PHYS465: Coursework Exercise 1

Deadline Tuesday 20th January 2026 @ 4pm.

- **Overall value: 20%**

This coursework assesses the learning outcomes from Week 11, and in particular model regression.

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 check that every every cell runs before submitting**
    - As part of the assessment your code will be run offline.
    - You *must* also respond to the mandatory GenAI self-assessment questionaire.
- The estimated workload for this is 4-6 hours.

Tips

- The last question of this exercise asks you write an interpretative statement.
    - This assessment is designed to test your reflections on your learning and ability to summarise it succiently for a non-specialist audience.
    - This question is worth 25% of the overall grade.
        - To obtain this mark, additional work beyond the scope of the exercise is expected. All working must be included in the .ipynb submission.
        - This additional work is at your own discretion. Any exploration of dataset beyond the scope of the worksheet presented is suitable.
            - **NB**: the estimated workload for the entire worksheet is 4-6 hours.
        - Markers have been asked to consider **both** additional work and insightful reflections.
            - i.e. extensive work does not guarantee a high mark.
    - The interpretative statement will be marked based on your reflections on your learning across the worksheet.
        - It is expected to include both the values that you have found and an interpretation of it in the wider context.
        - If you have not completed all exercises (or an extension) then the interpretative statement can focus on your learning:
            - e.g. which techniques were difficult, and how might you address them.
- 10% of marks are award for 'good coding practice'.
    - A particular focus for this worksheet will be on annotations, such as doc-strings, comments and markdown notes.
    - Pythonic coding is not expected, rather code that is accessible, and likely to be understandable **by you** after an extended break.
    - Marks will be deducted for unnecessary steps (e.g. `for` loops) and inaccessible coding practices.
    - Marks are also awarded for high quality visualisations, which be of extended focus later in the course.
    - Explain all your reasoning for each step. Marks are given for explanations (in markup format) and discussion, as they evidence understanding of the analysis.

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 (e.g. ChatGPT, Gemini) to produce code, but you must understand it. This module is an opportunity to learn key python libaries at the core of Data Science. Understanding these libaries now will enable you to use GenAI effectively when more advanced tasks are required.
- **GenAI cannot be used to write the final interpretative statement**
    - Grammatical and syntax checks may be performed.
- Should you use GenAI, 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_2025_resources/main/coursework/datasets/happiness_scores
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. Calculate the mean and median happiness levels across the globe.

[1 mark]

2. Use an establish python library to perform a straight line fit between `log10(GDP per capita)` and the measured `happiness` levels. Your algorithm should include using the measured uncertainties. Return the best-fitting values.

[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 and χ^2_{red} statistics between a given dataset (with uncertainties) and a input model (e.g. a straight-line)

[8 marks]

5. Calculate the reduced χ^2 statistic for your best-fitting happiness v `log (GDP per capita)` model. Return the result.

[4 marks]

6. Based on your visualisation, select another suitable functional form for this correlation. Perform a least-squares-fit and calculate the χ^2 statistic. Print and visualise the results. Which model is favoured, and why?

[10 marks]

7. Calculate the best-fit linear relationship between happiness and `social support`. Which model is favoured, and why?

[6 marks]

8. **Extension/Interpretative Statement**. Write a short statement (300 words max) summarising a key result from this work and the consequence of it. Up to two figures may be included.
   A. Suggestions for extensions include:
      a. Consider how the relationship between happiness and GDP varies globally
      b. Consider the correlation between other parameters (e.g. `life expectancy`) and happiness. Can a model with multiple parameters accurately be used to infer happiness?
      c. Note: Your tests need not be exhaustive, but results should be accompanied with plots and analysis.

[15 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.

[6 marks]

•

Total available: **[60 marks]**