# Computer Assignment (CA) No. 2: REGRESSION and HISTOGRAM

Truc Le

ECE 3512: Signals – Continuous and Discrete

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

# Problem Statement

The computer assignment number two we has been introduced were regression and histogram. Based on what we have done on the previous computer assignment, we will be building on the result of CA number 1 with the identical speech signal file. From this computer, we will be asked to plot different types of functions to compare results based on their statistical data. The assignment consists of two part, for the first part, we need to plot a linear regression model with frame size of 1 and a window size of 7 of a Google stock data, then compare it with the plot of the mean values. For the second part of the assignment, we need to plot a histogram of the speech file and the cumulative distribution function, CDF, of it with a bin size of 10.

# Approach and Results

For part one of the assignment, we need compare the distribution of different plots. From Figure 1, it displayed the statistical data for the close column from the excel worksheet in linear regression function plot along with its root mean square. The google stock price could be determine through a linear formula by: y = a\*x + b. But instead of x, it is considered to be in term of time, so denote SP to be stock price: so SP = a\*t + b, where,

And with the linear regression function, its plot produced all data in a smoother and increasing straight line pattern compared to the plot above with RMS vs. Time which is much rougher.

For part two of the assignment, we will be loading the speech signal file and observed the histogram and the Cumulative Distribution Function. The cumulative distribution function of a random variable X is the function given by: Fx(x) = P(X<= x), where the right side of the formula represents the probability that the random variable X takes on a value less than or equal to x. From Figure 2, we could observed that the probability of Cumulative Distribution Function tends to go to 1 as the values goes to positive. As for the histogram section, we need to acquire the histogram of the speech signal file with a set of range from -45 to +45. The plots of the histogram are shown from Figure 3 to Figure 11, we noticed from the histograms, the maximum value tend to occur at higher range.

