ME 3300 Experiment-01: Introduction to Data Analysis and Visualization in MATLAB

Dr. Christopher Bitikofer

Contents

1 Learning objective **2**
- 1.1 Objectives . . . 2
- 1.2 Check your Understanding . . . 2

2 Datasets & Code Resources **3**

3 Laboratory Introduction **3**

4 Part-1: Scatterplot and Histogram from Voltage data **3**
- 4.1 Detailed Instruction . . . 4
- 4.2 Plot Format Instructions . . . 5
- 4.3 Example Results . . . 6

5 Part-2: Linear Curve Fit on Scatter Data **7**
- 5.1 Detailed Instruction . . . 7
- 5.2 Plot Format Instructions . . . 8

6 Part-3: Generated Histogram and PDF Lines **9**
- 6.1 Detailed Instruction . . . 9
- 6.2 Plot Format Instructions . . . 9

7 Post-Lab Assignment **11**
- 7.1 General Instruction . . . 11
- 7.2 Submission Items . . . 11
- 7.3 Questions . . . 11

1 Learning objective

1.1 Objectives

Your objectives for this laboratory session are to:

- Learn how to read the data files acquired during the experiment
- Learn about plotting using MATLAB, including:
  - Setting proper text in x and y labels and setting font size
  - Setting the size of the figure to be used in reports
  - Saving the figure files in .pdf and .png format with 600 dpi
  - Learning about different plot types and formatting options in MATLAB
- Learn about curve fitting in MATLAB
  - Compute and interpret the standard error of fit and the standard error of the linear fit slope.

1.2 Check your Understanding

By the end of this lab, you should be able to answer all of these questions.

- How can I access help documentation in MATLAB?
- How can I find the value of the 50^{th} data point in a vector or array?
- How can I view the values of data points 4 through 8 in a dataset?
- What function is used to open or load a data file in MATLAB?
- What is the standard file format for saving figure files in this class?
- How can I set the size (width and height) of a figure for reports and submissions?
- What command is used to set the DPI (dots per inch) of a figure?
- How can I change the font style and font size used in a figure?
- How do I position a legend in the northeast corner of a plot in MATLAB?
- How can I find the size (dimensions) of a data matrix?
- What MATLAB function is used to perform linear fitting or linear regression?
- What functions do I use to calculate the average and standard deviation of a dataset?

2 Datasets & Code Resources

In this laboratory session, you will use MATLAB for data analysis and creating visually appealing and communicative plots. Before the lab, please review MATLAB's online learning resources. The MATLAB On-Ramp and Core Skills courses both contain useful background on using data to produce figures. You should refer to these courses as necessary throughout the semester.

Throughout this course, you'll be working with code and some example datasets posted on Canvas. Weekly lab sections will often include example codes to help you get started with each lab. While it might be tempting to copy and paste from the example files, You are strongly encouraged to **hand-type the code**, especially in the beginning, as a way to build familiarity with MATLAB syntax and tools. Once you understand how the code works, you'll find it easier to modify and expand the examples to suit the specific goals of each lab throughout the semester.

3 Laboratory Introduction

This lab is divided into 3 parts, each covering key tasks you need to master as an engineer to understand and communicate about data.

- In Part-1, you'll learn how to load experimental data from a past experiment and use it to create three types of plots: **lines**, **scatters**, and **histograms**. These are basic but powerful tools that engineers and scientists use to understand their data better and spot useful patterns.
- In Part-2 you'll import tabular data, and use some of MATLAB's curve fitting tools to find a trend between two related variables. You'll also apply statistical formulas (covered in class) to estimate **uncertainty** or **confidence intervals** for the fit and the slope. Learning how to fit curves and evaluate confidence helps engineers build models that can make predictions and understand how **reliable** those predictions are.
- For Part-3, you'll work with an example set of data that follows a normal (bell curve) distribution. You'll start by generating artificial data and creating histograms to see how it's spread out. Then, you'll calculate a few key statistics: the mean, median, and standard deviation. After that, you'll use the equation for the normal distribution to recreate a smooth curve that represents your data. Engineers often use this kind of analysis to turn a set of measured values into a continuous model that helps with design decisions. For example, it can help answer questions like: What percentage of parts will fit together based on the variation of a key dimension?

