Statistics and Data Analysis

Dr. Anna McLeod

Michaelmas, 2024/25

E-mail: anna.mcleod@durham.ac.uk
Web: http://www.astro.dur.ac.uk/ dwvw48/
Office Hours: open door policy (stop by any time)
Office: Zoom or OCW120

Lecture Hours: Tue 2-4 pm
Class Room: TLC025/026

Workshop Hours: Fri 9-11 am
Workshop Room: TLC025/26

Course Description

The course is divided into four segments, each of which lasts 1 week. Each week we will have 1 lecture (Tuesdays 1-3 pm). The lecture will be in-person. After each lecture, I will release the weekly Problem Sheet. Therefore, there are four Problem Sheets (one per week), which you will have to hand in. The first three Problem Sheets are **formative** and will be released weekly each Tuesday at 3 pm, and will be due by Tuesday 1 pm the following week. The fourth Problem Sheet will be **summative**, more information about this is given below. It is mandatory to hand in all four Problem Sheets.

In addition, there is a weekly workshop. This will take place on Fridays, 9-11 am. During the workshops, we will discuss the previous week's Problem Sheets as a group (the tutorial in the first week will be on a drop-in basis with the aim of providing assistance with the first Problem Sheet).

Course Materials

- The course is based on the book *Measurements and their Uncertainties* by I G Hughes and T P A Hase, Oxford University Press. Copies are available in the Bill Bryson Library Library (shelfmark 511.43HUG) and some college libraries. An e-book version is available here.
- Lecture slides will be available on Learn Ultra (Course Materials → Data Analysis)
- Weekly formative assessments will be released on Learn Ultra as Jupyter Notebooks (Course Materials → Data Analysis → Notebook Server). Once you have finished the Problem Sheet, you will submit your Jupyter Notebook via the same server.

Prerequisites/Corequisites

Prerequisites
Programming knowledge in at least one programming language and commitment to learning C and Python independently if not known before.
Corequisites
PHYSPGNEW03 Core Ib: Introduction to Scientific and High-Performance Computing.

Course Structure

Week 1
The first lecture (chapters 1-3 in the book) will cover:

- **Introduction** and **Precision & Accuracy**. Related to these are questions 1 to 2 in the Problem Sheet.
- **Gauss & Poisson** and **Central Limit Theory**. Related to these are questions 3 to 6 in the Problem Sheet.

The slides about the *Incredible Goal* help with this week's material, and are needed to complete question 3 in Problem Sheet 2.

Week 2
The second lecture (chapters 4-5 in the book) will cover:

- **Error propagation**. Related to this are questions 1 to 3 in the Problem Sheet.
- **Residuals** and **Method of Least Squares**. Related to these are questions 4 to 6 in the Problem Sheet.

Week 3
The third lecture (chapters 6-7 in the book) will cover:

- **Arbitrary Function Least Squares** and **Errors in fit parameters**. Related to these is question 1 in the Problem Sheet.
- **How do Codes Minimise?** and **Fits Errors and Correlations**. Related to these are questions 2 to 4 in the Problem Sheet.

Week 4
The second lecture (chapters 8-9 in the book) will cover:

- **What is a Good Fit?** and **Testing Distributions**. Related to these are questions 1 to 4 in the Problem Sheet.
- **Occam's Razor** and **Visualization**. This gives you more insight into the Problem Sheet.

Remember that the last Problem Sheet is summative.

Problem Sheets

There are four problem sheets, all of these are Jupyter notebooks. *Formative* means that you do not get a grade for it but must hand it in to complete the course. *Summative* means that this is an exam, it will be graded and your grade will count towards the final grade of Core Ia.

Here is a summary of the problem sheets:

- Problem sheet 1: formative, due by Tuesday, October 15th, 1 pm
- Problem sheet 2: formative, due by Tuesday, October 22nd, 1 pm
- Problem sheet 3: formative, due by Tuesday, October 29th, 1 pm
- **Problem sheet 4: summative, this is an in-person exam on Friday, November 8th, 2-5 pm, in TLC025/026**

Like the first three formative problem sheets, the summative one will be released as a Jupyter Notebook via the same server (see above). You will have **3 hours** to complete and submit the assignment. Further details regarding the date of the assignment will be communicated in class and via Learn Ultra.