Homework 01: Data Analysis

Katha Haldar

Department of Physics, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202

1 Introduction

Gnuplot is a versatile and potent tool that we have used extensively for data visualization and comprehensive data analysis. In this comprehensive guide, we will explore the intricacies of plotting data using Gnuplot and delve into the multifaceted aspects of data analysis, including fitting data, regression analysis, statistical analysis, standard deviation calculation, and error analysis.

2 Plotting Data using Gnuplot

Gnuplot is renowned for its adaptability in handling various data types and producing high-quality visualizations. When we plot data using Gnuplot, we consider the following aspects:

- **Data Import:** Gnuplot can import data from various sources, including text files, spreadsheets, and even live data streams.
- **Plot Customization:** We customize our plots by specifying titles, labels, axes, colors, and line styles. Gnuplot's flexible syntax allows for detailed control over plot aesthetics.
- **Plot Types:** We create a wide range of plot types, such as line plots, scatter plots, bar charts, histograms, and 3D plots, to effectively represent our data.

3 Fitting Data

Fitting data is invaluable for modeling experimental results with mathematical functions. Gnuplot offers several fitting options:

- **Basic Fit:** We start by fitting our data to a pre-defined function or equation, adjusting parameters to minimize residuals.
- **Determination of Function:** Discovering the most suitable mathematical model for our data can be challenging. Gnuplot provides tools to help us identify the best-fitting functions by visualizing fits with various equations.
- **Best Fit:** Once we've determined the function, Gnuplot employs optimization algorithms to find the best-fitting parameters, minimizing the sum of squared residuals.

4 Regression Analysis

Regression analysis allows us to quantify the relationships between variables. Within Gnuplot, we can perform:

- **Quadratic Regression:** This is appropriate when data exhibits a curved trend. Gnuplot helps us fit quadratic functions to data points, providing coefficients and goodness-of-fit metrics.
- **Linear Regression:** This is ideal for modeling data with a straight-line equation ($y = mx + b$). Gnuplot calculates regression coefficients, residuals, and correlation coefficients.

5 Statistical Analysis

Gnuplot isn't limited to just graphing; it can also assist in statistical analysis:

- **Descriptive Statistics:** We can easily compute and visualize statistical measures like mean, median, variance, and skewness to gain insights into data distributions.
- **Box-and-Whisker Plots:** We use Gnuplot to create box plots that reveal data dispersion and outliers.

6 Standard Deviation & Physical Interpretation

Standard deviation, a crucial statistic, quantifies the spread of data points around the mean. Gnuplot aids in:

- **Standard Deviation Calculation:** Gnuplot computes standard deviations, helping us understand the variability of our data.
- **Error Bars:** We include error bars in our plots to visualize the uncertainty associated with data points, making it easier to assess data reliability.

7 Error Analysis

Error analysis is pivotal in scientific research and engineering. Gnuplot assists in:

- **Error Propagation:** We compute propagated errors when combining multiple measurements with uncertainties.
- **Visualizing Errors:** We represent errors in plots using error bars, helping us make informed decisions regarding data quality.

8 Case Study

Table 1: Our dataset. We choose a fitting function $f(x) = (x - 0.5)^2 + 1$ for our data, against which the error is calculated.

x	y	$\Delta y = \mid f(x) - y \mid$
-3.022460938	19.6511187	6.24338764
-1.251831055	5.972167361	1.903255316
2.623291016	23.97354176	18.46517702
3.712768555	45.31074309	33.9888613
2.349243164	18.96341096	14.54371068
5.571289063	68.60618677	41.88821401

x	y	$\Delta y = \mid f(x) - y \mid$
-1.869506836	7.383182036	0.76861939
0.471191406	4.392371255	3.39154132
0.697021484	6.715382153	5.676564688
-6.327514648	75.13253214	27.51757587
-4.390258789	37.61193692	12.6973059
-0.762329102	2.971377592	0.37790283
5.474243164	59.42879433	33.68569928
4.918823242	53.86350985	33.33751101
-1.098022461	4.661798613	1.108122827
-2.138671875	13.06635378	5.103764516
6.865844727	143.6846349	102.1606558
-3.243408203	20.74656512	5.733460146
-2.944946289	22.72672544	9.859070506
6.060791016	108.9985779	77.07618118
6.866455078	136.4553321	94.92358184
-2.085571289	8.553788797	0.868609906
5.653686523	80.59419285	53.03370807
-2.624511719	17.1857301	6.423156618
6.857299805	116.7170289	75.30176809
-3.395385742	27.12694667	10.95291659
0.645751953	5.410118703	4.388875071
-6.383056641	67.82400362	19.4475349
7.684936523	144.0262053	91.40289246
1.228637695	9.216766568	7.685853677
4.670410156	54.00218479	35.60986392
1.826171875	9.99271433	7.233982488
2.238769531	14.55620856	10.53288908
-6.421508789	80.92304465	32.01576073
5.888671875	95.75933598	65.7215514
3.72253418	32.01828569	20.63355915
0.531616211	5.729715182	4.728715597
-5.459594727	72.56585615	36.04908684
9.307250977	206.5814274	128.0137576