4 Part-1: Scatterplot and Histogram from Voltage data

In this part of the lab, you will read real voltage data from a file, visualize it using a scatter plot and histogram, and calculate basic statistics (mean and standard deviation). These are fundamental skills for working with real-world experimental data.

During most experiments, we read measurements and record them. You can do this with pen and paper, but using a programming tool such as MATLAB and a DAQ (data acquisition) system, we can automate this process to efficiently collect measurements and store them in a digital file.

Data can be stored in many different file formats (e.g. **.csv**, **.mat**, **.dat**, etc). Most of these store data as plain ASCII text, and each uses a **filename** and an **extension** to indicate its type for your computer (e.g. **my_data.csv**). Different formats offer various advantages depending on application, and you'll likely encounter many of them throughout your engineering career. In this course, we will use **.csv** (comma-separated values) format. CSV files organize data into columns and can include headers to clearly label variables. They are human-readable and can be opened in any plain text editor, such as Notepad++. They also work well with programs like Excel, MATLAB, and Python, making them an excellent format for both inspection and analysis.

A complete **example solution** and a well-commented **starter script** are available on Canvas. The starter script will help guide you through each step with hints on which MATLAB commands to use. You will be editing and running that script to complete this lab.

4.1 Detailed Instruction

1. Set Up Your Working Folder
   - An example blank file structure is also located on the **Lab-01 Canvas page**. You should use this folder setup for each lab.
2. Create a new MATLAB script file and save it in the code folder in "/ME3300/Lab_01/Code/" and name the script file with "FirstNameLastName_Lab_01.m". Always change the MATLAB script file names to something that allows you to understand the purpose of that script file. Use this file structure to organize all future experiments and data for this course.
3. Download the sample data file "time_voltage_data.csv." from the Lab_01 page on Canvas and save it to your local data folder.
4. Inspect the Data File using Notepad++
   * *I will use **Q** in these manuals to indicate questions you should answer in your lab notebook!* ***These will be checked.***

   Q How many columns of data do you see in the document?

   Q Does the data include headers? If so, do the values below the headers make sense (e.g., time counts up, and the scale is reasonable for the measurement units)?
5. Import the data into MATLAB using one of two methods.
   - Method 1: Use **readtable**
   - Method 2: Use **readmatrix**
   - Note in each method, you need to use a relative path to access data e.g.**readmatrix(".._volts_data_example.csv");**
   * Syntax for each method is included in the example code. *

   Q What is the difference between the readable and readmatrix methods?
6. Compute Statistics on the data
   - Compute the mean, median, standard deviation, and variance of the voltage. Record these values in your notebook.
   * You may use built-in MATLAB functions for these stats, or implement calculations yourself. *
7. Visualize the Data
   - Create two separate figures, one scatterplot and one histogram plot.
   - Your complete plots should look similar to figures 1 and 2.
   - See this page for more information on formatting markers and lines for your plots.
   - Include the measured sample mean and standard deviation on the scatter plot as annotations.
8. Save Your Plots
   - Use the "exportgraphics" command to save a 600 DPI **.pdf** render of your figure to file. Repeat to save a **.png** render as well. You will regularly save figures in both formats as **.pdf** files are better for submitting on Canvas, but **.png** retains transparency and is more appropriate for inserting in reports and presentations.

   Q What is the importance of figure DPI? When might you need to use a higher DPI? How about a lower DPI?

4.2 Plot Format Instructions

Your complete figures need to match the example formatting. Use the following list to make sure you properly format your figures. Note that in future labs, you will be expected to continue using similar settings (and more) to generate high-quality figures.

