

Sistemas Lineares

Dr^a Marise Miranda
Msc Eduardo Verri

Seguindo com o exercício...

Contexto

Você é um administrador de sistema responsável pela manutenção de um servidor. Nos últimos meses, você coletou dados sobre o uso da CPU do servidor (em porcentagem) e o número de usuários ativos durante diferentes horários do dia. Você suspeita que o uso da CPU pode ser influenciado pelo número de usuários ativos. Você deseja analisar os dados para determinar se há uma relação entre o uso da CPU e os usuários ativos e criar um modelo preditivo para o uso da CPU com base no número de usuários ativos.

Dados:

Hora do dia	Nº ativo de usuários	Uso de CPU (%)
09:00	10	20,0
10:00	12	25,2
11:00	15	30,0
12:00	25	45,1
13:00	22	42,7
14:00	18	33,6
15:00	15	31,5
16:00	20	45,0
17:00	28	53,1
18:00	30	60,2

2 b) Se você acredita que existe uma relação linear, calcule a inclinação e a interceptação da linha de melhor ajuste que descreve a relação entre o uso da CPU e os usuários ativos

Na aula anterior utilizamos retas distintas para trabalhar com esse exercício, mas não sabemos qual o melhor modelo. Veremos apenas utilizar um conceito, de regressão linear, aplicando python, mas que será discutido posteriormente.

Lembrando:

$$\hat{y} = a * x + b$$

- \hat{y} : Variável dependente de uma relação estatística, na qual a distribuição está baseada em estimativas de dados colhidos por amostragem
- a : valor constante, coeficiente angular da reta de regressão (ou inclinação da reta)
- x : variável independente, parte de nossa base de dados
- b : valor constante, ponto de intersecção entre a reta e o eixo vertical y (é o valor de y quando x é zero)

Até aí tudo bem, mas como fazemos uma reta que melhor represente nosso conjunto de dados? Estatisticamente falando temos as seguintes maneiras de calcular os coeficientes a e b :

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad b = \bar{y} - a * \bar{x}$$

- n : tamanho da amostra
- \bar{y} : valor médio da variável dependente (y)

$$\bar{y} = \frac{\sum y_i}{n}$$

Sistemas Lineares

Dr^a Marise Miranda
Msc Eduardo Verri

- \bar{x} : valor médio da variável independente (x)

$$\bar{x} = \frac{\sum x_i}{n}$$

Mas, vamos fazer isso na mão? **Não**

Precisamos sim conhecer a origem, porém já conseguimos trabalhar computacionalmente com essas fórmulas estatísticas.

Vamos utilizar duas formas, uma mais simples, que resolve o problema. E outra com mais pompa e circunstância!

Método 1

Utilizando a boa e velha biblioteca **statistics**, módulo já presente no python que apenas precisa ser invocado no script. ([statistics – Mathematical statistics functions – Python 3.11.5 documentation](#))

```
>>>import statistics as st
>>>usuario = [10, 12, 15, 25, 22, 18, 15, 20, 28, 30]
>>>uso_cpu = [20, 25.2, 30, 45.1, 42.7, 33.6, 31.5, 45, 53.1, 60.2]
>>>a_coef_angular, b_coef_linear = st.linear_regression(usuario, uso_cpu)
>>>print('Equação da reta: y = {:.04f}x + {:.04f}'.format(a_coef_angular, b_coef_linear))
Equação da reta: y = 1.8472x + 2.6187
```

Pronto!

Não precisou nem baixar biblioteca diferente para resolver o problema. É óbvio que não temos a visualização nem a pompa e a glória que iremos ver no método 2.

Método 2

Queremos algo grandioso, que nos ajude na visualização, ao mesmo tempo que resolvemos o problema. Então além de ser pomposo, tem que ser funcional!

Utilizaremos a biblioteca **scipy** para a fórmula de regressão linear, ela faz basicamente a mesma coisa que a função utilizada da biblioteca **statistics**, contudo, ela nos trás mais informações. Onde numa análise mais aprofundada pode nos enriquecer com mais dados 😊. ([SciPy](#))

Utilizaremos a biblioteca **numpy**, uma biblioteca poderosa para criar vetores com n-dimensões e nos fornecer ferramentas para computação numérica! ([NumPy](#))

