Exercise 1: Salmon Runs

Download the file **salmon_dat.csv** included with the homework. This file contains the annual Chinook salmon counts taken at Bonneville on the Columbia river from the years 1938 to 2014 (*www.cbr.washington.edu*). You can load this file into MATLAB using `>>load salmon_data.csv`. **Do not upload this file to Scorelator, your code will be tested using the counts for another species of salmon.**

(a) You should begin by creating a time vector `>>t=(1:length(salmon_data)).'` and plotting the salmon counts against the year in which they were taken (plotting usually helps you get a better understanding of the data), but make sure to comment out the plot before submitting. Note that we have let 1938 be represented by year 1 (making 2014 year 77), continue with this convention in the rest of the problem. In the video lectures, you learned that the following matrix equations could be used to determine the coefficients of a linear best-fit:

$$\underbrace{\begin{bmatrix} \sum_{k=1}^{N} t_k^2 & \sum_{k=1}^{N} t_k \\ \sum_{k=1}^{N} t_k & \sum_{k=1}^{N} 1 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} A \\ B \end{bmatrix}}_{P} = \underbrace{\begin{bmatrix} \sum_{k=1}^{N} t_k y_k \\ \sum_{k=1}^{N} y_k \end{bmatrix}}_{R}$$

where y_k is the k-th index of the salmon count data. The solution provides A and B such that $y = At + B$ is the RMS best-fit line. Compute the Q, R, and P matrices and save them in **A1.dat**, **A2.dat**, and **A3.dat**. You can confirm that your A and B are correct by finding a first order fit using `polyfit`.

(b) Use `polyfit` to find the best-fit polynomials of order 2, 5, and 8, and save these coefficients in **A4.dat**, **A5.dat**, and **A6.dat**, respectively. You should plot each of these best-fits, but make sure to comment out the plot before submitting.

(c) Using each polynomial fit from (b), predict the salmon counts in 2015. Save the predictions from each polynomial fit in a column vector with 3 elements in **A7.dat**. If you are curious, look at the website provided above to see if the predictions were accurate!

(d) We will now use this data to study interpolations. Create coarse vectors of time and the salmon data which contain the first element and every fourth subsequent element. For example, the coarse time vector would begin with `[1; 5; 9; 13; ...]`. Save the coarse salmon data as **A8.dat**.

(e) Now, interpolate this coarse data onto the original time vector using each of the following methods: nearest neighbor, linear, cubic, and spline. Save the interpolated salmon count values for these methods as column vectors in **A9-A12.dat**.

(f) For each of these interpolation methods, compute the RMS error between the interpolated values and the true values using

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}$$

where $\hat{y}$ represents the interpolated values. Save the RMS errors for each method in a single column vector with four elements in **A13.dat**.