Universidade do Estado do Rio de Janeiro – UERJ

Campus Regional Instituto Politécnico do Estado do Rio de Janeiro - IPRJ

Curso de Graduação em Engenharia da Computação

TRABALHO 2 DE MODELOS LINEARES

Leonardo Simões

Leonardo T. Muzí de Carvalho

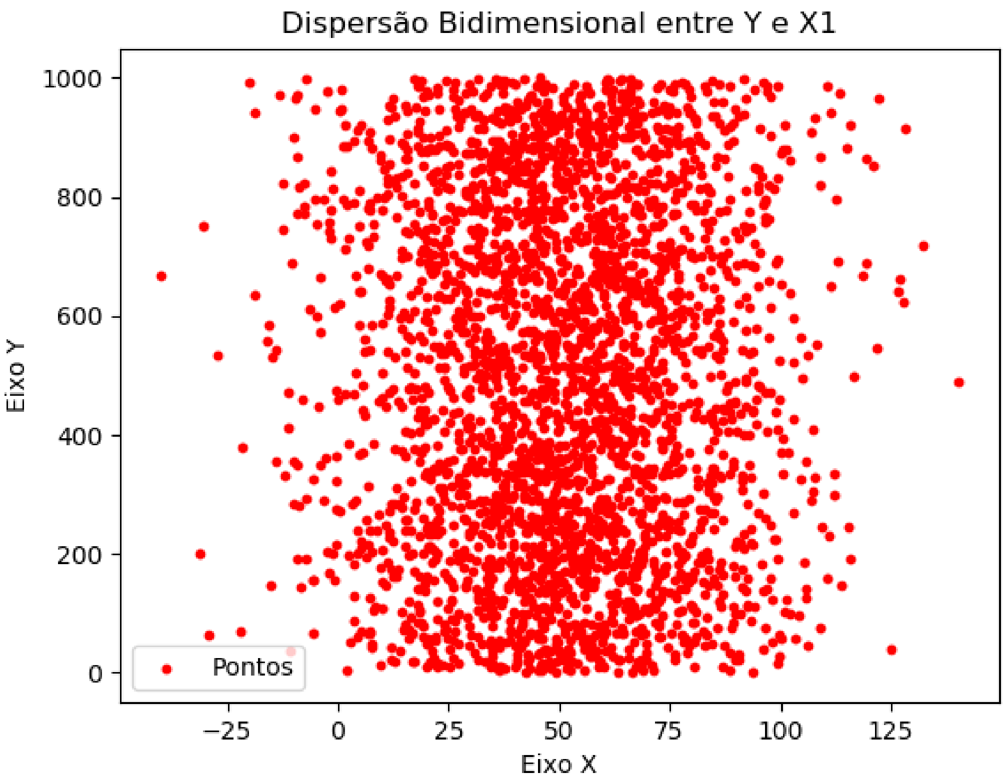
Professor:

