Jose Dixon

November 6, 2016

Dr. Rahman

COSC 251.001

**Assignment 3**

Question 1- Correlation

**Source Code**

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

n = 12

cor\_data = {'x':[35, 50, 75, 95, 120, 130, 145, 155, 160, 175, 185, 190],

'y':[5.88, 5.99, 6.74, 6.1, 7.47, 6.93, 6.42, 7.97, 7.92, 7.62, 6.89, 7.9]}

frame = pd.DataFrame(cor\_data, columns = ['x', 'y'])

frame

plt.scatter(frame.x, frame.y, c='red', s=30)

plt.xlabel('Depth at Which Drilling Begins')

plt.ylabel('Time to Drill')

plt.title('Scatter plot')

plt.show()

frame['x^2']=frame['x']\*\*2

frame['y^2']=frame['y']\*\*2

frame['xy']=frame['x']\*frame['y']

print sum(frame['x']), "Sum of X"

print sum(frame['y']), "Sum of Y"

print sum(frame['x^2']), "Sum of X^2"

print sum(frame['y^2']), "Sum of Y^2"

print sum(frame['xy']), "Sum of X\*Y"

ssx = sum(frame['x^2']) - (sum(frame['x'])/n), "SS(x)"

print ssx

ssy = sum(frame['y^2']) - (sum(frame['y'])/n), "SS(y)"

print ssy

ssxy = sum(frame['xy']) - (sum(frame['x'])\*sum(frame['y'])/n), "SS(xy)"

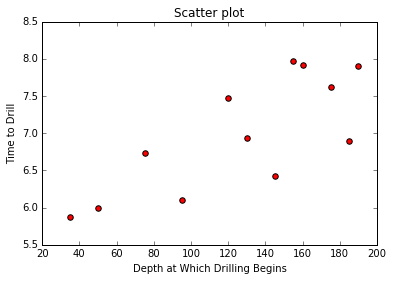
print ssxy

r = ((sum(frame['xy'])) - ((sum(frame['x']))\*sum(frame['y'])/n))/(sum(frame['x^2']) - (sum(frame['x'])/n))\*((sum(frame['y^2']) - (sum(frame['y'])/n))\*\*0.5)

print "The linear correlation is", r, "so there is little or no correlation to determine if is it positive or negative."

frame

**Output**

****

1515 Sum of X

83.83 Sum of Y

221275 Sum of X^2

592.3281 Sum of Y^2

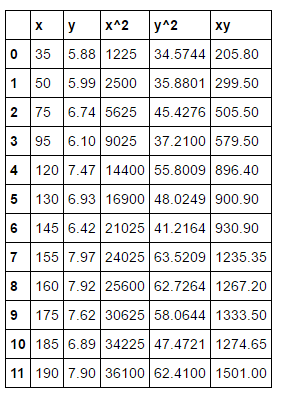
10930.2 Sum of X\*Y

(221149, 'SS(x)')

(585.34226666666666, 'SS(y)')

(346.66249999999854, 'SS(xy)')

The linear correlation is 0.0379251080913 so there is little or no correlation to determine if is it positive or negative.

****

Question 2 – Regression

**Source Code**

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

n = 5

cor\_data = {'xi':[95, 85, 80, 70, 60],

'yi':[85, 95, 70, 65, 70]}

frame = pd.DataFrame(cor\_data, columns = ['xi', 'yi'])

frame['(xi-x)']=frame['xi']-(sum(frame['xi']/5))

frame['(yi-y)']=frame['yi']-(sum(frame['yi']/5))

frame['(xi-x)^2']=(frame['xi']-(sum(frame['xi']/5)))\*\*2

frame['(yi-y)^2']=(frame['yi']-(sum(frame['yi']/5)))\*\*2

frame['(xi-x)\*(yi-y)']=frame['(xi-x)']\*frame['(yi-y)']

print sum(frame['xi']), "Sum of xi"

print sum(frame['yi']), "Sum of yi"

print sum(frame['(xi-x)^2']), "Sum of (xi-x)^2"

print sum(frame['(yi-y)^2']), "Sum of (yi-y)^2"

print sum(frame['(xi-x)\*(yi-y)']), "Sum of (xi-x)\*(yi-y)"

print sum(frame['xi'])/5, "Mean of xi"

print sum(frame['yi'])/5, "Mean of yi"

print "b1 is", sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2'])

print "b0 is", sum(frame['yi'])/5-(sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2']))\*(sum(frame['xi'])/5)

print "The regression equation is y =", (sum(frame['yi'])/5-(sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2']))\*(sum(frame['xi'])/5)), "+", (sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2'])), "x"

print "The students estimated statistics grade would be", (sum(frame['yi'])/5-(sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2']))\*(sum(frame['xi'])/5))+((sum(frame['(xi-x)\*(yi-y)'])/sum(frame['(xi-x)^2']))\*80)

print "R^2 is", (((0.2)\*sum(frame['(xi-x)\*(yi-y)'])/((sum(frame['(xi-x)^2'])/5)\*\*0.500\*(sum(frame['(yi-y)^2'])/5)\*\*0.500))\*\*2), "indicates it fits the data very well."

frame

**Output**

390 Sum of xi

385 Sum of yi

730.0 Sum of (xi-x)^2

630.0 Sum of (yi-y)^2

470.0 Sum of (xi-x)\*(yi-y)

78 Mean of xi

77 Mean of yi

b1 is 0.643835616438

b0 is 26.7808219178

The regression equation is y = 26.7808219178 + 0.643835616438 x

The students estimated statistics grade would be 78.2876712329

R^2 is 0.480321809089 indicates it fits the data very well.

