Project 5: Avg Temperature in NYC

By: Connie Zhang, Grace Qu, Tyler Neath, Juliet Giraso, Helen Cao

Introduction

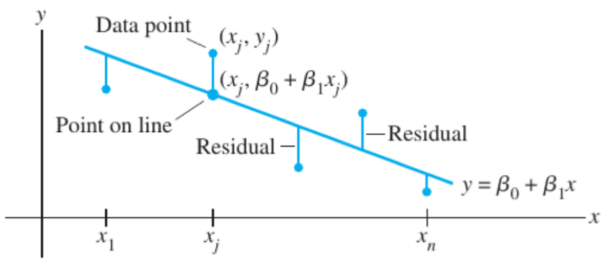


FIGURE 1 Fitting a line to experimental data.

What is Curve Fitting?

Curve fitting is the process of finding the line of best fit of a given set of data points.

What is Least Squares?

Least squares is a way to find the line of best fit.

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

Data Collection Process

SOURCES:

U.S. climate data
Temperature - Precipitation - Sunshine - Snowfall

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

- SELECTED NEW YORK AS OUR STUDY AREA
- CHOOSE THE MONTH OF AUGUST FOR THE ANALYSIS ACROSS THE YEARS (1895-2019)
- USED THE AVERAGE TEMPERATURE OF AUGUST FOR EACH YEAR
- REPRESENTED THE DATA IN A TABLE

Dataset: Sample Average August Temps

1800s

Year	*Average*
1895	74.6
1896	74.9
1897	71.8
1898	74.4
1899	74.4

1900s

Year	*Average*
1922	72.2
1923	72.1
1924	73.8
1925	72.9
1926	73.5

2000s

Year	*Average*
2015	79
2016	79.2
2017	74
2018	78.1

Code Dev

Decided Programming tool: MATLAB

Reasons:

1) Flexibility in Code/ Data import
2) Built-in Functions
3) Data visualisation

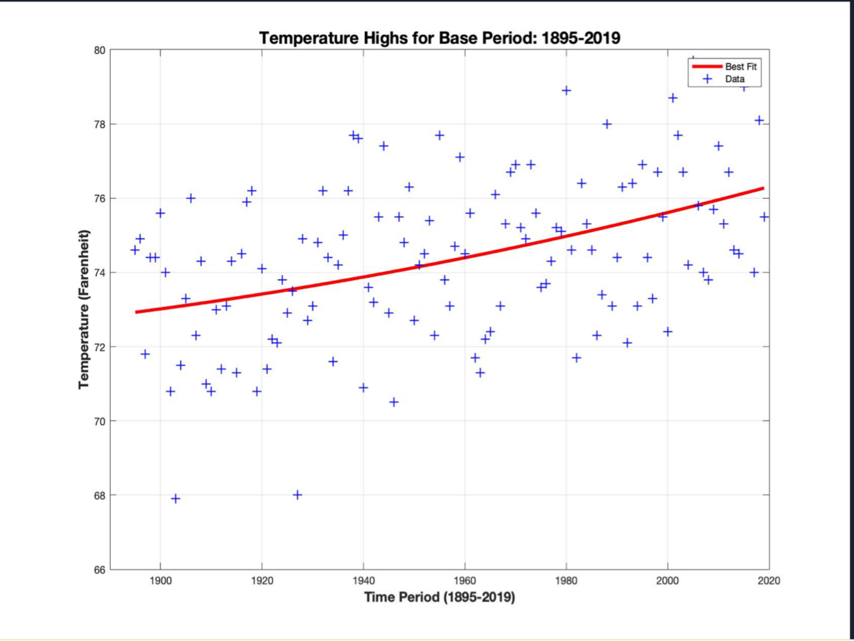
Least Squares Implementation

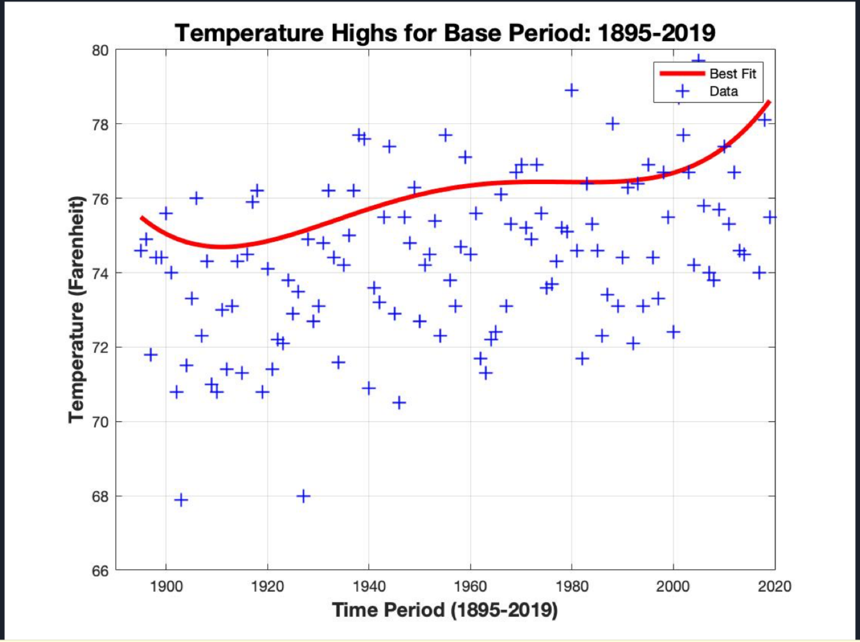
$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$$

```
Y = temp'; % tranpose
X = time';
H = [ones(length(Y),1),X];
Astar = inv(H'*H)*H'*Y;

Y2 = H*Astar;
```

Curve Fitting

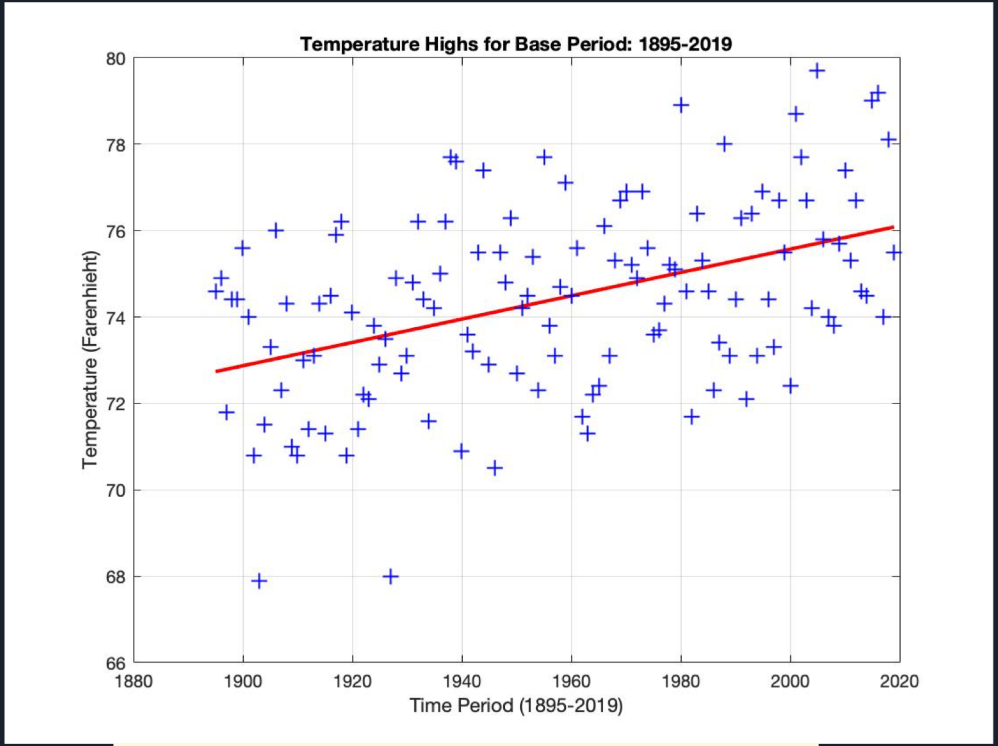





```
H_2nd = [ones(length(Y),1),X,X.^2];
Astar_2nd = inv(H_2nd'*H_2nd)*H_2nd'*Y;
```

```
H_fifth = [ones(length(Y),1),X,X.^2,X.^3,X.^4,X.^5];
Astar_fifth = inv(H_2nd'*H_2nd)*H_2nd'*Y;
```

Final Linear Result



$y = 0.0270x + 21.5816$

Conclusion

Generally, there is a positive correlation between the time and temperature.

Factors affecting data analysis:

1. Sample size; Allowing more variation;

2. Curve-fitting method chosen;