- Both Plots
  - Figure size: 6.5" wide x 3.5" tall
  - Figure background color: white
  - All fonts: "Times"
  - Min/Max font sizes: 10 pt / 12 pt
  - Turn on major and minor grids
  - Set the outer box off
  - Include appropriate axis labels (always include units when needed),
  - Add a short, accurate title, including your name and today's date.
- Scatter Plot
  - Match example horizontal line colors (red for standard deviation lines, black for the mean)
  - Horizontal linewidths: 1.5 pt
  - Scatter Marker Size: 25
  - Scatter Marker face and edge color: blue
  - Marker face alpha: 0.8
- Histogram Plot:
  - Choose an appropriate number of bins: between 10 and 40.
  - Set the face color to blue

4.3 Example Results

You will be submitting .pdf figure files for your post-lab. Your submitted files should look similar to Figures 1 and 2.

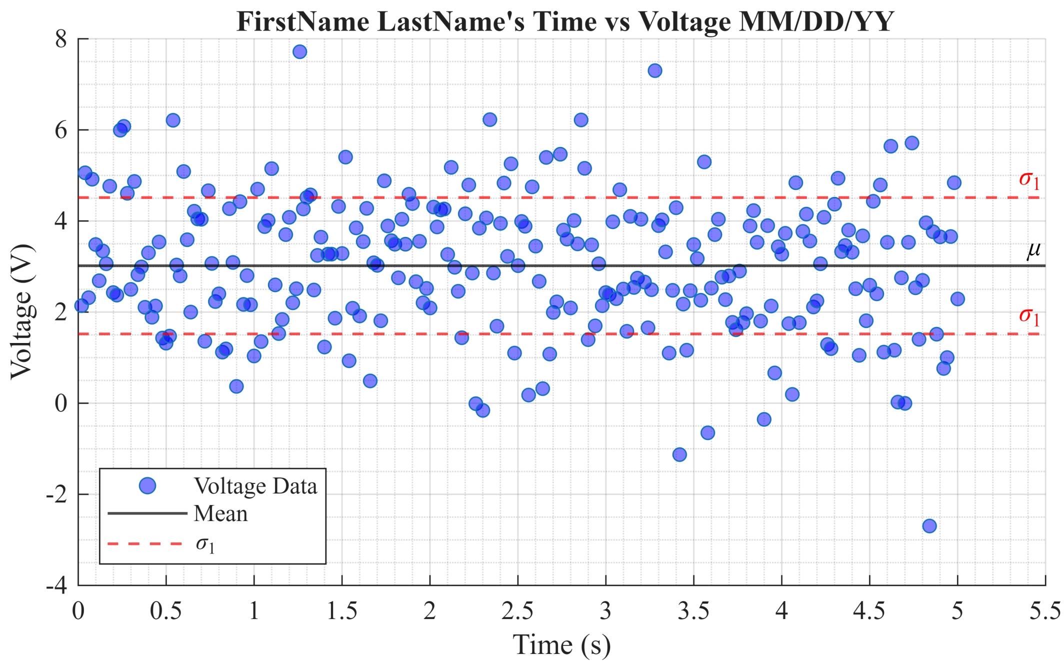

Figure 1: Sample figure showing scatter of the experimental data around the mean value.

