

a)

x1 = work hours

x2 = gender

x = [work hours, gender]

y = income

t-statistic and p-value are showed on the following figure.

```
PS D:\NTNU\111-2\ML\assigment3> python assigment3.py
t-statistic1: 3.5664619263968755
p_val1: 0.004570033229912673
Reject the null hypothesis

t-statistic2: 0.13638234135990554
p_val2: 0.44767924871009757
Accept the null hypothesis

t_statistic: -0.2517677790422141
p_val: 0.4042261512679057
Accept the null hypothesis
PS D:\NTNU\111-2\ML\assigment3> █
```

b)

the result shows that there is some relationship between x1 and y but there is no relationship between x2 and y.

the last t-statistic on figure is t-statistic of whole model. And the result shows there is no relationship between x and y.

Code:

```
from sklearn import linear_model
from scipy.stats import t
import numpy as np
import pandas as pd
data = {
    'income' : [100, 90, 99, 62, 123, 165, 111, 50, 165, 200],
    'gender' : [0, 1, 0, 1, 1, 1, 0, 0, 1, 0],
    'working_hours' : [8, 6, 7, 5, 9, 13, 9, 5, 8, 10]
}

df = pd.DataFrame(data)
x1 = df[['working_hours']].values

x2 = df[['gender']].values
x = df[['working_hours', 'gender']].values
y = df['income'].values

regr = linear_model.LinearRegression()
regr.fit(x, y)
regr_predict = regr.predict(x)

RSS = np.sum((y - regr_predict) ** 2)
n = len(data['income'])
var = np.sqrt(RSS/(n-3))

SE1 = var/np.sqrt(np.sum((x1-np.mean(x1))**2))
# T-statistic
t_statistic1 = regr.coef_[0] / SE1
print("t-statistic1: ", t_statistic1)

if t_statistic1 < 0:
    p_val1 = t.cdf(t_statistic1, df = n-3)
else:
    p_val1 = 1-t.cdf(t_statistic1, df = n-3)
print('p_val1: ', p_val1)
```

```

if(p_val1 < 0.05):
    print('1st variable \'Reject\' the null hypothesis')
else:
    print('1st variable \'Accept\' the null hypothesis')

print('')
SE2 = var/np.sqrt(np.sum((x2-np.mean(x2))**2))
# T-statistic
t_statistic2 = regr.coef_[1] / SE2
print("t-statistic2: ", t_statistic2)

if t_statistic2 < 0:
    p_val2 = t.cdf(t_statistic2, df = n-3)
else:
    p_val2 = 1-t.cdf(t_statistic2, df = n-3)
print('p_val2: ', p_val2)
if(p_val2 < 0.05):
    print('2st variable \'Reject\' the null hypothesis')
else:
    print('2st variable \'Accept\' the null hypothesis')
print('')

a = np.mean(x)**2
b = np.sum((x - np.mean(x))**2)
SE = var * np.sqrt((1/n) + (np.mean(x1)**2) / np.sum((x1 -
np.mean(x1))**2) + (np.mean(x2)**2) / np.sum((x2 - np.mean(x2))**2))

t_statistic = regr.intercept_ / SE
print("t_statistic: ", t_statistic)
if t_statistic < 0:
    p_val = t.cdf(t_statistic, df = n-3)
else:
    p_val = 1-t.cdf(t_statistic, df = n-3)
print('p_val: ', p_val)
if(p_val < 0.05):
    print('Reject the null hypothesis')
else:
    print('Accept the null hypothesis')

```