Data Analysis

xxxxxx–xxxxxxxx & Fynn Krebser–f.krebser@olympiad.ch

Experimental Training Camp 2026

1 Motivation

Measurements are prone to errors at heart. We distinguish between SYSTEMATIC and STATISTIC errors. Systematics error happen due to an imperfect model used to describe a system. They happen systematically every time we measure. Statistical errors happen, at the level of Physics Olympiads due to imprecise measurements. They can be corrected through multiple measurements.

Exercise 1.1:
Are the following errors systematic or statistical?

- Neglecting the damping of an oscillator
- Measuring length only with centimeter precision
- Noise of a oscilloscope

In reality we usually use computers to do error analysis. However you do not have access to these resources. Properly doing data analysis by hand is also rather difficult. Hence, we resort to graphical data analysis.

2 Some statistical vocabulary

The most important quantity for you is most likely going to be the MEAN.

Definition 2.1: Mean
The mean of a set of data $x_1, x_2, \ldots, x_n$ is defined as

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The spread of the data is quantified by the STANDARD DEVIATION σ. The calculation of it is not essential to know. Usually at Physics Olympiads it is given or you can estimate it.[1] However what is good to know is the error of the mean

Theorem 2.2: Standard Deviation of the mean
Given N measurements with mean $\overline{x}$ and standard deviation σ. Then, the mean has a error given by

$$\Delta\overline{x} = \frac{\sigma}{\sqrt{N}}.$$

[1] For example if you have a ruler with millimeter markings, the error is 0.5 mm

The most important conclusion from this is, that with more measurements you can get a more precise result.[2]

Exercise 2.3:
Compute the mean of the following data

$$1.3\,\text{cm}, 1.5\,\text{cm}, 1.9\,\text{cm}, 1.6\,\text{cm}, 1.7\,\text{cm}, 1.6\,\text{cm}.$$

Assuming this was measured with a millimeter scale, estimate the error of the mean.

In the above exercise, using conventional methods of computing the standard deviation, one would find the error of the mean to be 0.08 cm. This is vastly larger than using the rule of thumb before. How can we account for this? The problem lies in the nature, that the above data was not very well measured.

In practice, you have to estimate your uncertainty. For example, given a stopwatch, despite it having a precision of hundredths, you probably can say you have an error of 0.1 s to 0.2 s. When in doubt choose the more conservative option.

Notice that you most likely also cannot determine σ graphically. (You may try in the figure below)

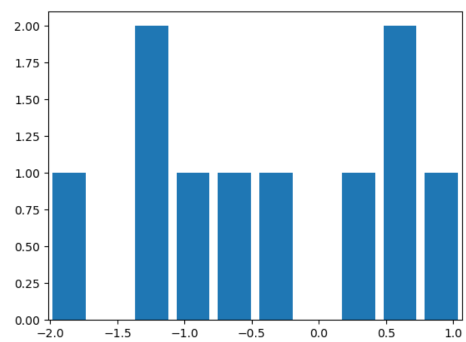

Figure 1: Histogram of a small data sample ($N = 10$)

3 The Art of Plots

3.1 Scatter Plots

Nonetheless, plotting data is still very useful to draw conclusions. The most common plot you will use is a scatter plot. When drawing a scatter plots, there are a few things to keep in mind:

[2] From this one can motivate the strategy of inventing data. Do not use this in real science.

- Always label your axes with the quantity and the unit.
- Mark your data points with a cross instead of a dot.
- Make your plot big and use a ruler.
- Do Reasonable Axes numberings. Use the 1-2-5 rule when appropriate.

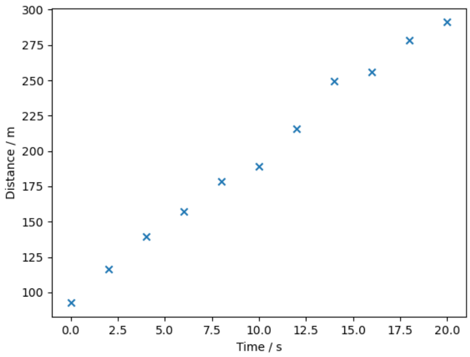

Figure 2: First example of a scatterplot.