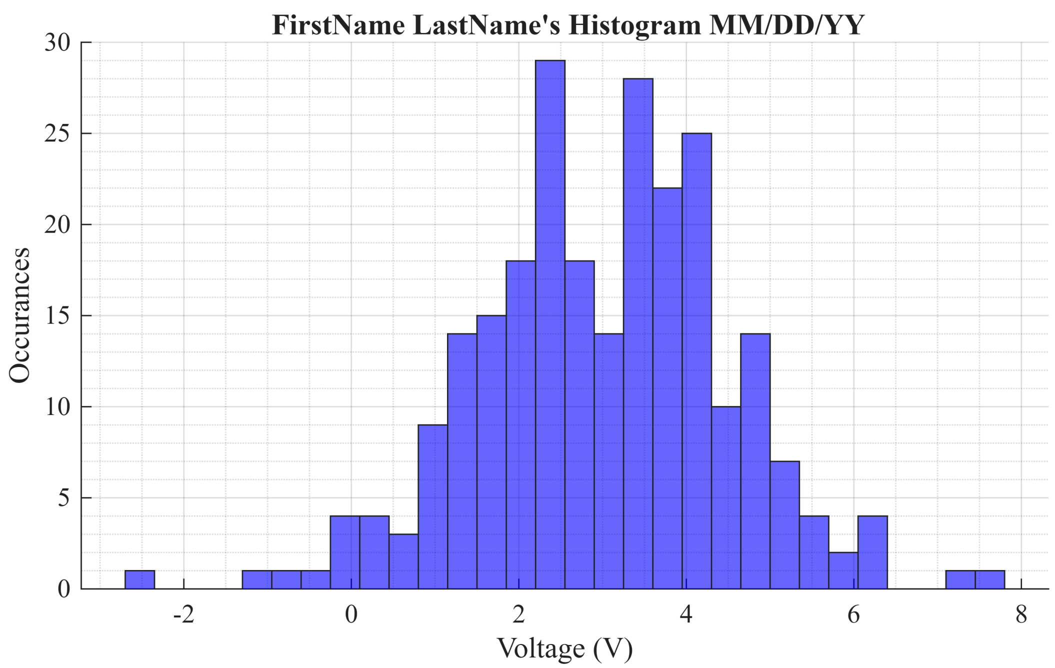

Figure 2: Sample figure showing histogram.

5 Part-2: Linear Curve Fit on Scatter Data

For part 2 of this lab, you will use tabular data to generate a scatter plot and then perform a curve fit on the data. You will compute and add confidence level lines to the plot. You will also learn how to annotate your plot by adding the linear fit equation and the standard deviation of the slope to the plot as text.

To calculate the confidence level in the curve fit, we will use the following equations

$$\begin{aligned} y_{cl} &= y_{fit} \pm t_{\nu,P} s_{yx} \quad (P\%) && (1) \\ s_{yx} &= \sqrt{\frac{\sum_{i=1}^{N}(y_i - y_{c_i})^2}{\nu}} && (2) \end{aligned}$$

where $t_{\nu,P}$ is the students t value, ν is degree of freedom ($\nu = N - 1$), P is confidence level (use 95%), s_{yx} is the standard error of fit, y is a vector of sample data, y_c is a vector of curve fit data, and N is the total number of data points.

To compute the standard deviation value of the slope, you will implement the following formula.

$$S_{a1} = s_{yx}\sqrt{\frac{1}{\sum_{i=1}^{N}(x_i - \overline{x})^2}}, \tag{3}$$

where s_{yx} is provided in equation 2 & $\overline{x}$ is the mean value of x data points. Please review the sample code to calculate these values. We will cover equations these parameters in further detail in class.

5.1 Detailed Instruction

Use the data provided in table 1 and equations 1 to 2 create a figure similar to Figure 3.

1. Transfer the data to MATLAB manually, or by using another program like Excel to make a **.csv** to import.
2. Plot the raw data on a figure using red filled circle markers with a marker size of 75 using scatter.
3. Curve fit the data using the MATLAB curve fitting functions **polyfit** and **polyval**.
4. Plot the resulting linear curve fit line as a solid blue line with **lineweight** = 2.
5. Compute the confidence interval using equations 1 and 2. Estimate the t statistic using **tinv(0.95,ν)** Plot CI lines on the plot using black dashed lines with line widths of 1.
6. Compute the slope fit error using equation 3.
7. Annotate the plot with the fit line equation and slope error using the text command and **sprintf**.
8. Include a legend, title, labels, and update formatting to closely match example figure 3.
9. Save a **.pdf** of the figure again for your post-lab submission.

