C What is the expected number of disk I/O's for a program with a CPU time of 40 milliseconds?
- D What bounds would you put on your answer in c if you wanted to take less than 10% chance of error on a single program to be measured tomorrow?
- E Repeat d for the case that you want to take a less than 10% chance of error on the mean of a large number of programs to be measured tomorrow?

Respostas:

A) Calculando os parâmetros, obtemos:

$$b_0 = (-5.5563, 7.8313)$$

$$b_1 = (3.3699, 4.6009)$$

O modelo de regressão obtido será:

$$y = 1.1375 + 3.9854x$$

¹ ahmoura@live.com, yanvictor_ds@hotmail.com, samuelcoouto@hotmail.com, victormartorelli@outlook.com

² genaina@cic.unb.br

Onde y é o número de E/S de disco e x o tempo de CPU.

Em um intervalo de confiança de 90%, é possível observar que apenas b_1 é significativo, dado que b_0 contém 0 em seu intervalo.

B) A regressão explica um percentual de 97.15%.

C) Para um valor de $x = 40$, teremos:

$$1.1375 + 3.9854 \cdot 40 = 160.5535$$

D) Para uma única predição, o intervalo de confiança de 90% fica: (139.5045 , 181.6038).

E) Para a previsão da média, o intervalo de confiança de 90% fica: (141.4449,179.6634).

Questão 14.4

The designers of a database information system that allows its users to search backward for several days wanted to develop a formula to predict the time it would take to search. Actual elapsed time was measured for several different values of days. The measured data is shown in Table below. Prepare a simple regression model for this data to predict elapsed time as a function of the number of days and interpret results.

Number of Days	Elapsed Time
1	0.65
2	0.79
4	1.36
8	2.26
16	3.59
25	5.39

Resposta:

O modelo de regressão obtido será:

$$y = 0.5105 + 0.196x$$

Onde y é o tempo estimado e x é o número de dias. A 90%, teremos os seguintes intervalos de confiança:

$$b_0 = (0.3576, 0.6635)$$

$$b_1 = (0.184, 0.2081)$$

Ambos coeficientes são significativos, pois nenhum intervalo contém 0. Valor de R^2 : 0.997.

Questão 14.5

In the second phase of Exercise 14.4, the analyst measured the elapsed time as a function of the complexity of the query. The complexity was measured by the number of

keywords in the query. The number of disk-read operations were also measured, as shown in Table below. For this data prepare regression models to predict the elapsed time as a function of number of keywords and interpret the results.

Number of Keywords	Elapsed Time	Number of Disk Reads
1	0.75	3
2	0.70	6
4	0.80	7
8	1.28	78
16	1.69	92

Resposta:

O modelo de regressão obtido será:

$$y = 0.635 + 0.0631x$$

Onde y é o tempo estimado e x é o número de chaves. A 90%, teremos os seguintes intervalos de confiança:

$$b0 = (0.4611, 0.8089)$$

$$b1 = (0.042, 0.0841)$$

É possível concluir que os parâmetros são significativos, pois nenhum inclui 0 em seu intervalo. Valor de R2: 0.943.

Questão 14.7

The time to encrypt a k-byte record using an encryption technique is shown in Table below. Fit a linear regression model to this data. Use visual tests to verify the regression assumptions.

Record Size		Observations	
1		2	3
128	386	375	393
256	850	805	824
384	1,544	1,644	1,553
512	3,035	3,123	3,235
640	6,650	6,839	6,768
768	13,887	14,567	13,456
896	28,059	27,439	27,659
1024	50,916	52,129	51,360

Resposta:

Temos os coeficientes:

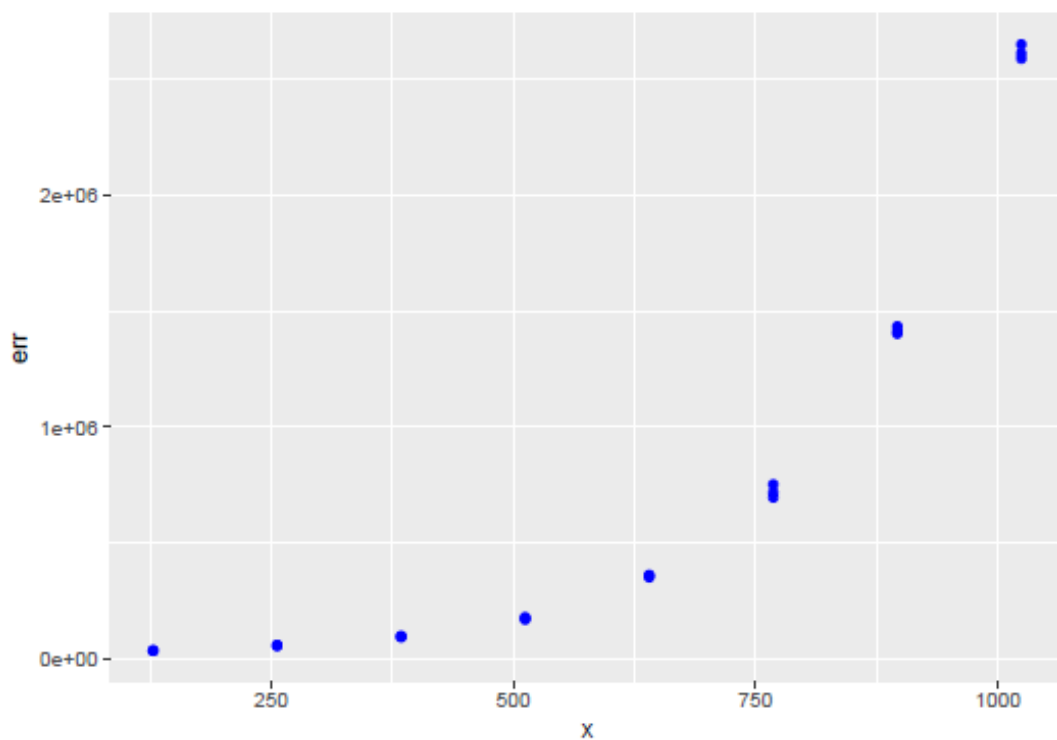
$$b0 = -1.5315964 \cdot 10^4$$

$$b1 = 49.5572$$

O modelo de regressão obtido será:

$$y = -1.5315964 \cdot 10^4 + 49.5572x$$

Onde y é o tempo e x é o tamanho do registro. Ambos b0 e b1 se mostram significativos, com um R2 de 0.744. Apesar disso, fazendo um plot dos resíduos pelos valores estimados, é possível notar que ambos possuem uma relação não linear entre si, indicando que os erros não são independentes das variáveis de predição. O seguinte gráfico mostra essa relação:



Questão 15.2

The time to encrypt or decrypt a k-bit record was measured on a uniprocessor as well as on a multiprocessor. The time in milliseconds are shown in Table below. Using a log transformation and the method for categorical predictors, fit a regression model and interpret the results.

Resposta:

k	Uniprocessor	Multiprocessor
128	93	67
256	478	355
512	3,408	2,351
1024	25,410	17,022