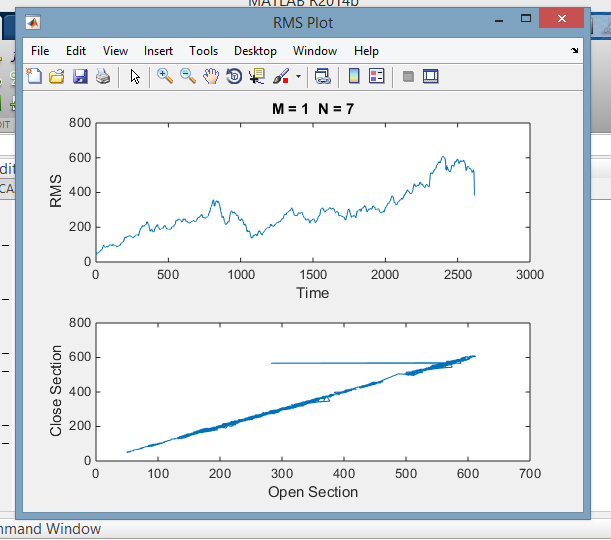


Figure 1: Signal Data Plot vs. Linear Regression Function

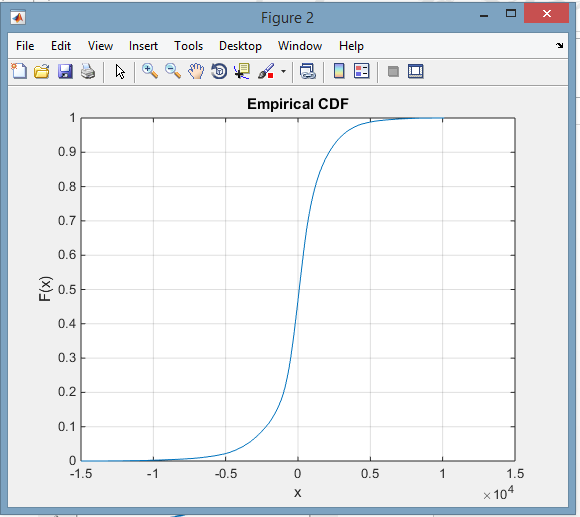


Figure 2: CDF of the Speech Signal File

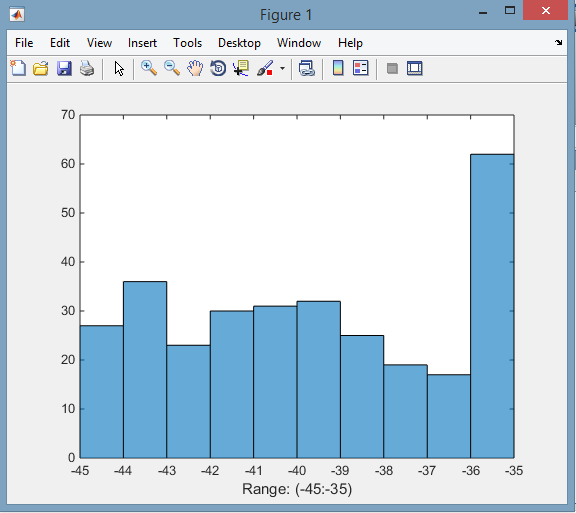


Figure 3: Range from (-45, -35)

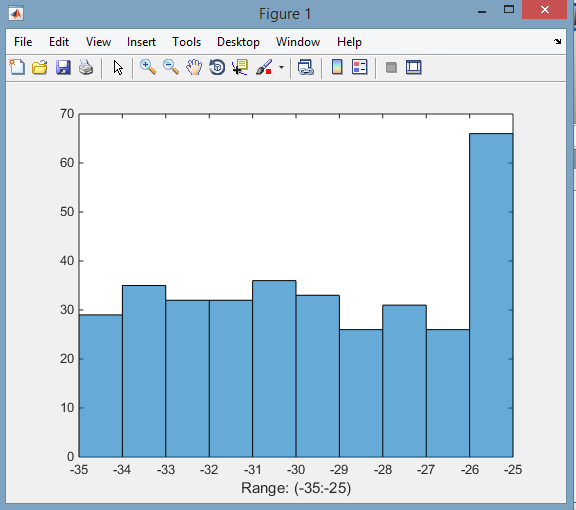


Figure 4: Range from (-35, -25)

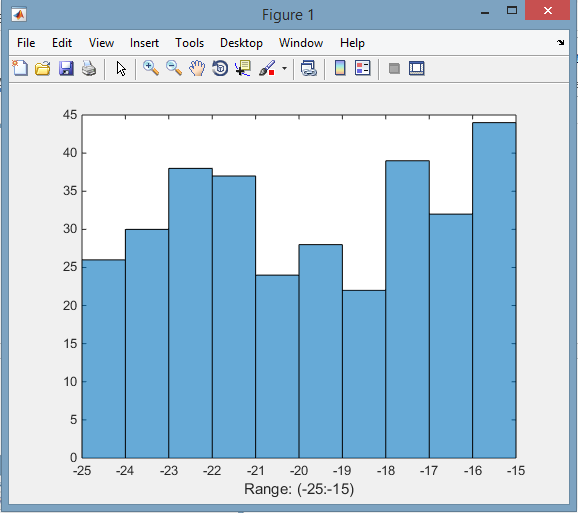


Figure 5: Range from (-25, -15)

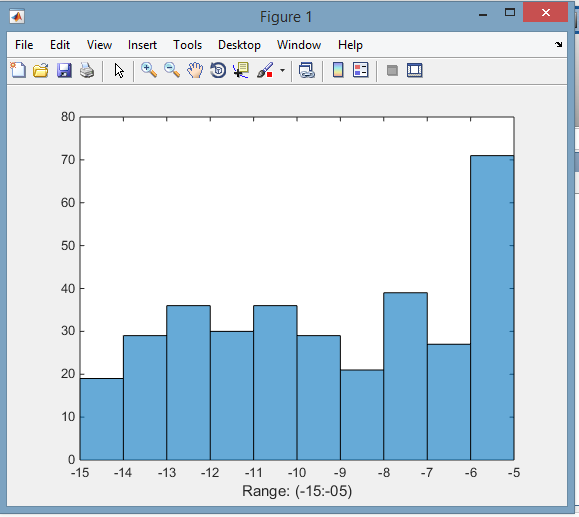


Figure 6: Range from (-15, -05)