Table 1: Experimental Data.

Velocity (m/s)	Volts (V)
0.00	0.7036
1.00	1.0096
2.00	1.3907
3.00	1.8867
4.00	2.2557
5.00	2.6313
6.00	2.9504

5.2 Plot Format Instructions

Use the following list to make sure you properly format this figure.

- Figure size: 6.5" wide x 3.5" tall
- Figure background color: white
- Text: font "times", Min/Max font sizes: 10 pt / 12 pt
- Turn on major and minor grids
- Set the outer box off
- Include appropriate axis labels (always include units when needed),
- Add a short, accurate title, including your name and today's date.
- fit line: solid blue with 2 pt width
- CI Line: dashed black with 1 pt width
- Scatter Markers: Red Face and edge, size 75

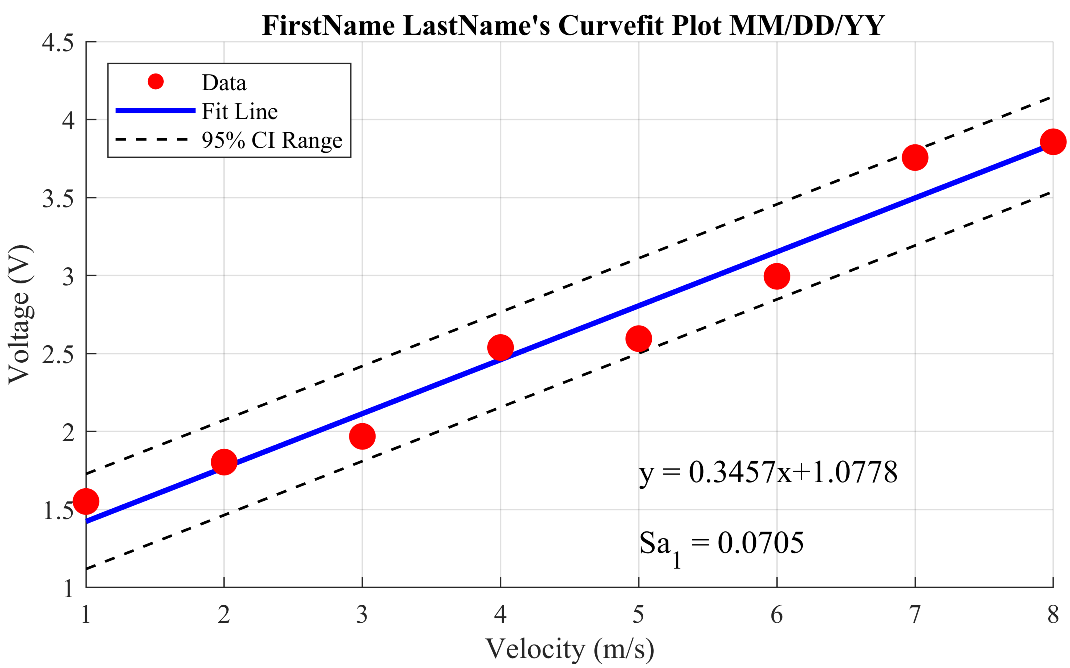

Figure 3: Sample figure showing experimental data and curve fit.

6 Part-3: Generated Histogram and PDF Lines

For the final part of this lab, you are going to generate your own synthetic data, then use it to create visualizations of probability distribution functions (PDF) using both histograms with normalization and the standard normal PDF equation.

The probability density function for normal distributions is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-0.5 \times \left(\frac{x - x'}{\sigma}\right)^2\right], \tag{4}$$

where σ is the standard deviation of the population and x' is the mean value of the population.

