PHYS465: Coursework Exercise 1

Deadline Monday 20th Jan @ 4pm.

This coursework assesses the learning outcomes from Week 11. It is worth 20% of the overall module mark.

Instructions

- Submit your work via Moodle.
- You must submit a fully compiled `.ipynb` file which includes all codes required to replicate your results
  - **Dont forget to run every cell before submitting**
  - You must also respond to the mandatory GenAI self-assessment questionaire.
- Your submission must include text (in markup format) that describes what each cell does and summarises the conclusions
- The estimated workload for this is 4-6 hours.

Tips

- The last question of this exercise asks you identify a key result. **To do this you do not have to have completed all exercises**. This assessment is designed to test your reflections on the problem undertaken.
- Don't worry about how your code looks - marks are not given for pretty code, but rather for the approach used in solving the problem, your reasoning, explanation and answer.
- Conversely, marks will be given for well presented plots
- Explain all your reasoning for each step. A *significant fraction* of the marks are given for explanations and discussion, as they evidence understanding of the analysis.
- Include all relevant lines of code including import statements and read statements. As part of the assessment your code will be run offline.

WARNING

- This submission must be your own work. Please note the university's policy on plagiarism.
- While it is acceptable (and indeed encouraged) to share ideas, you must ensure that you do not use other people's code or text, and that the reflections are your own.
- It is acceptable to use GenAI tools for guidance on how to approach this exercise, but you must ensure that all code is written by you.
  - Should you use GenAI in this work, then answer yes to the GenAI self-assessment. You will not be penalised for this.

The Problem

The World Happiness Report is published every year by the Sustainable Development Solutions Network (SDSN). This dataset uses polling to estimate the 'happiness' of people across the globe. Download the catalogue (from 2023) here:

https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/happiness_score.csv
or from Moodle.

The catalogue contains the happiness scores (and an associated uncertainty) for each country along with 6 other variables :

- `Logged GDP per capita` : The logarithm of the countrys gross domestic product (GDP) per capita.
- `Social support` : A measure of the perceived social support in the country.
- `Life expectancy` : The average number of years a person is expected to live in the country.
- `Freedom to make choices` : The degree to which individuals think they have freedom.
- `Generosity` : The level of generosity reported by individuals in the country.
- `Corruption` : The perception of corruption within the country
- `Dystopia` : The hypothetical happiness score in a dystopian society. This is used as a reference point.

Exercise

1. Load this dataset into a pandas dataframe.

   [2 mark]

2. Perform a straight line fit between `GDP per capita` and the measured `happiness` levels using the measured uncertainties.

   [4 marks]

3. Produce a plot of `GDP per capita` v `happiness`. Include both the associated error bars and line-of-best-fit.
   A. NB: 50% of the marks will be awarded for clarity of presentation

   [6 marks]

4. Write **your own function** to calculate the χ^2 statistic between the dataset and your line-of-best-fit and calculate the reduced χ^2 statistic between the dataset and best-fitting model.

   [6 marks]

5. Based on the observed correlation, select another functional form of your choosing, perform a least-squares-fit and compare the χ^2 statistics to determine the most likely function. Plot the results.

   [7 marks]

6. Either:
   A. Split the dataset into different regions, and determine the region where GDP is most strongly correlated with happiness
   B. include multiple input parameters (e.g. `Social Support`) into the linear fit and determine the strongest correlation
   C. Note: You do not need to test all regions or all input parameters, but results should be accompanied with plots.

   [8 marks]

7. **Summary Statement**. Write a short statement (200 words max) summarising a key result from this work and the consequence of it. You may include a maximum of one figure.

   [9 marks]

Additional Marks Marks will be awarded for notebooks, codes and plots that are well explained and well formatted. In particular, attention will be given to sensible variable names, easy to follow comments and notebook structure.

   [8 marks]