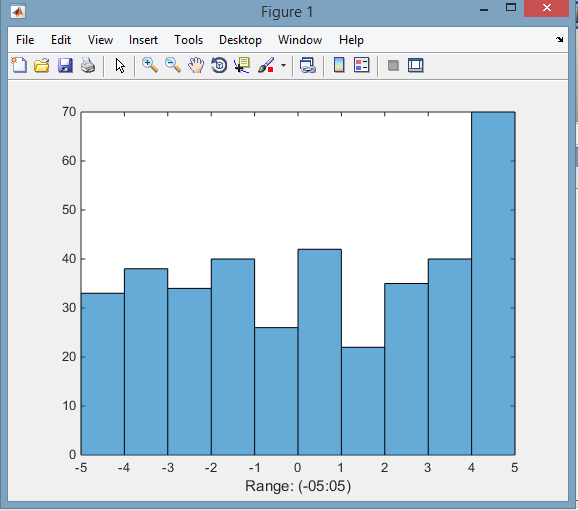


Figure 7: Range from (-05, 05)

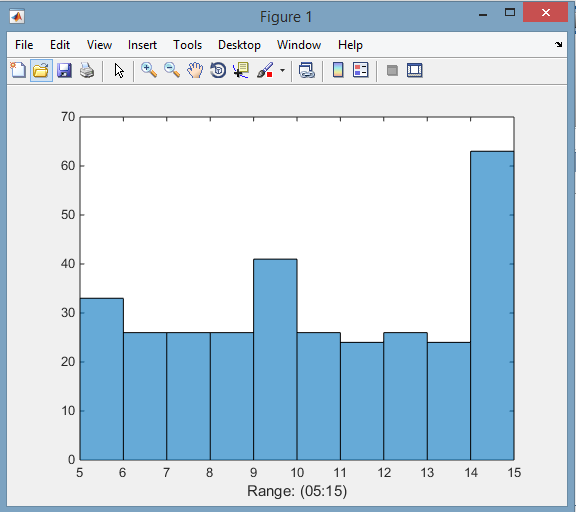


Figure 8: Range from (05, 15)

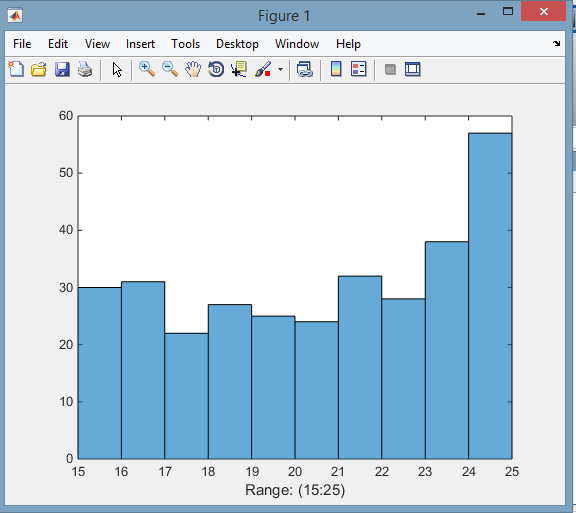


Figure 9: Range from (15, 25)

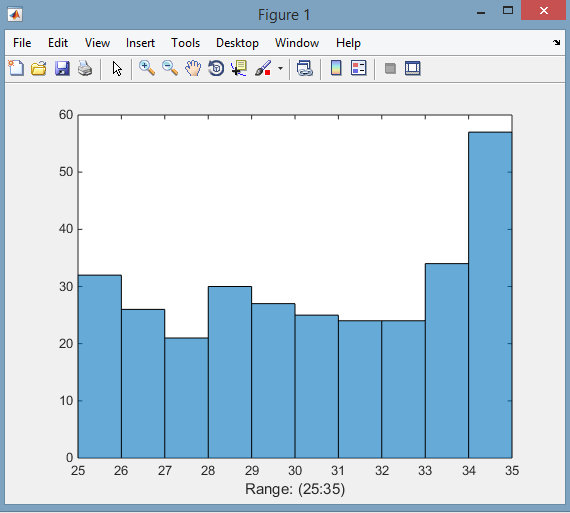


Figure 10: Range from (25, 35)