For your plot, create data with the following true mean and standard deviation values

1. $x' = -2$ and $\sigma = 0.80$
2. $x' = 0$ and $\sigma = 0.50$
3. $x' = 0$ and $\sigma = 2.00$
4. $x' = 3$ and $\sigma = 2.00$

6.1 Detailed Instruction

1. Generate synthetic datasets. Use N = 1000 points for each set. You can use MATLAB's **randn** function to help with this step.
   - hint: data = randn(N,1)*$\sigma + \mu$.
2. Implement your PDF function using equation 4 to generate a curve of samples for each dataset.
   - Use range -10 to 10 for the input x. One way to generate a vector of sample points for x is to use **linspace**
   - You can create a separate function **.m** file or use the @ symbol to make a local function handle in your main script.
3. Setup a 2 row, 1 column figure using **tiledlayout** and **nexttile** commands, or **subplot** commands.
4. Use **histogram** and set **normalization** to "pdf" to generate PDF histograms on the top chart.
   - Set the histogram colors to black, red, green, and blue in order.
   - Set the number of bins, **nbins** = 50 for each histogram.
5. Use your PDF function to generate continuous lines and plot them on your bottom chart.
   - Set the line colors to black, red, green, and blue in order.
6. Include a legend, title, labels, and update formatting to closely match example figure 4.
7. Save a **.pdf** of the figure again for your post-lab submission.

6.2 Plot Format Instructions

Use the following list to make sure you properly format your figure.

- Figure size: 6.5" wide x 7" tall
- Figure background color: white
- Text: font "times", Min/Max font sizes: 10 pt / 12 pt
- Turn on major grids
- Set the outer box off
- Include appropriate axis labels (always include units when needed),
- Add a short, accurate title, including your name and today's date.
- linewidths: 2 pt
- Match x and y axis limits so that plots are comparable and called identically
  - This is very important for understanding when making multi-axis figures!

