

Data Set chosen here is [Graduate Admission](#)

Code

Written in python can be found at [GitHub](#)

```
import math
import pandas

def selectattribute(csv_file):
    index = 0 # default value of index
    for row in csv_file.keys():
        print(index, ': ', row)
        index += 1
    index = int(input('input : '))
    return csv_file.keys()[index]

if __name__ == '__main__':
    target_csv_file_name = 'dataset/Admission_Predict.csv'
    csv_file = pandas.read_csv(target_csv_file_name)

    print('select one of the Y attribute')
    y_attr = selectattribute(csv_file)
    print(' Y attribute set to : ', y_attr)

    print('select one of the X attribute')
    x_attr = selectattribute(csv_file)
    print(' X attribute set to : ', x_attr)

    # print('table creation')
    y = list(csv_file[y_attr])
    x = list(csv_file[x_attr])

    y_sum = csv_file[y_attr].sum()
    x_sum = csv_file[x_attr].sum()

    y_mean = csv_file[y_attr].mean()
    x_mean = csv_file[x_attr].mean()

    y_minus_y_mean = []
    for ele in y:
        y_minus_y_mean.append(ele-y_mean)
    x_minus_x_mean = []
    for ele in x:
        x_minus_x_mean.append(ele-x_mean)

    y_sq = []
```

```
for ele in y:
    y_sq.append(ele**2)
x_sq = []
for ele in x:
    x_sq.append(ele**2)

xy = []
for p, q in zip(x, y):
    xy.append(p*q)

n = csv_file[x_attr].count()

r = ( n*(sum(xy)) - sum(x)*sum(y) ) / math.sqrt(
(n*sum(x_sq)-sum(x)**2) * (n*sum(y_sq)-sum(y)**2) )
print('r = ',r)
```

Dependencies

Python 3+

Pandas

Run via `python3 main.py`

References

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html