Tutorial: Data Analysis Practice

Nikolas Provatas, Anh-Khoi Trinh, Cesar Daniel Rodriguez Rosenblueth

1 Learning objectives

- Learn how to use Excel.

- Compute certain basic mathematical statistics.

- Introduction to data analysis.

2 Introduction

This lab will be an interactive demo lab where teaching Assistants will help you learn how to use Excel and to apply it to the concepts presented in the **Data analysis** chapter. We recommend that you read the **Data Analysis** and familiarize yourself to Excel before attending this session. The first part is a mandatory tutorial session and will be scheduled in the Lab information document on MyCourses.

3 In-class tutorial

During the in-class tutorial session, you will familiarize yourself with Excel. You will be guided through this exercise by the TAs.

- Open the Excel document found at [location].

- Explore the user interface. Note that data is best presented in columns.

- Learn how to resize, split and merge cells.

- In a adjacent column, calculate the resistance R.

- Calculate the sum of resistances R by using =SUM(x).

- Calculate the average resistance $\bar{R}$ in two ways: by using the integrated function =AVERAGE(x) and by dividing the sum that you calculated previously by the total number of cells.

- For each value of resistance R_i, calculate the difference

$$R_i - \bar{R}, \tag{1}$$

  in a separate column by "dragging" the first cell downwards.

- Calculated the square of the difference calculated above in another column.

- In two other columns, calculate the standard deviation by using eq. (??) with the values that you calculated previously, and by using the function =STDEV.P(x).

Now we will learn how to plot data, and how to obtain a linear regression.

- Select the columns of I and V, perform a scatter plot.

- Put in manual error bars.

- Display a linear trendline with the fit equation and its R^2 value. The fit coefficient is your "fit resistance" R_{fit}.

Now we will compare this fit equation to our data.

- To start, copy the I column to another area of your spreadsheet.

- In the column next to it, write the fit equation such that this will be your "Fit V" column.

- Plot V vs I, and "Fit V" vs I on the same graph. Add error bars only on the V vs I dataset and compare the two graphs qualitatively.

Let us now compare resistances.

- On another area of your spreadsheet, write a column with the current values I.

- In an adjacent column, rewrite the resistance R_i calculated previously with its corresponding standard deviation.

- In another column, write the average resistance $\bar{R}$ for all values of I.

- In another column, write the fit resistance R_{fit} for all values of I.

- Plot R vs I for all three datasets: the measurement R_i, the mean $\bar{R}$ and the fit R_{fit}. Compare them.
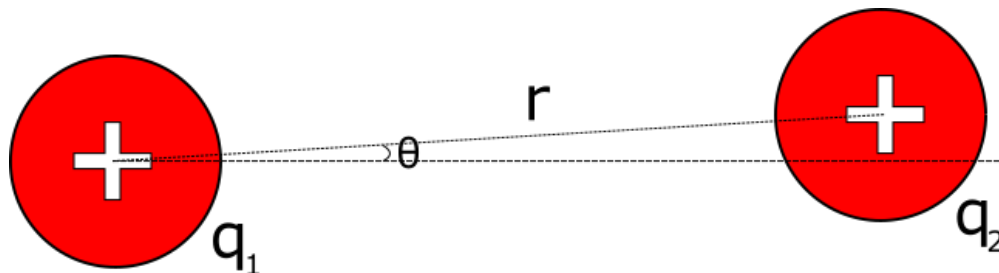
Figure 1: Setup for the data of Session 2. Two charges q_1 and q_2 are separated by a distance r, and at an angle θ with respect to some axis horizontal to q_1.

4 Homework Exercise

As homework, you will analyse a given dataset. Consider a setup where you can measure the electric force between two charges as shown in Fig. 1.

Given this dataset, you must find how the electric force on q_1 is affected by the following parameters:

- Charges q_1 and q_2.

- Distance r between the charges.

- The angle θ with respect to a horizontal axis separating q_1 from q_2.

You are given a .txt files from which you are to import them into Excel. For all the .txt files, the charges are in Coulomb units, the forces are in Newton, the angle is in radians and the distance is in meters. [Change .txt files to have the constant variables and the standard deviations.]

We suggest that you start your analysis for fixed r, θ, but varying q_1 and q_2. Find how the force F depends on q_1 and q_2. Support your conclusion by performing appropriate fits to your equation and comparing them to the error bars. If you find that you can simplify their dependence by

$$f(x, y) = x \times y, \tag{2}$$

or

$$f(x) = \frac{1}{x}, \tag{3}$$

perform the appropriate error propagation equation to find the effective error of this composite quantity.

After you have found the appropriate q_1 and q_2 dependence, find the dependence on θ by using appropriate data analysis techniques.

Finally, find the appropriate dependence on r. Note that you have 3 different trials for this experiment: you must analyse all three and compare your findings to see if your analysis is self-consistent.

After analysing the effect of the all previous variables, state the equation of the electric force $F(q_1, q_2, r, \theta)$ that best matches your findings.

Given your equation, what happens if the radius r is much larger or smaller than $q1$ and q_2? Is this what you expect physically to occur?

Your methodology can only find the dependence on the explicit variables that you analyse, but the physics can be shifted by an overall constant:

$$F(q_1, q_2, r, \theta) = k \times f(q_1, q_2, r, \theta) + c. \tag{4}$$

What is the significance of c? Given all your previous datasets, find the constant factor k. If you found the right dependence on your variables, this constant should be the same for all your datasets.