T3.1 / N2 e Presença em aula 24

Atividade de Presença para Probest.

Para nossa análise, X são os casos e Y são as mortes.

Primeiramente iremos pegar os dados das 20 primeiras semanas epidemiológica.

In [3]:
```python
import numpy as np
```

In [4]:
```python
import pandas as pd
dados = pd.read_csv('./Dados semana 1 a 20 - Covid 2021 - Página1.csv')
```

Amostragem dos 5 primeiros dados:

In [5]:
```python
dados.head()
```

Out[5]:

	Semana	Casos	Obitos
0	1	359593	6906
1	2	379061	6665
2	3	361195	7149
3	4	360721	7500
4	5	320820	7067

2: Encontre a reta de regressão linear Y em função de X (para as semanas ímpares das 20 semanas), se ela existir.

Vamos primeiramente pegar os dados das semanas ímpares.

In [6]:
```python
impares = dados.loc[dados.Semana % 2 != 0]
# 5 primeiros dados.
impares.head()
```

Out[6]:

	Semana	Casos	Obitos
0	1	359593	6906
2	3	361195	7149
4	5	320820	7067
6	7	329394	7445
8	9	421604	10104

Temos que encontrar a fórmula da regressão: Ya = a + b*X

Onde:
a = (média de Y) - b * (média de X)
b = Covariância(X,Y) / (desvio padrão(X))^2

Vamos descobrir B:

B = Covariância(X,Y) / (desvio padrão(X))^2

Como todos os dados são diferentes, a esperança de X e Y é igual à média aritimética.

Covâriancia (X, Y):

Cov(X,Y) = E(X*Y) - [E(X)* E(Y)]

In [7]:
```python
# E (X), onde X são os casos.
E_X = impares.Casos.sum() / len(impares.Casos)
print('Esperança/Média de  X (Casos)', E_X)
```

```
Esperança/Média de  X (Casos) 408086.5
```

In [8]:
```python
# E(Y), onde Y são os Obitos
E_Y = impares.Obitos.sum() / len(impares.Obitos)
print('Esperança/Média de  Y (Óbitos)', E_Y)
```

```
Esperança/Média de  Y (Óbitos) 12466.3
```

In [9]:
```python
# E(X*Y)
X_Y = impares.Casos * impares.Obitos
print("Valores de X*Y:", X_Y.values)
E_X_Y = X_Y.sum() / len(X_Y)
print('\nEsperança/Média de X*Y:', E_X_Y)
```

```
Valores de X*Y: [2483349258 2582183055 2267234940 2452338330 4259886816 8010977364
 9099325105 9258249240 7078943200 5904336345]

Esperança/Média de E*Y: 5339682365.3
```

Então temos:

In [10]:
```python
cov_X_Y = E_X_Y - (E_X * E_Y)
print('Covâriancia de X e Y:', E_X_Y)
```

```
Covâriancia de X e Y: 5339682365.3
```

(Desvio Padrão(X))^2

Desvio Padrão(X) = Raiz de Variância (X) = raiz de $E(X^2) - E(X)^2$
Logo, (Desvio Padrão(X))^2 = (Raiz de Variância (X))^2 = Variância(X) = $E(X^2) - E(X)^2$

In [62]:
```python
# Calcular E(X²), onde X são os casos.
x_ao_quadrado = impares.Casos**2
print('Valores de X²:', x_ao_quadrado.values)
E_X_2 = x_ao_quadrado.sum() / len(x_ao_quadrado)
print('\n E(X²) = ', E_X_2)
```

```
Valores de X²: [129307125649 130461828025 102925472400 108500407236 177749932816
 261656802576 214586665225 207102357225 174523417600 194176829025]

 E(X²) =  170099083777.7
```

In [12]:
```python
from math import sqrt
# Variância = E(x²) - E(x)²
var_X = E_X_2 - (E_X**2)
print('Variância de X:', var_X)
dp_X = sqrt(var_X)
print('Desvio padrão de X:', dp_X)
```

```
Variância de X: 3564492295.450012
Desvio padrão de X: 59703.36921355454
```

Agora podemos calcular: B = Covariância(X,Y) / (desvio padrão(X))^2

In [13]:
```python
B = cov_X_Y / (dp_X**2)
print('Temos B:', B)
```

```
Temos B: 0.07079651446353907
```

