Name: Hoyoung Kim, Christopher Keating

Google the average daily temperature for the last year in Orem or SLC, UT. This will serve as the

population.

1) Compute the mean and variance of the population.

Mean = 55.5616

Var = 437.7194

2) Use the method of sample of convenience and take 20 samples. Then, compute the sample mean

and sample variance.

Var = 50.0526

Mean = 31.5000

3) Use the Simple Random Sample (SRS) method and randomly select 20 samples. Then,

compute the sample mean and sample variance.

mean = 49.2857

var = 228.9143

4) Discuss the results obtained in parts (1), (2) and (3).

As we can see, the mean and variance of the whole year of 2022 is

Mean = 55.5616

Var = 437.7194

While the mean and variance of sample of convenience of the 20 data is

Var = 50.0526

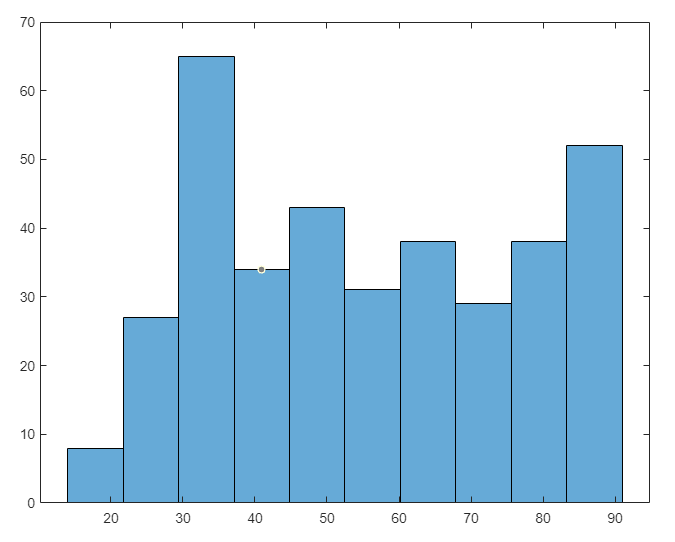
Mean = 31.5000

And mean and variance of 20 random data of the graph are

mean = 49.2857

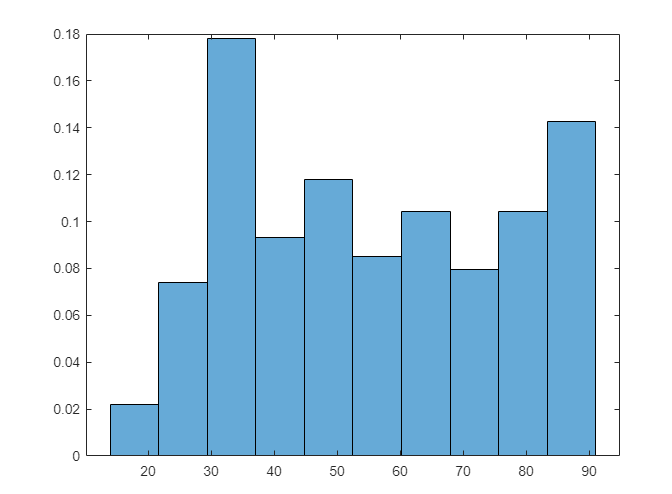
var = 228.9143

5) Plot the histogram of the population (set the length of bins to 10 degrees). Discuss what it shows.



This shows unnormalized data of the temperatures in 2022. This does not show by month but rather by the amount of the temperature that happened in the year

6) Normalize the histogram data and plot it.



7) Plot the temperature population data. Use appropriate labels for the x- and y-axis (x demonstrates

the day (eg., 1, 2, …) and y-axis shows the temperature). In the title of this figure, state when is

the starting and ending date of your data.

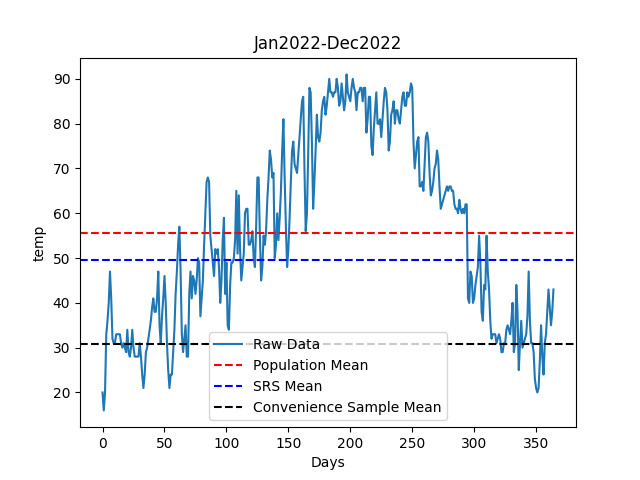
A picture containing text, plot, screenshot, font

Description automatically generated

8) On the same figure, show the population mean with a red line, the mean obtained from the SRS

method with a blue line, and the mean from the method of sample of convenience with a black

line. Use appropriate legends.



9) Predict the average temperature for Sept. 1, 2023 based on the collected data using any

polynomial that you think better fits the temperature data, and then find the coefficients using the

least squares method. Include your code in your report and state what model you considered.

Using 3rd degree polynomial, we get “a \* day\*\*3 + b \* day\*\*2 + c \* day + d” with this, we can find the temperature of the particular day (September 1) to be 244th day. Using this, we find the polyfit with 3rd degree, days (365) and the temperature of the days and find the coefficient of the least square method and when the values are added to the equation, we get 76 degrees.

Linear regression could be considered as well but this would assume that the temperature and days would have a linear relationship which would not be true.

def predict\_temperature(day, coefficients):

    a, b, c, d = coefficients

    return a \* day\*\*3 + b \* day\*\*2 + c \* day + d

def main():

    temp = []

    with open("avetemp.csv","r") as data:

        for line in data:

            temp.append(float(line.split("\n")[0]))

    day = 244

    days = np.arange(len(temp))

    coef = np.polyfit(days, temp, 3)

    predicted\_temp = predict\_temperature(day, coef)

    print(predicted\_temp)