x	y	$\Delta y = \mid f(x) - y \mid$
-2.340087891	10.69866589	1.632566661
7.274169922	136.235914	89.34653587
2.297973633	15.06007755	10.82736837
0.611572266	4.549782641	3.53733427
-6.92199707	125.2685676	69.18252709
-8.044433594	163.8085778	89.80123236
3.035888672	32.82373315	25.39300179
2.490234375	22.43314123	17.47210836
-2.544555664	11.27462746	1.005308269
-8.578491211	126.851984	43.43298133
-5.858764648	67.29699097	25.86310312
3.463745117	42.70444936	32.92066424
-4.235229492	29.86170506	6.439306718
4.951171875	73.90422439	53.09129333
-0.091552734	3.769785725	2.419851088
-0.785522461	4.628015698	1.9754477
2.424316406	23.62205799	18.91906436
1.835327148	13.46727385	10.68417526
-0.257568359	3.177298187	1.603388368
6.412353516	84.56864694	48.61272284
5.749511719	99.92098112	71.36360783
1.456298828	7.875843756	5.961336308
8.860473633	154.5560294	83.65851003
9.045410156	173.5181721	99.49413737
4.91027832	59.57886348	39.12830862
5.864257813	83.69814908	53.9228872
-8.50402832	163.9802578	81.90773181
-2.753295898	20.15330904	8.56937484
-0.619506836	3.223096689	0.969801133
4.855957031	77.7353083	57.76094664
6.600952148	103.4784587	65.25684159
-1.37878418	6.517761986	1.987931991
-9.108276367	144.1453005	50.82632576

x	y	$\Delta y = \mid f(x) - y \mid$
-6.446533203	69.03789156	19.78356802
6.051635742	88.49802641	56.677367
2.075195313	17.64087398	14.15963371
-2.111816406	7.88880294	0.067218001
2.479858398	18.0403134	13.12047412
5.347900391	84.89219742	60.39005922
-5.853271484	84.36580045	43.0017419
2.477416992	19.55655483	14.64637687
0.142211914	2.813068853	1.685056539
6.819458008	102.5455117	61.60996219
7.828369141	170.3306906	115.6256963
9.886474609	288.7411145	199.6352089
8.103027344	167.9783032	109.1722784
-4.673461914	52.55953188	24.7948237
3.899536133	44.17353339	31.61668747
0.410766602	4.590633533	3.582670934
8.565673828	189.2091368	123.1540425
6.465454102	98.41898688	61.83234424
4.754638672	68.46379088	49.36184065
-6.127929688	59.90595347	14.97650152
4.446411133	49.50385192	32.92969109
7.269287109	154.3738264	107.5505784
0.727539063	4.977326461	3.925552436
6.567382813	134.8416388	97.0285046
-8.045043945	145.8954191	71.87764308
9.413452148	169.2085676	88.75893841
8.923950195	205.9429852	133.9800483
0.725097656	4.371552221	3.320883266