Agora calculamos A: a = (média de Y) - b * (média de X)

In [14]:
```python
A = E_Y - (B * E_X)
print('Valor de A:', A)
```

```
Valor de A: -16424.80179962504
```

Por fim, a fórmula da reta: Ya = a + b*X*

Temos: ***Y = -16424.80 + 0.070*** **X**

In [15]:
```python
# Definição da nossa formula Y = A + B*X
def previsao(X, a = A, b = B):
    Y = a + b * X
    return Y
```

In [16]:
```python
# Prever usando a regressão linear.
previsao(X=450_000)
```

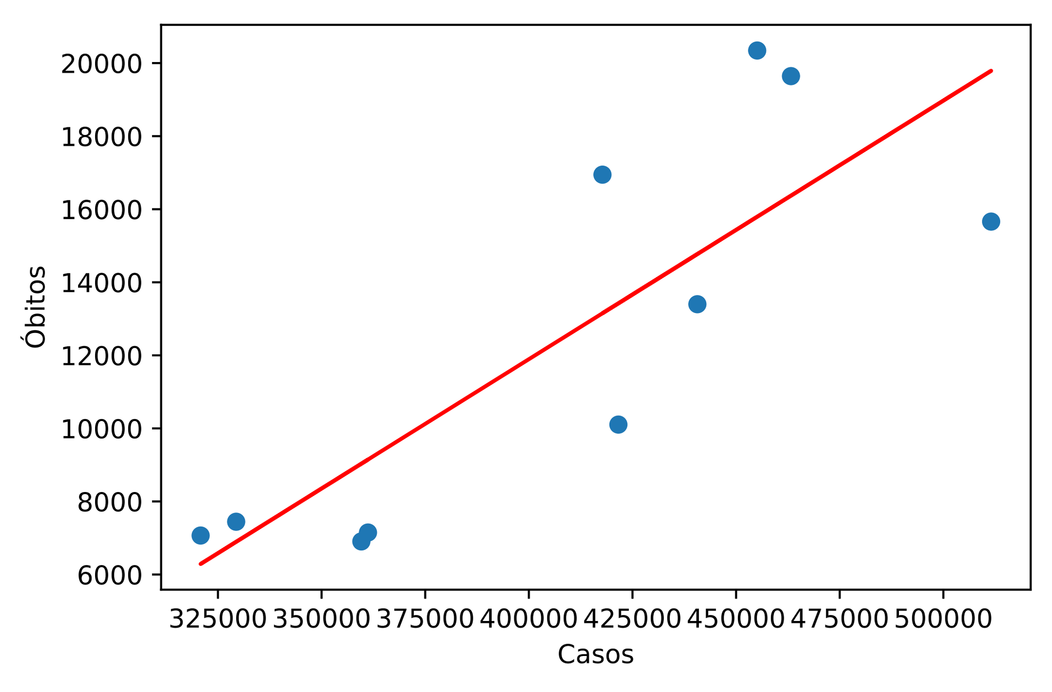
Out[16]:
```
15433.629708967546
```

3.1 Plote em um gráfico;

Vamos agora plotar os dados em um Gráfico de dispersão e a Reta de Regressão

In [17]:
```python
import matplotlib.pyplot as plt
plt.scatter(impares.Casos, impares.Obitos)
plt.plot(impares.Casos,[previsao(i) for i in impares.Casos], color='red')
plt.ylabel('Óbitos')
plt.xlabel('Casos')
```

Out[17]:
```
Text(0.5, 0, 'Casos')
```



3.2 Determine os coeficientes de determinação e correlação linear entre as variáveis X e Y;

Coeficiente de Determinação

Agora vamos calcular o Coeficiente de Determinação

R^2 = Soma(Yia - E(Y))^2 / Soma(Yi - E(Y))^2

In [18]:
```python
previsoes = [previsao(i) for i in impares.Casos]
previsoes_soma_2 = [(p - E_Y)**2 for p in previsoes]

real = impares.Obitos.values
real_soma_2 = [(r - E_Y)**2 for r in real]
```

In [19]:
```python
R_2 = sum(previsoes_soma_2) / sum(real_soma_2)
print('Coeficiente de Determinação:', R_2)
```

```
Coeficiente de Determinação: 0.6771566077941553
```

Sabemos que $R^2 = P^2$

Então:

In [20]:
```python
# p = raiz quadrada de R²
p = sqrt(R_2)
print('Coeficiente de Correlação linear:', p)
```

```
Coeficiente de Correlação linear: 0.8228952593095644
```

3.3 - Escolha 5 semanas pares e faça a previsão do número de mortes por semana, na reta ajustada de Y; determinando o erro cometido no processo para cada semana.

