

INSTRUCTIONS

This worksheet is not assessed! It has 6 questions.

In this exercise you will be examining the 'Poverty.csv' dataset to practice linear regression.

Download 'Poverty.csv' from 'Blackboard->Learning Materials->Worksheets Semester 2' save in an appropriate location. The dataset consists of 8 columns:

Variable Name	Description
BirthRt	Live birth rate per 1000 population
DeathRt	Death rate per 1000 population
InfMort	Infant deaths per 1000 population under 1 year
LExpM	Life expectancy at birth for males
LExpF	Life expectancy at birth for females
GNP	Gross National Product per capita US\$ 1 = Eastern Europe 2 = South America/Mexico 3 = Western Europe/US/Canada/ Australia/NewZealand/Japan 4 = Middle East 5 = Asia 6 = Africa
Country	Country

Write python code to perform the following tasks.

1. Load the dataset using Pandas. (You did some Pandas stuff in Practical 1).
2. Plot Female life expectancy against GNP. You will need to use Matplotlib.
3. Assuming that males and females each represent 50% of the population, calculate the average life expectancy for each country. Plot against GNP.
4. Perform a Pearson's R correlation of Female Life Expectancy against GNP. You should report your results as:
An appropriate hypothesis and null hypothesis
H0 —
H1 —
We performed a Pearson's R correlation for [First Variable] against [Second Variable]. We [rejected/did not reject] the null hypothesis at the [X]% significance level ($R^2 = X.XX$, $p=0.XXX$).
5. Compare the result from 3 to the scatter plot. Do the results of the regression match what you expect from the scatterplot.
6. Is the rate of Infant Mortality correlated with the Birth Rate (per 1000) population? Your analysis should consist of the following steps:

1. Formulate a Hypothesis and Null Hypothesis.
2. Plot your data. Are there any visible issues, e.g. outliers?
3. Carry out your statistical test. (A Pearson's R correlation)
4. Is $p < 0.05$. If so then your result is significant, otherwise it isn't
5. Report your result as shown above.
6. What does this result mean? If Infant Mortality is correlated with Birth Rate, is this relationship causal? Does this mean that Infant Mortality causes Births?