

# PHYS465: Coursework Exercise 1

Deadline Tuesday 27th January 2026 @ 4pm.

- Overall value: 20%

This coursework assesses the learning outcomes from Week 11, and in particular uncertainty estimation.

## 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 questionnaire.
- 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 sufficiently 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

### Dataset 1:

The warming of the planet as caused by human activities is one of the key aspects of climate change. The influence of human activity can be traced through the atmospheric content of carbon dioxide ( $\text{CO}_2$ ).

This project uses atmospheric carbon dioxide ( $\text{CO}_2$ ) measurements from the Mauna Loa Observatory in Hawaii, part of the Keeling Curve record that has monitored global background  $\text{CO}_2$  levels continuously since 1958. The aim of the project is to model the long-term evolution of atmospheric  $\text{CO}_2$ , quantify the rate at which it is increasing, and test whether the growth rate itself is changing over time using statistical model fitting and  $\chi^2$ -based inference.

Every year a new catalogue is released. The 2025 catalogue can be downloaded from here:

[https://raw.githubusercontent.com/MatSmithAstro/phys465\\_2025\\_resources/main/coursework/data](https://raw.githubusercontent.com/MatSmithAstro/phys465_2025_resources/main/coursework/data) or through Moodle.

This catalogue contains :

- `decimal_year` : the date of observation
- `co2_ppm` : the measured  $\text{CO}_2$  content
- `sigma_ppm` : the measured uncertainty

See [here](#) for additional details.

## Exercise

1. Write down a suitable null hypothesis for the expected relationship between CO<sub>2</sub> levels and time (in years)

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

[1 mark]

2. Load the dataset into a panda dataframes and visualise the data using matplotlib.

- Adjust the uncertainties to include in quadrature a systematic error-floor of 3.5 for each point.

[4 marks]

3. Considering only the data taken since the year 2000, fit a linear model. Visualise the results through the residuals from the best-fitting model.

[4 marks]

4. The dataset shows a strong seasonal dependence. To account for this we can consider the following model:

- $f(t) = a + bt + A \sin(2\pi t) + B \cos(2\pi t)$
- where  $(a, b, A, B)$  are constants. To start with, we assume fixed values of  $A = 3.0$  and  $B = -1.5$ .
- Considering only the data taken since the year 2000:
  - Write a function to calculate  $f(t)$  given input values.
  - Write a function to calculate both the  $\chi^2$  and  $\chi^2_{\text{red}}$  statistics for this model.
  - Calculate the best-fit values of  $(a, b)$  and calculate the  $\chi^2_{\text{red}}$  statistic between the data and the model considering uncertainties.
  - How do these values (and their uncertainties) compare to those determined above?
  - Visualise the data including results from the new model

[8 marks]

5. Using existing tools, calculate the Spearman Rank correlation between the time and CO<sub>2</sub> levels

[1 mark]

6. For the time dependence term, calculate the  $\chi^2$  statistic between the data and the model considering multiple values of  $b$

- Store the results in a table
- From these results calculate the uncertainty on  $b$ .
- Plot the results in a histogram.
- 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: To calculate the uncertainty in  $b$  consider the look-up table from the week 2 lecture notes, considering how many variables you are varying.

[9 marks]

7. Repeat the above test, but allow 2 parameters  $(a, b)$  to vary.

- Calculate the marginalised uncertainty on each.
- Plot the results as a contour plot.
- Visualise the results including your calculated uncertainties

[12 marks]

8. **Extension and 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:

- Consider all time-points and include an additional acceleration term to be constrained
- Vary all four parameters in the model and consider how the uncertainties vary, including the effect of the error-floor.
- 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, notebook structure and informative visualisations.

[6 marks]

•

Total available: [60 marks]