HW1

1.b

# load the data

x = pd.read\_csv("fatherson.csv")

# plot the data

plt.scatter(x["fheight"], x["sheight"])

plt.title("Relationship between Son's Height and Father's Height")

plt.xlabel("Father's Height")

plt.ylabel("Son's Height")

# calculate the scope

son = np.transpose(np.asmatrix(x["sheight"])) # change to matrix

father = np.transpose(np.asmatrix(x["fheight"]))

beta = np.linalg.inv(np.transpose(father) \* father) \* np.transpose(father) \* son #calculate the estimated beta

# fit the line

slope = np.asscalar(np.array(beta))

y=slope\*father

plt.plot(father, y, 'r');

1.c

On average, one unit increase of father's height will introduce 1.01 unit increase of son's height.

1.d

x\_bar = np.mean(x["fheight"])

y\_bar = np.mean(x["sheight"])

a= x["fheight"]-x\_bar

b= x["sheight"]-y\_bar

beta1=sum(a\*b)/sum(a\*\*2)

beta0=y\_bar-beta1\*x\_bar

beta1

beta0;

beta1=0.514

beta0=33.887

The estimated $\beta\_1$ is 0.514, the estimated $\beta\_0$ is 33.887.

2.a

#merge

train = pd.read\_csv("citibike\_train.csv")

test = pd.read\_csv("citibike\_test.csv")

weather = pd.read\_csv("weather.csv")

test1= pd.merge(test,weather, on="date")

del test1["date"]

test1.head(n=10)