Vamos primeiro pegar os dados pares:

In [21]:
```python
pares = dados.loc[dados.Semana % 2 == 0]
pares_5 = pares[:5]
pares_5
```

Out[21]:

	Semana	Casos	Obitos
1	2	379061	6665
3	4	360721	7500
5	6	311959	7520
7	8	378084	8244
9	10	500099	12766

Vamos encontrar a reta da regressão.

Como já foi apresentado detalhadamente antes, só passaremos por cima

Yp = a + b*Xp

In [34]:
```python
# E de X pora os pares.
X_PAR = pares_5.Casos
E_X_PAR = X_PAR.sum() / len(X_PAR)

# E de Y para os pares
Y_PAR = pares_5.Obitos
E_Y_PAR = Y_PAR.sum() / len(Y_PAR)

print(f'Esperança de X: {E_X_PAR}. Esperança de Y: {E_Y_PAR}')
```

```
Esperança de X: 385984.8. Esperança de Y: 8539.0
```

In [32]:
```python
# E de X*Y
X_Y_PAR = X_PAR * Y_PAR
E_X_Y_PAR = X_Y_PAR.sum() / len(X_Y_PAR)
print('\nEsperança/Média de X*Y (Par):', E_X_Y_PAR)
```

```
Esperança/Média de X*Y (Par): 3415793815.0
```

In [39]:
```python
# Desvio Padrão de X
X_2_PAR = X_PAR**2
E_X_2_PAR = X_2_PAR.sum() / len(X_2_PAR)
var_X_PAR = E_X_2_PAR - E_X_PAR**2
print('A variância de X se (par) dá por:', var_X_PAR)
dp_X_PAR = sqrt(var_X_PAR)
print('O desvio padrão de X (par) se dá por:', dp_X_PAR)
```

```
A variância de X se dá por: 3850098188.960022
O desvio padrão de X (par) se dá por: 62049.15945409754
```

In [40]:
```python
#Covariância de X e Y
cov_X_Y_PAR = E_X_Y_PAR - (E_X_PAR * E_Y_PAR)
print('Covariância de X e Y (Pares):', cov_X_Y_PAR)
```

```
Covariância de X e Y (Pares): 119869607.80000019
```

In [41]:
```python
# Valor de B
B_PAR = cov_X_Y_PAR / dp_X_PAR**2
print('O valor de B se dá por:', B_PAR)
```

```
O valor de B se dá por: 0.03113416903073296
```

In [42]:
```python
# Valor de A
A_PAR = E_Y_PAR - B_PAR * E_X_PAR
print('O valor de A se dá por:', A_PAR)
```

```
O valor de A se dá por: -3478.3160064936546
```

AGora podemos montar nossa formula:

Yp = Ap + Bp*Xp

In [43]:
```python
# Formula da previsão
def previsao_par(Xp, Ap = A_PAR, Bp = B_PAR):
    Yap = Ap + Bp * Xp
    return Yap
```

Agora sim, vamos fazer a previsão e encontrar o valor do erro:

In [61]:
```python
# Previsões pares
for index, caso in enumerate(X_PAR):
    previsao_item = previsao_par(caso)
    print(f'Valor real: {Y_PAR.values[index]}.\nValor obtido: {previsao_item}.\nDiferença Absoluta: {previsao_i
```

```
Valor real: 6665.
Valor obtido: 8323.433240465012.
Diferença Absoluta: 1658.4332404650122

Valor real: 7500.
Valor obtido: 7752.432580441369.
Diferença Absoluta: 252.43258044136928

Valor real: 7520.
Valor obtido: 6234.268230164769.
Diferença Absoluta: -1285.7317698352308

Valor real: 8244.
Valor obtido: 8293.015157321986.
Diferença Absoluta: 49.01515732198641

Valor real: 12766.
Valor obtido: 12091.850791606868.
Diferença Absoluta: -674.1492083931316
```

In [ ]: