

PHYS465: Coursework Exercise 2

Deadline Monday 27th Jan @ 2pm.

This coursework assesses the learning outcomes from Week 12. 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 **and** a .pdf version.
 - **Dont forget to run every cell before submitting**
 - You must also respond to the mandatory GenAI self-assessment questionnaire.
- 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 too much about how your code looks - while some marks will be given for sensible coding, the focus of this assessmnt is your approach used in solving the problem, your reasoning, explanation and answer.
- As data visualisation is a key outcome, 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

Dataset 1:

The warming of the planet as caused by human activities is one of the key aspects of climate change. The Met Office has been measuring the average temperature (combining land-surface air and sea-surface water measurements) since 1850.

Every year a new catalogue is released. The 2023 catalogue can be downloaded from here: https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/anomaly.csv or through Moodle.

For every year, this catalogue contains :

- `temp_diff` : the mean global temperature (in Celsius) compared to a baseline derived from data taken pre-1925.
- `temp_upper` : an upper estimate (temp+err)
- `temp_lower` : a lower estimate (temp-err)

Dataset 2:

Independently, the Global Carbon Budget Report has been collecting information on the total amount of greenhouse gases that are released each year.

The catalogue, tracaing how greenhouse gase emissions have changed over the past 150 years can be downloaded from here : https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/global-emissions.csv .

This catalogue is also released per country which can be found here : https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/emissions-bycountry.csv .

For every year the catalogue contains:

- `AnnualC02` : total emissions (from all sources) as measured in tonnes of carbon dioxide-equivalents
 - `AnnualC02_err` : an uncertainty on this value
 - `logC02` : the logarithm (in base 10) of this value
 - `logC02_err` : an uncertainty on this value
-

Exercise

1. Write down a suitable null hypothesis for the expected relationship between global temperature (T_{global}) and time (in years)

- *Hint: Consider not only trends but also the functional form that you are expecting.*

[2 marks]
2. Load both **global** datasets into pandas dataframes and merge them using the 'Year' key.

[1 mark]
3. Define two functions :

- $y = \text{scale} \times e^{(t-t_0)/\text{tau}} + \text{offset}$
 - $y = \text{intercept} + \text{slope} \times t$

A. Fit both to the relationship between T_{global} and time.

B. Return both the best-fit values and calculate the χ^2_{red} statistic between the data and the model.

C. Using the best-fit relationships calculate when the global temperature will first exceed 2\degree.

- *Hint: an exponential function is highly sensitive. Consider offsetting the time axis prior to fitting and using an initial estimate*
 - **NB (2025jan24)**: the uncertainty on t_0 will be huge. Either ignore it, or fix it to a value (e.g. 1850). It is also sensible to renormalise the time values: i.e. $t \rightarrow (t - 1850)/10$
 - *Hint: to include uncertainties, consider how to combine two estimates*

[6 marks]

4. Produce a plot with two axes: one showing excess temperature as a function of time, the other showing CO₂ emissions as a function of time. Include the best-fit models for both relationships.

- *Hint: Remember to include uncertainties and legends*
 - **NB (2025jan24)**: for this exercise, you need only plot one function per dataset

[5 marks]

5. Using existing tools, calculate the Spearman Rank correlation between the two datasets

[1 mark]

6. For one parameter in the exponential model, calculate the χ^2 statistic between the data and the model at multiple values, and using a look-up table, calculate the uncertainty on that parameter. Plot the results in a histogram.

- *Hint: An exponential function is sensitive, I would suggest leaving the exponent as a fixed value.*
 - *Hint: You can use the uncertainty derived from curve-fitting to estimate the values to be looped over*
 - *Hint: For speed, loop over no more than 300 values*
 - *Hint: The parameters in the loop can be stored in a dataframe or list*
 - **NB (2025jan24)** : the lookup table can be found in the week 2 lecture notes (slide 22). consider how many variables you are varying.
 - **NB (2025jan24)** : this exercise asks you to loop over a range of possible values for 1 parameter while keeping all other parameters to their best-fit values. The result will be a table of (parameter_values, chi2). The *histogram* you are required to make is a scatter plot of this table.
 - *Hint: The measured uncertainty will be small*

[15 marks]

7. Either:

A. Repeat the above test, but this time allow 2 parameters to vary. Calculate the marginalised uncertainty on each. Plot the results as a contour plot.

B. Repeat the above test, but for the relationship between log emissions and time. Vary both parameters in the linear function to calculate an uncertainty on each. Plot the results as a contour plot.

C. By identifying the countries responsible for the majority of global emissions in the 19th century, and fitting functional forms to their emissions as a function of time, determine when (with an uncertainty) their contributions to global emissions will likely fall.

D. **NB (2025jan24)**: this exercise asks you to extend the analysis of Q6 to 2 dimensions. If you find the exponential to be too unstable, you can instead use the (more stable) linear relationship between time and `LogC02` . You will not be penalised for doing this.

[12 marks]

8. **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. A significant fraction of the marks are awarded for physically interpreting the results.

[8 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]
