

Assignment 3

Introduction to Data Science



October 26, 2023

FA21-BSE-059

Ammar Ahmed

**Q:** Consider the following data about population

|  |  |  |
| --- | --- | --- |
| **Age** | **Income** | **Savings** |
| 20 | 100,000 | 50,000 |
| 22 | 120,000 | 40,000 |
| 32 | 80,000 | 5,000 |
| 40 | 350,000 | 10,000 |
| 50 | 50,000 | 0 |
| 36 | 150,000 | 30,000 |
| 47 | 250,000 | 8,000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Predicted Income | | Predicted Savings | |
|  | Linear Regression | Polynomial Regression | Linear Regression | Polynomial Regression |
| 15 |  |  |  |  |
| 55 |  |  |  |  |
| 65 |  |  |  |  |

You are required to generate the R and R2 factors of the Linear and Polynomial regressions respectively and fill in the following data:

By using the percentile calculation, please find out the income and saving of the following percentage of the population:

|  |  |  |
| --- | --- | --- |
| Percentile | Income | Savings |
| 25 |  |  |
| 50 |  |  |
| 75 |  |  |

**Solution**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Predicted Income | | Predicted Savings | |
|  | Linear Regression | Polynomial Regression | Linear Regression | Polynomial Regression |
| 15 | 105356.50 | 626323.59 | 49679.68 | 100141.17 |
| 55 | 207470.44 | -525540.81 | -7998.56 | -35045.77 |
| 65 | 232998.92 | -3191600.73 | -22418.12 | -187266.614 |

|  |  |  |
| --- | --- | --- |
| Percentile | Income | Savings |
| 25 | 9000 | 6500 |
| 50 | 12000 | 6500 |
| 75 | 20000 | 6500 |

**Regression code**

import numpy as np

from scipy import stats

from sklearn.metrics import r2\_score

# Given data

age = np.array([20, 22, 32, 40, 50, 36, 47])

income = np.array([100000, 120000, 80000, 350000, 50000, 150000, 250000])

savings = np.array([50000, 40000, 5000, 10000, 0, 30000, 8000])

ages\_to\_predict = [15,55,65]

#slope, intercept, r, p, std\_err = stats.linregress(age, income)

#slope, intercept, r, p, std\_err = stats.linregress(age, savings)

#mymodel = np.poly1d(np.polyfit(age, income, 2))

mymodel = np.poly1d(np.polyfit(age, savings, 2))

# Define a linear regression function

def linearRegression(x):

return slope \* x + intercept

#for age in ages\_to\_predict:

#print(linearRegression(age))

for age in ages\_to\_predict:

print(mymodel(age))

**Percentile code**

import numpy

age = [20, 22, 32, 40, 50, 36, 47]

income = [100000, 120000, 80000, 350000, 50000, 150000, 250000]

savings = [50000, 40000, 5000, 10000, 0, 30000, 8000]

print(numpy.percentile(income, 25))

print(numpy.percentile(income, 50))

print(numpy.percentile(income, 75))

print(numpy.percentile(savings, 25))

print(numpy.percentile(savings, 25))

print(numpy.percentile(savings, 25))