# -\*- coding: utf-8 -\*-

"""

Robert Cocker - 1441

DATA 3461 - Machine Learning

Dr. Jawad

2/14/2025

Quiz 4

"""

# Quiz 4

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import statsmodels.api as sm

# Load input data

df = pd.read\_csv("C:/Users/Robert/Desktop/longley.csv", index\_col=0)

print(df)

# Correlation

print("Correlation Coefficient\n", np.corrcoef(df.Employed, df.GNP)[0,1])

x = df.Employed # predictor (independent variabl)

y = df.GNP # response (dependent variable)

x = sm.add\_constant(x) # adds constant term to predictor

# Linear Regression Model

lr\_model = sm.OLS(y,x).fit()

print(lr\_model.summary())

# Predictions

y\_pred = lr\_model.predict(x)

# Plot the data points and regression line

plt.figure(figsize=(8,6))

plt.scatter(df.Employed, df.GNP, color='blue', label='Actual Data') # Scatter plot of actual data

plt.plot(df.Employed, y\_pred, color='red', linewidth=2, label='Regression Line') # Regression line

plt.xlabel("Employed Population")

plt.ylabel("Gross National Product (GNP)")

plt.title("Linear Regression: Employed vs. GNP")

plt.legend()

plt.grid(True)

plt.show()

"""

Inference:

We built a linear regression model with various Python libraries to predict gross national product

based on the Employed feature/variable. The model outputs summary statistics and plots utilizing matplotlib

etc. About 97% of the model's variation is explained (via R^2, Adjusted R^2) and there is statistical significance.

Based on initial observations, we can conclude that increasing employment leads to an increase in GNP based

on historical data. Further statistical inference and modeling can be done for a more robust model and conclusion.

"""

"""

# References

- Lecture, course materials

- ChatGPT

"""