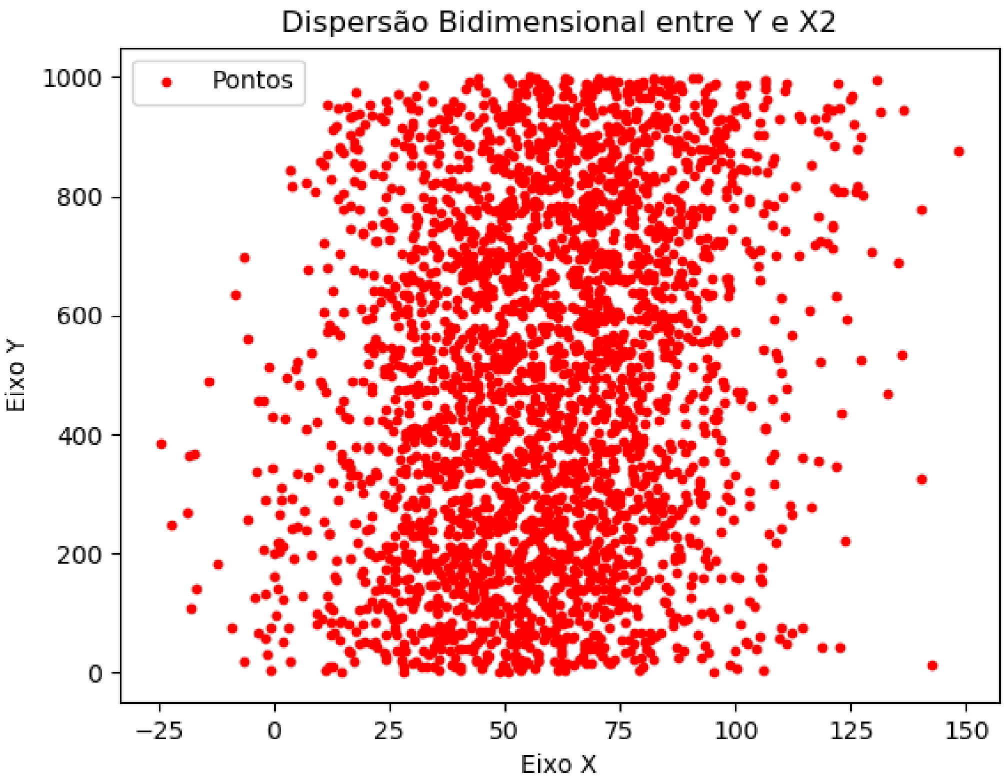
Bernardo Sotto-Maior Peralva

Nova Friburgo, 09 de Julho de 2018.

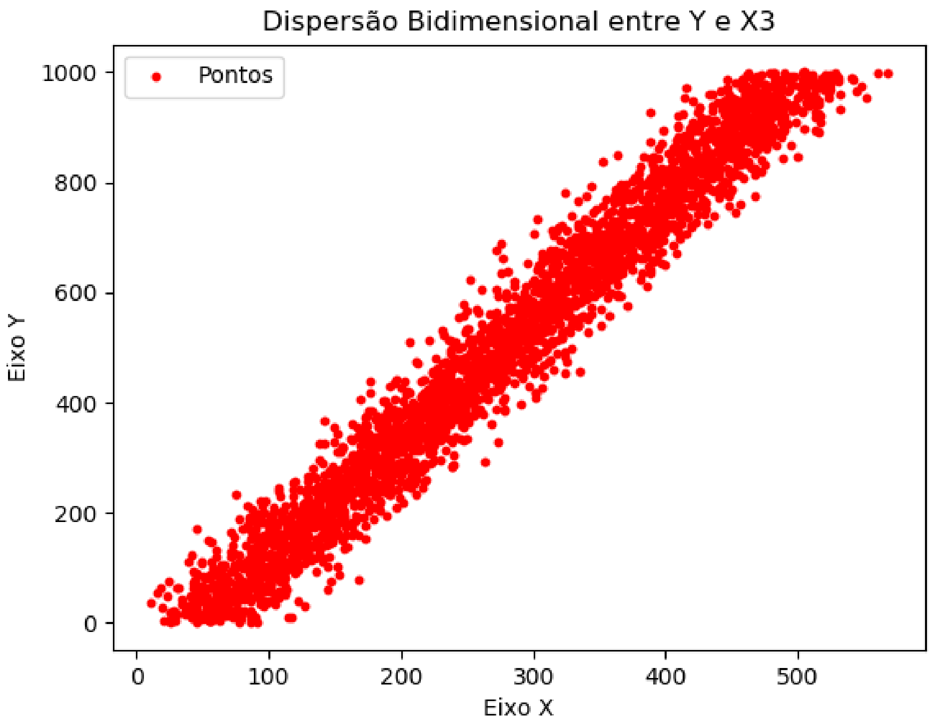
Item A: Gráficos de dispersão bidimensional (XY) entre a variável y e cada uma das variáveis regressoras.

As correlações podem ser observadas visualmente pelos gráficos de dispersão bidimensional e indicadas pelo coeficiente de correlação, sendo discutidas no Item B.