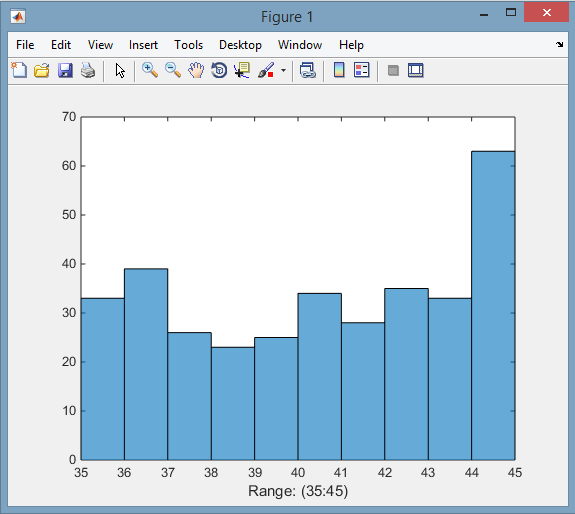


Figure 11: Range from (35, 45)

# MATLAB Code

Part l: Signal Plot & RMS & Linear Regression of Google Stock Data

First of all, we are going to load the google stock data from the excel worksheet with the command: data = xlsread('google\_v00.xlsx'); We then, created a number of loop to generate the window manually as wanted with the frame size of 1 and window size of 7. And plot the RMS vs Time with the Linear Regression Function plot of the open and close section of the excel worksheet.

MATLAB CODE:

%Part 1: Signal Plot & RMS & Linear Regression of Google Stock Data

%opening / reading / and closing files

close all;clear;clc;clf;

%loading google stock data from excel

data = xlsread('google\_v00.xlsx');

%define M and N

M = [1];

N = [7];

%loop for M

for m = 1:length(M)

Fig = figure('name', 'RMS Plot', 'numbertitle', 'off');

%M

%loop for N

for n = 1:length(N)

st = 1;

%N

%main loop for google data

for main = 1:round(length(data)/M(m))

%frame position

p = round(M(m)\*(main-1)+M(m)/2);

p\_Begin = round(p - N(n)/2);

p\_End = round(p + N(n)/2-1);

%check pointer begin

if(p\_Begin < 1)

p\_Begin = 1;

end

%check pointer end

if(p\_End > length(data))

p\_End = length(data);

end

%Generate windows of the begin pointer

%and end pointer

winvector = (p\_Begin : p\_End);

Average = sum(data(winvector).^2 / N(n));

%create repetitions of M

for M\_repeat = 1:(M(m))

RMS(st) = sqrt(Average);

st = st + 1;

% check that length doesn't not overpass start's value

if st > length(data)

break;

end

end

end

while st <= length(data)

RMS(st) = 0;

st = st + 1;

end

%plot all values

figure(Fig);

subplot(1 + length(N), 1, n);

plot(RMS);

Display = sprintf('M = %d N = %d',M(m),N(n));

title(Display);

xlabel('Time');

ylabel('RMS');

%plot of signal

figure(Fig);

subplot(1 + length(N), 1, n + 1);

plot(data);

title('Signal');

xlabel('Time');

ylabel('Amplitude');

end

end

%Linear regression

clear;clc;

data = xlsread('google\_v00.xlsx');

%set of data from open and close section

x = data(:,1);

y = data(:,4);

%Linear regression formula and plot

%[a0, a1] = func(x , y);

plot(x,y)

xlabel('Open Section');

ylabel('Close Section');

Part ll: plotting the histogram plots and the cumulative distribution file of the speech signal.

We would first open and read the speech signal file using the commands: Faudio = fopen('rec\_01\_speech.raw','r'); signal = fread(Faudio, inf, 'int16');

fclose(Faudio); Then, use the command: histogram(signal,35:45), to make a plot with a variety set of range. Finally use command: cdfplot(signal), to plot the cumulative distribution function of the speech signal file.

MATLAB CODE:

%opening / reading / and closing files

close all;clear;clc;clf;

%open and read the speech signal file

Faudio = fopen('rec\_01\_speech.raw','r');

signal = fread(Faudio, inf, 'int16');

fclose(Faudio);

%display figure 1

figure(1)

%to change the range of the bin (b,x:x)

histogram(signal,35:45)

xlabel('Range: (35:45)');

%display figure 2

figure(2)

cdfplot(signal) %Plotting CDF of speech file

# Conclusions

Linear regression was used extensively in real-world applications. It is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine. By using the linear regression, we would be able to examine a linear line behavior of unsorted collection of data and to observe the increasing of the statistical data of the linear line. Then again, by plotting the Cumulative Distribution Function of the speech signal file, we could observed the probability of it. As for the histogram plot, with different ranges, we could perceived how the histogram depicted and where the higher value are tend to located at.