Por fim a tradicionalíssima **matplotlib** que nos fornecerá uma maneira de criar gráficos (storytelling com gráficos é mais legal!) tanto estáticos quanto dinâmicos. ([Matplotlib – Visualization with Python, 12. An Application: Solving a Simple Regression Problem – Programming with Python for Engineers 1.0 documentation \(pp4e-book.github.io\)](#))

```
>>>from scipy import stats
>>>import numpy as np
>>>import matplotlib.pyplot as plt
>>>usuarios = np.array([10, 12, 15, 25, 22, 18, 15, 20, 28, 30])
>>>uso_cpu = np.array([20, 25.2, 30, 45.1, 42.7, 33.6, 31.5, 45, 53.1, 60.2])
>>>solucao = stats.linregress(usuarios,uso_cpu)
>>> a_coef_angular, b_coef_linear = solution.slope, solution.intercept
```

Sistemas Lineares

Dr^a Marise Miranda

Msc Eduardo Verri

Peça para imprimir o **solucao**, qual a diferença de informação que eu tenho do método 1 para o método 2??

```
>>>print('A equação da reta é: y = {:.04f}x + {:.04f}'.format(a_coef_angular, b_coef_linear))
```

A solução bateu com o método 1??

Vamos criar funções para podermos trabalhar de forma um pouco mais genérica

```
>>>def formula(a,b,x):  
    return a*x + b
```

```
>>> def graph(a,b):  
    x = usuario  
    y = formula(a,b,x)  
    plt.scatter(usuario,uso_cpu)  
    plt.plot(x,y)  
    plt.xticks(np.arange(5, 40, step = 5))  
    plt.xlabel('usuários')  
    plt.ylabel('uso da CPU [ % ]')  
    plt.title('uso da cpu por usuário logado')  
    plt.grid()  
    plt.show()
```

```
>>>graph(a_coef_angular, b_coef_linear)
```

Você agora pode criar um script para automatizar o processo, deixar ainda mais genérico o código!

Sistemas Lineares

Dr^a Marise Miranda

Msc Eduardo Verri

Vamos praticar!

Exercício 1: Tesouro direto

Você é um investidor individual que está pensando em fazer um pé de meia. Você analisou os rendimentos de um título do tesouro atrelado à inflação (IPCA+2045) e coletou dados sobre os rendimentos (Y) durante um período. Seu principal interesse está em compreender as relações entre os rendimentos desse título e seus períodos de vencimento (M).

Tempo até o vencimento (anos)	Yield (%)	Tempo até o vencimento (anos)	Yield (%)
18	2,9	15	4,7
5	4,2	7	4,3
11	3,2	20	2,7
9	3,8	19	2,5
14	4,0	16	4,1
6	4,5	21	2,3
13	3,4	10	3,5
8	3,7	17	3,2
22	2,1	12	3,6

- Calcule a média, mediana e desvio padrão dos rendimentos do título.
- Formule um sistema linear de equações para representar a relação entre os rendimentos do título e os períodos de vencimento.
- Analise os coeficientes obtidos na resolução dos sistemas lineares. Que insights você pode obter sobre a sensibilidade dos rendimentos aos períodos de vencimento para esse tipo de título?
- Com base na sua análise, forneça recomendações aos investidores interessados em títulos do tesouro com diferentes prazos de vencimento.

Sistemas Lineares

Dr^a Marise Miranda

Msc Eduardo Verri

Exercício 2: Fotografia

Como uma amante da fotografia que gosta de explorar a relação entre as configurações da câmera e o brilho das fotografias, você coletou dados sobre as configurações de uma abertura (A) específica ($f1/4$), configurações de velocidade do obturador (S) e os valores correspondentes de brilho da imagem (B). Você deseja entender como essas configurações afetam o brilho geral de suas fotos.

Velocidade do obturador (s)	Brilho
1/2	200
1/4	190
1/8	175
1/15	171
1/30	168
1/60	150
1/125	148
1/250	140
1/500	131
1/1000	127

- Calcule a moda, mediana, os decils, a variância e desvio padrão dos valores de brilho.
- Crie um gráfico de dispersão para visualizar a relação entre as configurações de abertura e o brilho da imagem.
- Formule um sistema linear de equações para representar a relação entre as configurações de velocidade do obturador e brilho da imagem.
- Analise os coeficientes obtidos na resolução do sistema linear. Que informações você pode obter sobre como as configurações de abertura e velocidade do obturador afetam o brilho da imagem?
- Com base na sua análise, forneça recomendações aos fotógrafos que buscam atingir níveis específicos de brilho em suas fotos.

Faça um script em python para cada exercício!