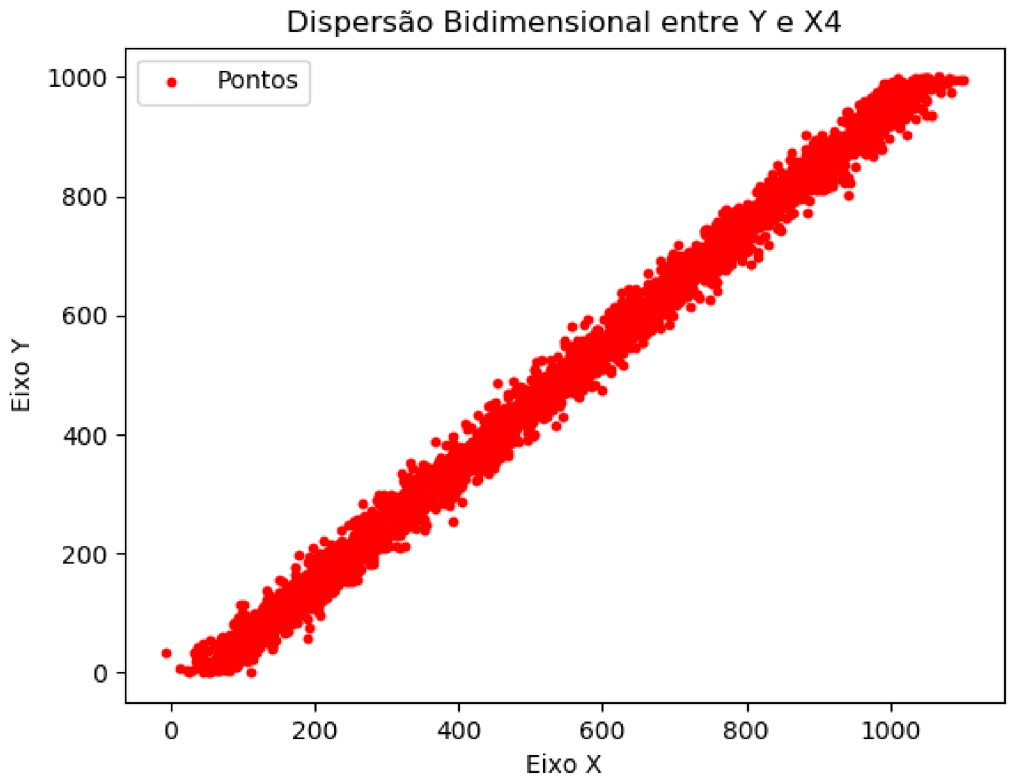


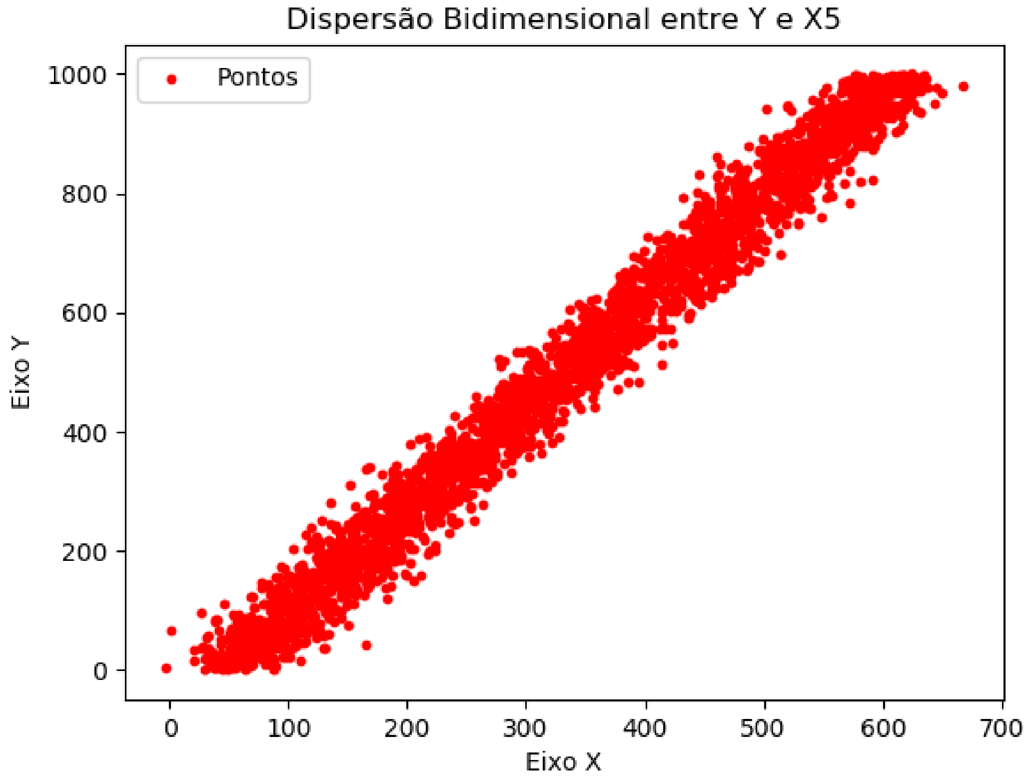


