Least Squares Regression 2 **Math 342**

You can use the method of least squares to find the coefficients of an approximating function, as long as it is a linear function of the coefficients: $\hat{y} = b_0 + b_1 f_1(x) + \ldots + f_{n-1}(x)$.

1. Over the last 60 years, the number of people killed by lightning in the United States has dropped dramatically. You can find the data in this spreadsheet:

   `http://people.hsc.edu/faculty-staff/blins/StatsExamples/Lightning.xlsx`

   Load the data using the Pandas library:

   ```
   import numpy as np
   import pandas as pd
   import matplotlib.pyplot as plt
   df = pd.read_excel("http://people.hsc.edu/faculty-staff/blins/StatsExamples/Lightning.xlsx")
   x = np.array(df.year)
   y = np.array(df.deaths)
   plt.plot(x,y,"o")
   ```

   Once you have the data loaded and plotted (hint you can copy and paste the code above from my website), then solve the normal equation to find the coefficients of the least squares regression line.

2. A more accurate model for the trend in lightning deaths might be an exponential decay model where $\hat{y} = Ce^{bx}$. Apply an appropriate log-transform to the data to find the least squares exponential decay model.

3. A power law model is one with the formula $\hat{y} = kx^{\alpha}$. Find the best fit power law model for this data from the 2000 Census which shows the top 10 most common last names in the United States. Use rank as the x-value and the number of occurrences as the y-value.

Name	Rank	Frequency
Smith	1	2,376,206
Johnson	2	1,857,160
Williams	3	1,534,042
Brown	4	1,380,145
Jones	5	1,362,755
Miller	6	1,127,803
Davis	7	1,072,335
Garcia	8	858,289
Rodriguez	9	804,240
Wilson	10	783,051

You can find this data in this spreadsheet:

`http://people.hsc.edu/faculty-staff/blins/StatsExamples/Surnames.xlsx`