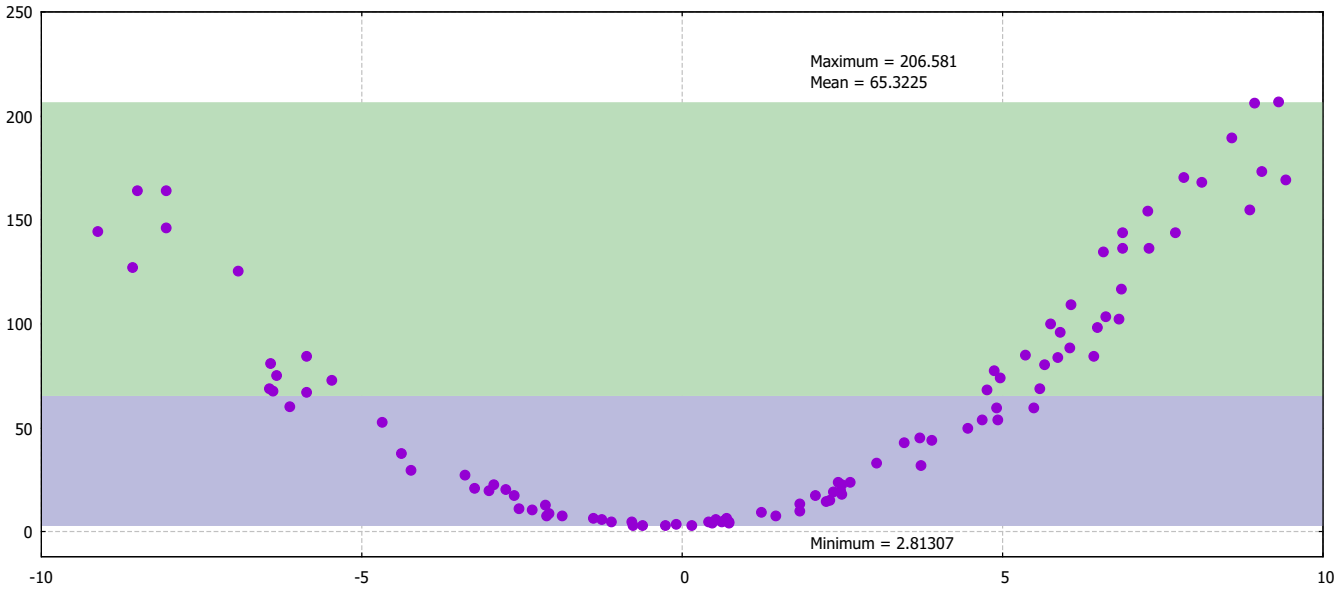

Figure 1: Determining the minimum, maximum, and the mean

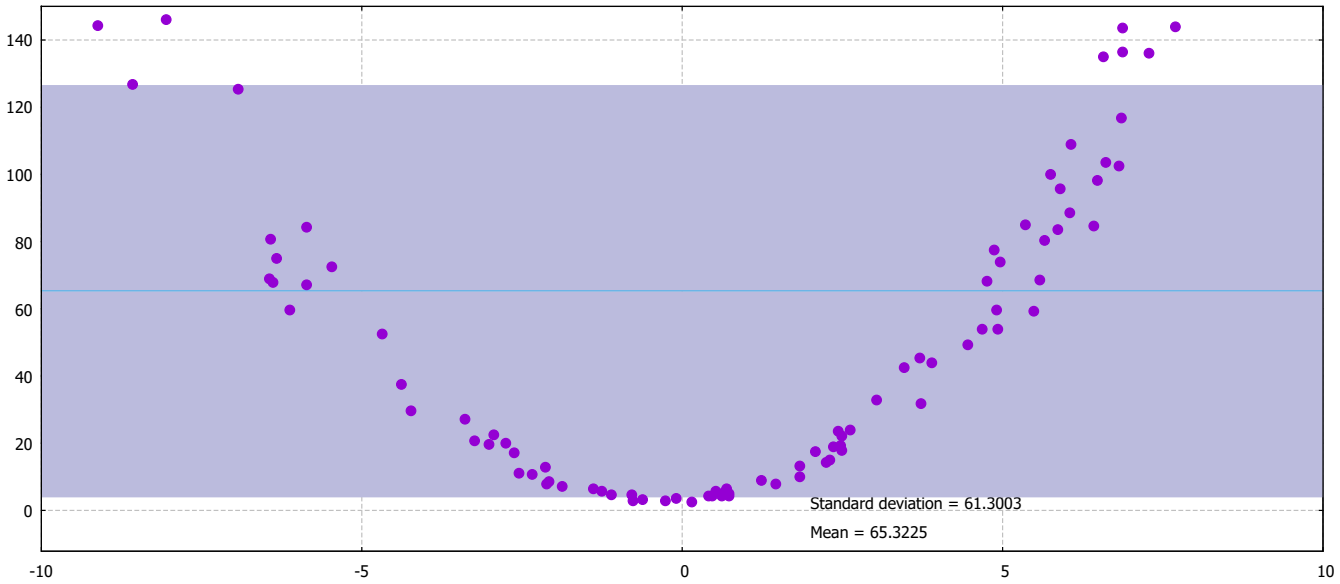

Figure 2: Calculating the standard deviation

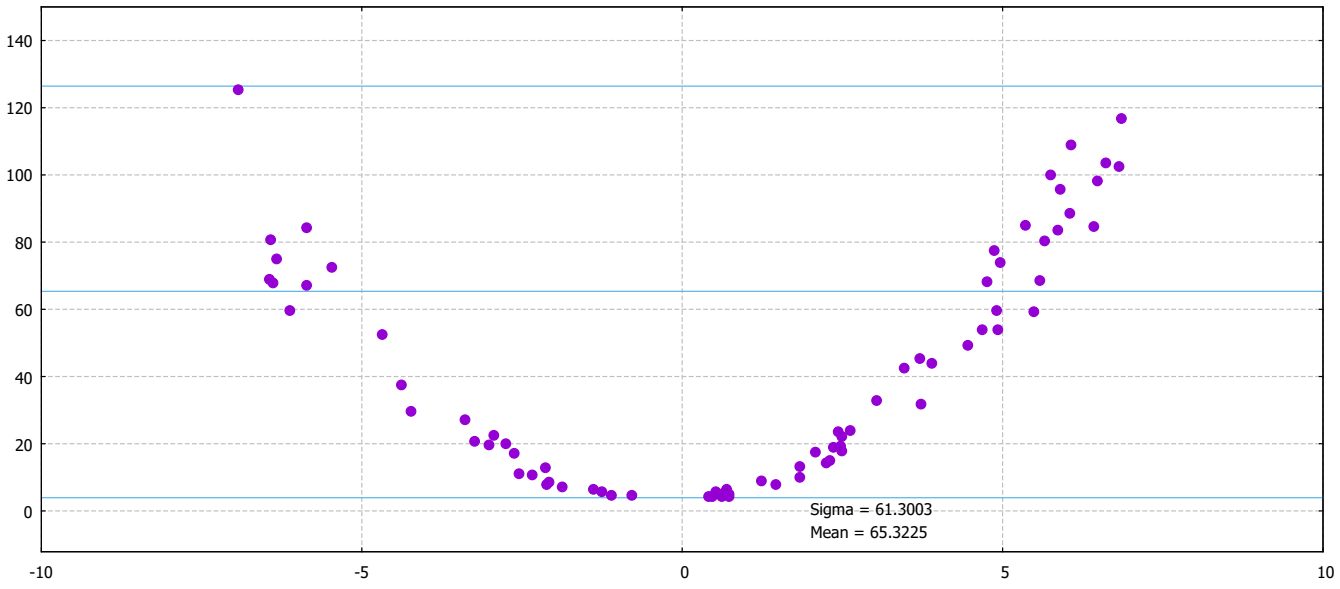

Figure 3: Discarding the Points w.r.t. Standard Deviation

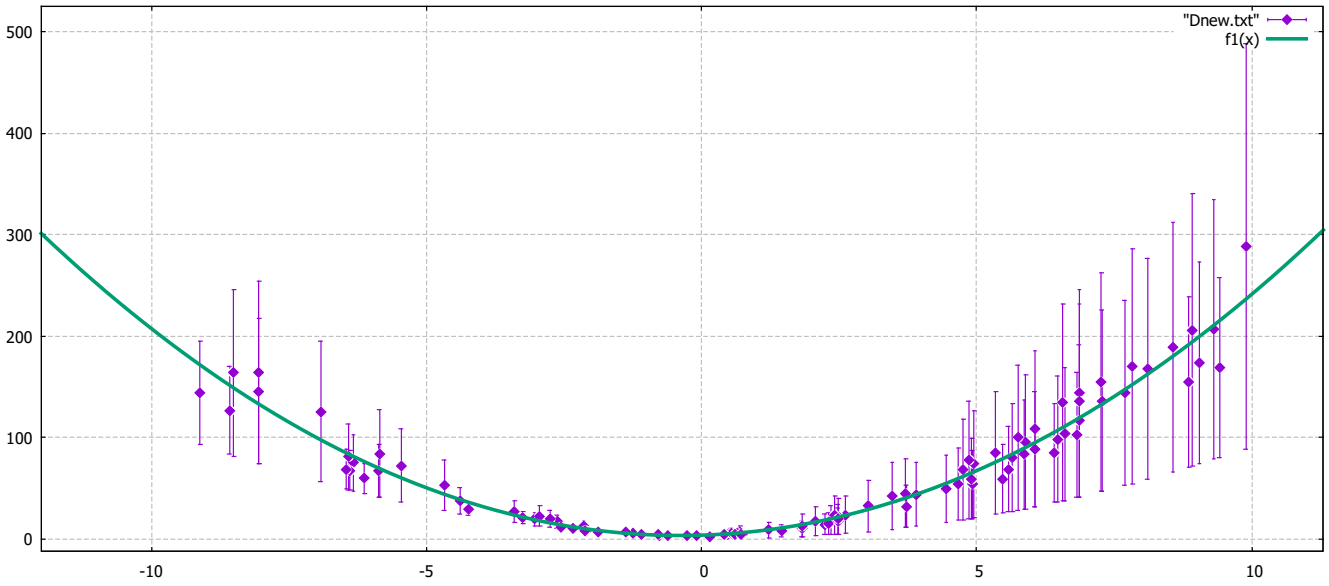

Figure 4: Quadratic Fit & Analysis of Data

9 Conclusion

Gnuplot emerges as an indispensable tool for plotting data and conducting in-depth data analysis. By leveraging its multifaceted features encompassing data fitting, re-

gression analysis, statistical examination, standard deviation computation, and error analysis, we can gain profound insights into the intricacies of our datasets. As a scientist, engineer, data analyst, or researcher, Gnuplot empowers to present, scrutinize, and interpret data effectively, contributing to informed decision-making and a deeper understanding of our subject matter. Gnuplot's adaptability and versatility make it an essential asset for researchers and professionals across various fields.

References