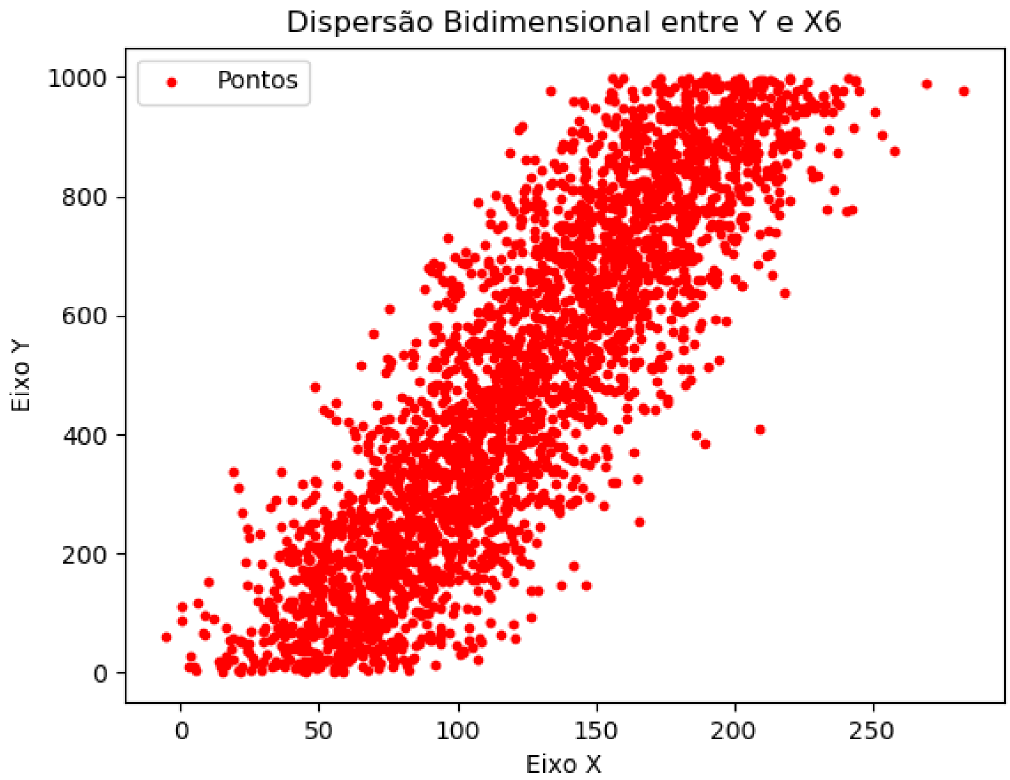
Dispersão Bidimensional entre Y e X2
Pontos
Eixo Y
Eixo X