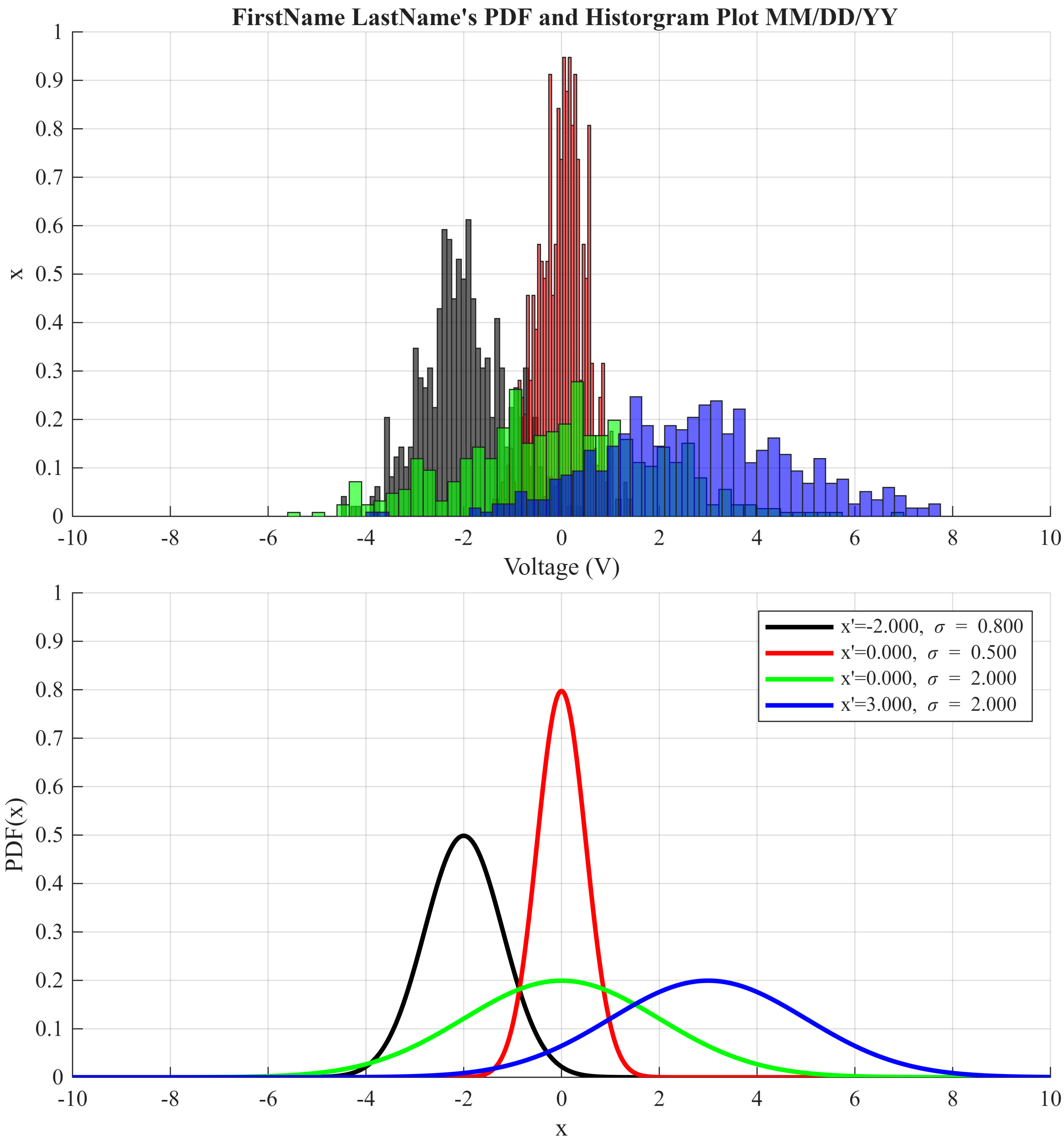

Figure 4: Sample figure showing PDF for different true mean and standard deviations.

7 Post-Lab Assignment

Following each lab, you will upload submissions to Canvas.

7.1 General Instruction

- Make sure to name the files your FirstName_ LastName_ ExptNoXX _ PartXX. You will follow this file-naming format for naming figures in all future experiments.
- Make sure your submitted figures have updated titles with your name and the date. Include the apostrophe!
- As you can see, MATLAB provides a great deal of power for formatting high-quality scientific figures. Here are a few best practices to remember throughout the semester:
  - Sampled data points should be plotted using markers without lines connecting the points.
  - Continuous lines/curves from theory (or curve fit lines) should be plotted using solid colored lines.
  - Remember to set the x and y-axis limits. When preparing multiple-axis plots, limits should usually match to make data easy to compare.
  - Make sure the major grid lines are visible. When needed, set axis ticks to be dense enough to make it easy to read the figure.
  - In this class, all plot text should be in Times font. You will need to specify this for the title, labels, and legends. The axis labels and numbers should typically be in 10-point font.
  - Titles are not always needed if labels are clear enough, but should be set at a 10-point font.
  - Make sure to add a legend when there are multiple elements on the plot. Keep in mind that the legend should not hide the plotted data.
  - You should always refer to example figures in the lab manual to make sure your figures match expected formatting.
  - If in doubt, make sure to request help from the lab TAs. They can help confirm if your plot matches expectations.

7.2 Submission Items

- Your final **.m** script file (named **Lab_01.m** or similar).
- Part-1
  - Your Time vs. Voltage Scatter plot in **.pdf** file format.
  - Your Voltage Histogram in **.pdf** file format.
- Part-2
  - Your Voltage vs. Velocity scatter and curve-fit plot in **.pdf** file format.
- Part-3
  - Your multi-axis PDF plot in **.pdf** format.

7.3 Questions

Be prepared to answer the following post-lab questions (make sure you can answer these questions before leaving the lab!) **Post-labs are due Fridays at 10:00 pm.**

1. What happens to the shape of a normal distribution (PDF) when the standard deviation is decreased?
2. What happens to a normal distribution (PDF) when the mean value is increased?
3. What is the observed standard deviation of the slope in part-2?