Homework #1b: Model Fitting

Math 4334: Mathematical Modeling
Dr. Scott Norris

August 20, 2023

Problem 1. Suppose we are looking at data sets $\{x_i, y_i\}$ that seem to display *proportionality* – i.e. they seem to be well-described by a model of the form

$$y = Ax$$

(a) write down an expression for the residual components $\{r_i\}$ associated with this model

(b) by minimizing the sum of squares $E_2^2(A)$ of the residual, obtain a general formula for A

(c) apply your formula to the data set shown below, to determine A. You may leave A as a fraction.

x	0	1	2	3	4	5
y	0.1	0.4	0.6	1.0	1.1	1.6

Problem 2. For each of the four files of the form 'hw-01b-datasetX.xlsx':

(a) plot the data together with a best-fit linear model of the data.

(b) calculate the UVR/EVR for the model, and plot the standardized residual

(c) discuss whether any of the data points might be considered outliers

(d) characterize the "magnitude / structure quadrant" of the residual

(e) describe whether you would seek a better model: why or why not?

Problem #3. In this problem we will obtain and explore some data related to climate change.

(a) Find and download data on CO2 levels for the years 1850-present. Then find and download data on the global mean temperatures for the years 1850-present.

(b) In the field of climate science, the convention is to express temperature as the "Global Mean Temperature Anomaly" (GMTA) which is the *difference* in degrees Celsius between the global mean temperature now from its average value over some reference period. Different data sources use different reference periods, but we will follow the convention of the IPCC and use the years 1850-1900 for this purpose. Thus, if your data source uses a different range, just compute the average temperature over the years 1850-1900, and subtract that number from our temperature data. In addition, we know that pre-industrial CO2 levels were about 280 ppm, so we will subtract 280 from our CO2 values to obtain "excess CO2" (ECO2).

(c) Create a plot with the ECO2 levels vs time on the left, and GMTA levels vs time on the right. Note that the data on GMTA is much noisier over time than are the data on ECO2 levels, because GMTA is affected by several important factors besides ECO2 (such as El Nino cycles).

(d) Next, eliminate the time variable by plotting the relationship **between GMTA vs ECO2**. The data is now in the same format as other exercises we have done in Unit #01. Does there appear to be a strong correlation, that could be fit by a simple, linear model?

(e) For any years that look significantly different than their neighbors, determine whether unusual circumstances may have applied. If you find such combinations, you may remove those years from your data. Justify each removal.

(f) Obtain a linear fit between ECO2 levels and the GMTA. State the fitted model and its coefficients. Discuss the value of R^2, and analyze the residual as we did in unit 1b.