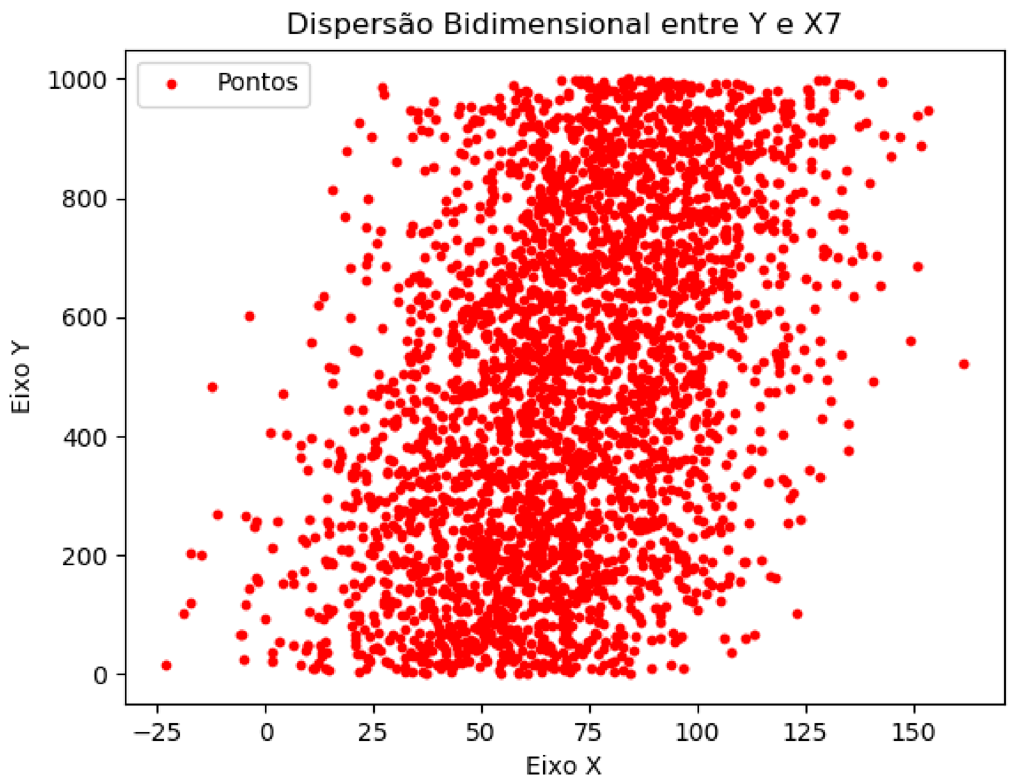
Dispersão Bidimensional entre Y e X3
Pontos
Eixo Y
Eixo X


Dispersão Bidimensional entre Y e X4
Pontos
Eixo Y
Eixo X
0
200
400
600
800
1000


Dispersão Bidimensional entre Y e X5
Pontos
Eixo Y
Eixo X
0
100
200
300
400
500
600
700
800
1000



Dispersão Bidimensional entre Y e X6
Pontos
Eixo Y
Eixo X
0
200
400
600
800
1000
0
50
100
150
200
250



Dispersão Bidimensional entre Y e X7
Pontos
Eixo Y
Eixo X
0
200
400
600
800
1000
−25
0
25
50
75
100
125
150

Item B: Estimar a correlação entre a variável Y e cada uma das variáveis regressoras.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$\boldsymbol{\rho(Y, X_i)}$	-0.000369	0.208	0.982	0.996	0.989	0.874	0.430

O tipo e intensidade da correlação entre as variáveis x_i e y são determinadas pelo coeficiente de correlação $\boldsymbol{\rho(Y, X_i)}$ e são visualizados pelos gráficos de dispersão entre x_i e y, apresentados no item A. As variáveis x_3, x_4, x_5 e x_6 estão fortemente e positivamente relacionadas com a variável y, já que apresentam valores relativamente próximos de 1 para o coeficiente de correlação. As variáveis x_1 e x_2 estão fracamente relacionadas com a variável y, já que apresentam valores relativamente próximos de 0 para o coeficiente de correlação. A variável x_7 possui uma correlação moderada com a variável y.

Item C: Calcular o coeficiente de determinação entre a variável Y e cada uma das variáveis regressoras.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$\boldsymbol{R^2(Y, X_i)}$	$1.36 \cdot 10^{-7}$	0.043	0.964	0.992	0.977	0.763	0.185