From personal experience, error bars are not needed, as your usually not expected to measure the same quantity a lot of times (This may be different for some experiments).

How to draw conclusion from these plots is covered in the section 4.

Exercise 3.1: EuPhO 2025 E1

Find below data I measured. Plot the data in a scatter plot.

x/V	$-\ln\left(\frac{A}{y}-1\right)$
0.00	2.02
0.25	1.86
0.50	1.68
0.75	1.55
1.00	1.38
1.25	1.23
1.50	1.05
1.75	0.92
2.00	0.77
2.25	0.60
2.50	0.45
2.75	0.29
3.00	0.15

3.2 Histograms

Although rarely used in Physics Olympiads, histograms are a very useful tool to visualize the distribution of your data. In particular, they can be used to estimate the mean if you have a lot of data. For this however, you would always need to measure the same value.

For example, if you measure the gravitational acceleration g by dropping a ball from a certain height, you can do this multiple times and plot a histogram of your results. The mean of the histogram will give you an estimate for g.

Exercise 3.2: EuPhO 2024 Experiment

Find below data I measured. Plot the data in a histogram and estimate the mean for both sets of data.

h_1 / cm	$1-\frac{h_2}{h_1}$	h_1 / cm	$1-\frac{h_2}{h_1}$
20	0.35	30	0.27
20	0.40	30	0.23
20	0.35	30	0.27
20	0.25	30	0.23
20	0.30	30	0.27
20	0.25	30	0.27
20	0.30	30	0.27
20	0.25	30	0.30

4 Linearization

Working with plots drawn by hand, there are two quantities we can easily determine: The slope and the intercept. Hence, linear functions are extremely handy to work with. Fortunately, many functions can be plotted in a way that they look linear. This process is called LINEARIZATION. The following are the most common linearizations you will encounter:

- Exponential functions: $y = ae^{cx}$ can be linearised by plotting $\ln y$ against x. The slope will be c and the intercept will be $\ln a$.
- Power laws: $y = ax^q + b$ can be linearised by plotting y against x^q. The slope will be a and the intercept will be b.
- Power laws II: $y = ax^q$ can be linearised by plotting $\ln y$ against $\ln x$. The slope will be q and the intercept will be $\ln a$.

Using these linearizations, one can generate a linear relationship between the data. This can then be plotted in a scatter plot in which you can eyeball the slope and intersection.

On a side note: When measuring the slope, measure points as far apart as possible to get a more accurate result.

Exercise 4.1:

A mass m is attached to a spring with spring constant k. The period of the oscillation is given by

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

How can you determine k from a plot?

Exercise 4.2:

Add a trendline to the scatter plot in Exercise 3.1 and determine the slope and the intercept.

5 Gaussian Error Propagation

If one calculates a quantity Q from measured quantities $x, y, z, \ldots$ then estimating the error of Q is not trivial. To do this, we can use the Gaussian error propagation.

For this we first determine our error in $x, y, z, \dots$ by any method we like. Then, we can calculate the error of Q by the following formula

Theorem 5.1: Gaussian Error Propagation

Given a quantity Q which is a function of measured quantities $x, y, z, \dots$. Then, the error of Q is given by

$$\Delta Q = \sqrt{\left(\frac{\partial Q}{\partial x}\Delta x\right)^2 + \left(\frac{\partial Q}{\partial y}\Delta y\right)^2 + \dots}.$$

The notation $\frac{\partial Q}{\partial x}$ is the partial derivative of Q with respect to x. For our purposes, it is just a derivative of Q treating everything other than x as a constant. Δx is the error of x.

Exercise 5.2:

A mass m is attached to a spring with spring constant k. The period of the oscillation is given by

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

If we measure T and m with errors ΔT and Δm, how can we determine the error of k?

This is a grid to practice plotting.

This is a grid to practice plotting.

This is a grid to practice plotting.