Exercise 4: Data analysis

The report must be delivered within 15 days after the corresponding practical lessons.

1. **Analysis of electoral data:** according to the paper of Araújo et al (PLOS ONE 5(9): e12446), the majority of elections are decided by a small number of votes. Using data in annex (column 1: Obama, column 2: McCain):
    a. Compute the average, median, and variance of votes for each candidate.
    b. Reproduce the histograms in figure 2 of the paper (plot both histograms in the same figure).

2. **Evolution of Sun spots:** Sun spots are measured by the Wolf number, given by $k(10g+f)$, where k is a scaling factor, f is the number of spots, and g the number of groups of spots. The monthly measures of Wolf number is found in an annex (columns: year, month, Wolf number).
    a. Plot the autocorrelation function (original series plus autocorrelation).
    b. Identify the periodicity of the series.
    c. Apply the seasonal differences in order to subtract the identified periodicity. Plot the original Wolf number series and the seasonal difference series (both curves in the same plot).

3. **Animal metabolism:** the following table presents a measure of metabolism of some animals as a function of their weight.

Animal	Mass (kg)	Metabolism (W)
Cow	400	270
Human	70	82
Sheep	45	50
Chicken	2	4,8
Mouse	0,3	1,45
Pigeon	0,16	0,97

    a. Compute a linear regression, applying the least squares method, with power function with metabolism as a function of mass. Compute the same function with the LinearModelFit[] function in Mathematica and compare.
    b. Plot the input data and fitted function.
    c. Compute a least squares fit with a power function, to the same data, using the gradient descent method. Use a maximum of 100 iterations, a precision goal of 10^{-6}, and a step size of $\lambda=\{0.01,0.05,0.1\}$. Discuss the resulting coefficients.

4. **Analysis of porosity of some material:** consider the image given in annex (rocks.jpg).
    a. Import the image, in RGB format, and convert it to grey scale. Plot both images, input and grey scale, side by side.
    b. Compute the histogram of intensity (of grey). Compute a binary version of the grey scale image by applying a threshold that separates objects from background. Determine the threshold from the histogram.
    c. Compute the porosity of this material knowing the porosity is given by $n=\frac{A_{empty}}{A_{total}}\times 100$, where A_{empty} is the area of empty space and A_{total} is the total area.

Exercise 4 (optional): Data analysis

This part is optional and doesn't need to go into report.

5. **Analysis of electoral data:** according to the paper of Araújo et al (PLOS ONE 5(9): e12446), the majority of elections are decided by a small number of votes. Using data in annex (column 1: Obama, column 2: McCain):
    a. Verify is the same happened in the last American presidential elections.

6. **Evolution of Sun spots:** Sun spots are measured by the Wolf number, given by $k(10g+f)$, where k is a scaling factor, f is the number of spots, and g the number of groups of spots. The monthly measures of Wolf number is found in an annex (columns: year, month, Wolf number).
    a. Compare the temporal series of the Sun spots with the temporal series of the average temperature on Earth and measure the Pearson correlation.

7. **Frequency of earthquakes:** Making use of the function EarthquakeData[] in Mathematica:
    a. Import the earthquake data of an arbitrary year.
    b. Find the amplitude of the largest earthquake in that year.
    c. Compute the histogram of earthquakes and verify the Gutenberg–Richter law.
    d. Explore the financial data available in Mathematica and seek for similar correlations.

8. **Animal metabolism:** the following table presents a measure of metabolism of some animals as a function of their weight.

Animal	Mass (kg)	Metabolism (W)
Cow	400	270
Human	70	82
Sheep	45	50
Chicken	2	4,8
Mouse	0,3	1,45
Pigeon	0,16	0,97

    a. Compute a fit directly to the power function with the gradient descent. Compare with previous results.

9. **Analysis of porosity of some material:** consider the image given in annex (rocks.jpg).
    a. Use the Sobel edge detector to locate the contours in the image. Filter with a Gaussian to improve the result. Show the resulting image.