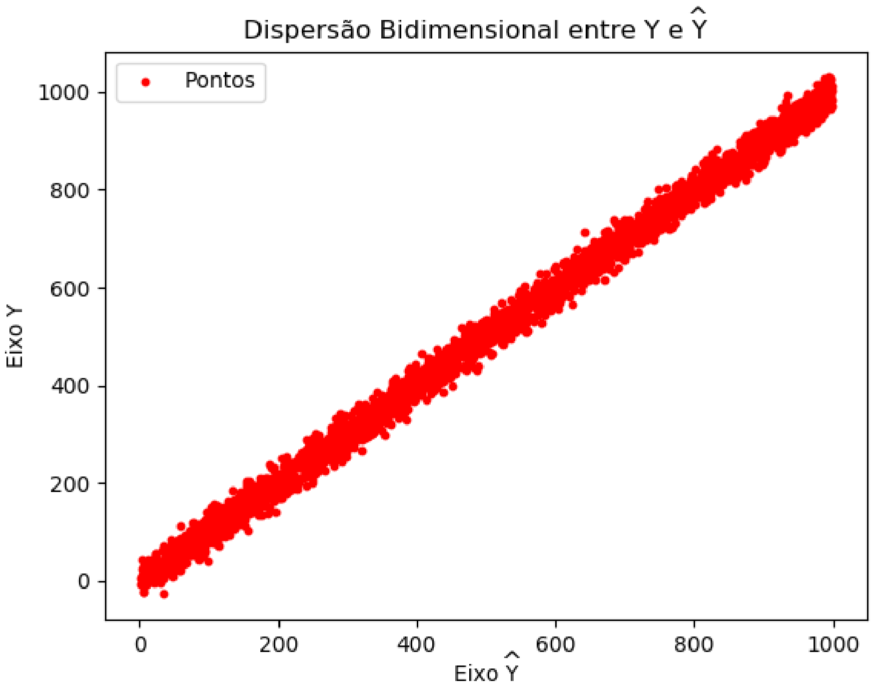
O coeficiente de determinação $\boldsymbol{R^2(Y, X_i)}$ é uma medida variabilidade de Y em relação a variabilidade da variável x_i. Pelos valores de R^2 percebe-se que Y é altamente explicado por x_3, x_4 e x_5, bem explicado por x_6 e muito pouco explicado por x_1, x_2 e x_7, sendo x_1 e x_2 praticamente insignificantes para o modelo.

Item D: Encontrar o hiperplano de quadrados mínimos:

$$\hat{\beta} = \begin{bmatrix} -7.032 \cdot 10^{1} \\ 9.763 \cdot 10^{-3} \\ 3.118 \cdot 10^{-2} \\ 2.803 \cdot 10^{-1} \\ 6.410 \cdot 10^{-1} \\ 3.680 \cdot 10^{-1} \\ 1.056 \cdot 10^{-1} \\ 3.470 \cdot 10^{-2} \end{bmatrix}$$

Item E: Calcular os valores estimados de y

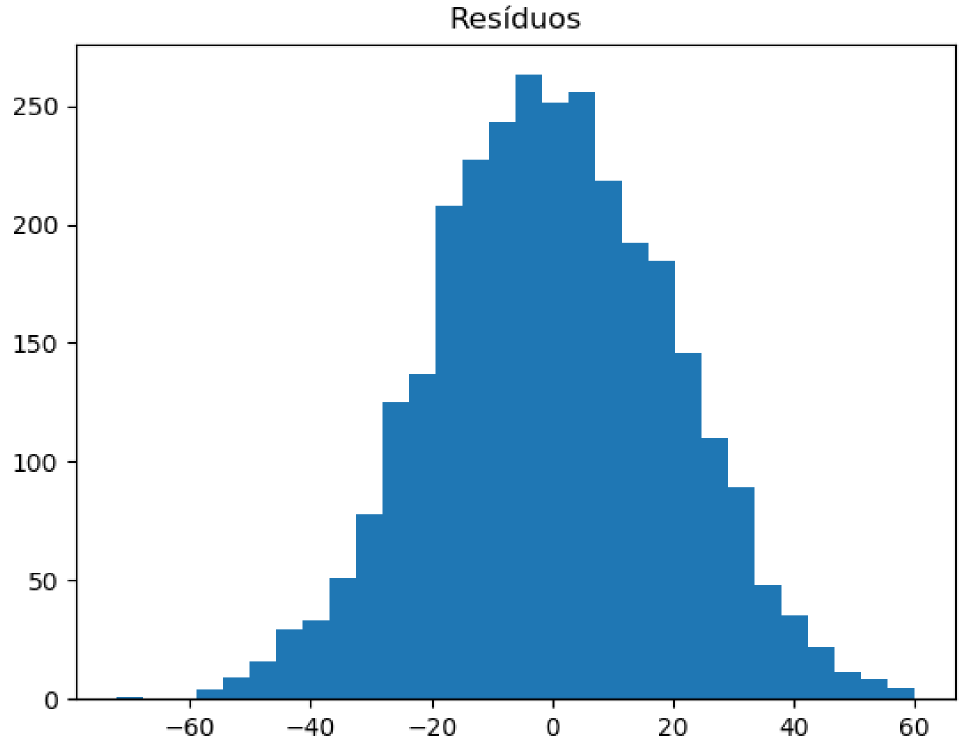
Foram calculados os $\widehat{y_i}$ de acordo com a expressão $\hat{y} = x\hat{\beta}$, sendo $\hat{\beta}$ apresentado no Item D, e sua dispersão com y_i é apresentada no gráfico a seguir, demonstrando uma relação linear e proximidade entre os valores no relação de aproximadamente 1:1.



Item F: Calcular os resíduos

Seus valores foram calculados e apresentados graficamente no Item G.

Item G: Histograma dos resíduos



Pelo histograma observa-se que os resíduos possuem distribuição normal de média zero, assim, obteve-se ótimas aproximações pelo modelo gerado.

```python
"""
Trabalho 2 de Modelos Lineares
Leonardo Simões e Leonardo T. Muzí de Carvalho
"""

import numpy as np
from numpy.linalg import inv
import matplotlib as mpl
import matplotlib.pyplot as plt

def lerArquivo(diretorio):
    arquivo = open(diretorio, 'r')
    X, Y = [], []
    for linha in arquivo:
        Y.append(float(linha.split()[0]))
        X.append([1.0]+[float(x) for x in linha.split()[1:]])
    arquivo.close()
    return (np.array(X),np.array(Y))

def QuadradosMinimos(X, Y):
    Xt = X.transpose()
    return (inv(Xt.dot(X)).dot(Xt)).dot(Y)

def plotarDispersãoBidimensional(X, Y, eixox):
    fig = plt.figure(1)
    plt.xlabel('Eixo '+ eixox[0])
    plt.ylabel('Eixo Y')
    plt.title('Dispersão Bidimensional entre Y e '+ eixox)
    plt.scatter(X, Y, label='Pontos', color='r', marker='o', s=10)
    plt.legend()
    plt.show()
    fig.savefig('Y vs ' + eixox +'.png')

def dispersãoBidimensional(X, Y):
    for i in range(1, len(X[0])):
        Xi = X[:,i]
        eixox = 'X'+str(i)
        plotarDispersãoBidimensional(Xi, Y, eixox)

def plotarResiduos(R):
    fig = plt.figure(1)
    plt.hist(R, bins=30)
    plt.title('Resíduos')
    plt.show()
    fig.savefig('Residuos.png')

#Coeficientes de correlação e determinação
def Sxx(X):
    soma = 0.0
    xm = np.average(X)
    for x in X:
        soma += (x - xm)**2
    return soma

def Syy(Y):
    soma = 0.0
    ym = np.average(Y)
    for y in Y:
        soma += (y - ym)**2
    return soma
```

```python
def Sxy(X, Y):
    soma = 0.0
    xm = np.average(X)
    ym = np.average(Y)
    for i in range(len(X)):
        soma += (X[i] - xm)*(Y[i] - ym)
    return soma

def coeficienteDeCorrelacao(X, Y):
    return Sxy(X, Y)/((Sxx(X)*Syy(Y))**(1/2))

def coeficientesDeCorrelacaoMultiplos(X, Y):
    p = []
    for i in range(1, len(X[0])):
        p.append(coeficienteDeCorrelacao(X[:,i], Y))
    return p

def coeficienteDeDeterminacaoMultiplos(X, Y):
    R2 = []
    for i in range(1, len(X[0])):
        R2.append((coeficienteDeCorrelacao(X[:,i], Y)**2))
    return R2

#Função Principal - MAIN
if __name__ == '__main__':
    diretorio = 'D:\\data6_t2.txt'
    (X,Y) = lerArquivo(diretorio)

    #Item A
    dispersãoBidimensional(X, Y)

    #Item B
    p = coeficientesDeCorrelacaoMultiplos(X, Y)
    print('p = ', p)

    #Item C
    R2 = coeficienteDeDeterminacaoMultiplos(X, Y)
    print('R2 = ', R2)

    #Item D
    B = QuadradosMinimos(X, Y)
    print('B = ', B)

    #Item E
    Yc = np.dot(X, B)
    plotarDispersãoBidimensional(Y, Yc, 'Y^')
    print('Yc = ', Yc)

    #Item F
    R = np.subtract(Y, Yc)
    print('R = ', R)

    #Item G
    plotarResiduos(R)
```