Experimental Physics Laboratory 2: Calculating the Value of Water Density using Metal Rod and Water Container

Osvaldo L. Santos-Pereira * †[1,2]

[1] *Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil*

February 17, 2025

Abstract

This article presents a detailed analysis of an undergraduate physics laboratory experiment designed to determine the density of water using fundamental measurement techniques and data analysis methods. The experimental setup consists of a precision scale, a graduated container filled with water, and a suspended metal rod held by a crank, allowing for controlled displacement measurements. The primary objective of this experiment is to reinforce essential concepts in experimental physics, particularly in deriving physical models that correlate measurable quantities, performing precise measurements, and analyzing data using regression techniques via ordinary least square methods for fitting data into linear models.

A common challenge students face is formulating a theoretical model that links directly observable variables, such as mass and volume, to the target physical quantity, in this case, water density. Furthermore, many students struggle with linearization techniques, which are crucial for applying linear regression to experimental data via ordinary least squares methods. The ability to extract meaningful physical parameters from the slope and intercept of a fitted linear model is often poorly understood despite its fundamental importance in experimental physics. Also, the difficult-to-understand simple physical phenomena can alter the results of the experiments, such as friction forces due to the contact of the metal rod with the surface of the container. Or even the main differences between selecting which model to use, a $M \times V$ or a $V \times M$ model, since some of the students fail to grasp the propagation of errors due to the mathematical relation they are using.

By analyzing the collected data and fitting a least-squares regression line to the mass-volume relationship, students can determine the density of water as the slope of the best-fit equation. This study discusses students' common misconceptions and difficulties throughout this process, providing insights into pedagogical strategies that can enhance their understanding of experimental physics and data analysis. The results emphasize the importance of integrating theoretical modeling, systematic measurement techniques, and statistical data analysis to improve students' ability to interpret and extract meaningful physical quantities from experimental observations.

This article aims to provide students with a theoretical and computational aid to dive deep into the physical interpretations of this experiment. Python codes are provided to fit and graph the linear model using an experimental table.

*olsp@if.ufrj.br

†ORCID: 0000-0003-2231-517X

Introduction

Hydrostatics is a branch of fluid mechanics that studies the equilibrium of fluids at rest and the forces exerted by or upon them. The fundamental principle governing hydrostatics is Pascal's law, which states that a change in pressure applied to an enclosed incompressible fluid is transmitted undiminished throughout the fluid [4, 6, 5]. Another crucial concept is Archimedes' principle, which states that a body submerged in a fluid experiences an upward buoyant force equal to the weight of the displaced fluid [7, 4, 5]. These physical principles are widely applied in engineering, geophysics, and biological systems, forming the theoretical foundation for determining fluid densities experimentally.

In this experiment, an undergraduate-level physics setup is used to determine the density of water through buoyancy measurements. The setup consists of a submerged cylindrical object connected to a spring system, allowing precise volume displacement control. By analyzing the equilibrium conditions before and after submersion, the density of water can be inferred using force balance equations. The experiment demonstrates the practical application of hydrostatic principles and provides students with hands-on experience in fluid mechanics experimentation in the laboratory.

This work is organized as follows: Section one discusses and explains the experimental setup, presenting the key physical variables. The second section summarizes the main pitfalls and caveats of the experimental procedure faced by the students. The third section presents the theoretical framework and the derivation of the main equations used to model the physical phenomena analyzed in this experiment. The fourth section presents the results and data analysis from the experiment and how the student should investigate the physical phenomenon in this experiment. The fifth section presents a pedagogical discussion on the difficulties faced by the students during the experimentation and the writing of the report. Lastly, the conclusion of this work is presented. The appendix presents the derivation for the slope and the intercept for the Ordinary Least Squares, as well as the errors for each of those estimators.

1 Experimental setup

Hydrostatics studies fluids at rest and the forces acting on them. In this experiment, we analyze the hydrostatic forces exerted on a submerged object to determine the density of water using the principles of buoyancy. The setup consists of a graduated cylinder filled with water, a digital scale, and a metal bar suspended by an adjustable support. By recording variations in mass and volume as the bar is gradually submerged, we can quantify the buoyant force exerted by the liquid.

The experimental apparatus, depicted in Fig. 1, consists of two distinct stages. In the first stage, the metal bar is outside the liquid, held by a support that ensures it does not interact with the fluid. The tension in the support balances the bar's weight, keeping it in equilibrium. In this state, the scale measures the combined mass of the graduated cylinder and the liquid, denoted by M_0. The initial liquid volume is V_0, providing a reference measurement.

In the second stage, the bar is partially submerged in the liquid. As the bar is lowered using the adjustable support, it displaces a volume of fluid, now represented as V_d. According to Archimedes' principle, the fluid exerts an upward buoyant force E on the submerged portion of the bar. Due to Newton's third law, action and reaction, the liquid also experiences an equal and opposite force, which alters the scale reading. Consequently, the new mass reading on the scale is $M > M_0$ due to the reaction force acting on the liquid. This setup allows us to quantify the buoyant force by analyzing the variations in mass and volume readings as the bar is submerged.

The materials used in this experiment include a graduated cylinder to measure liquid displacement, a scale to record mass variations, metal bars of different materials and cross-sections, water as the working fluid, and support with a crank for controlled vertical movement of the metal bar.

The following steps are followed: First, the mass of the empty container is measured and denoted as M_R. The scale's precision is checked, and the most minor measurable division is noted. Ensure the support and scale are leveled for accurate readings. The liquid's initial volume V_0 in the graduated cylinder is recorded. The liquid level is adjusted to ensure that the bar can be fully submerged without overflowing.

For data collection, the values of M_0 and V_0 are measured with the metal bar completely outside the liquid. The bar is lowered incrementally into the liquid using the crank, displacing a volume V_d each time the experiment is executed. The new mass M_1 and volume V_1 are recorded. This process is repeated for additional measurements $(M_2, M_3, \dots)$ and $(V_2, V3, \cdots)$ while ensuring that the bar remains suspended and does not touch the graduated cylinder. Mass M and volume V measurements must be annotated in a proper table with their respective measurement errors, σM for the mass, and σ_V for the volume.

The experiment is conducted using two different metal bars. For the second bar, measurements are taken only for volumes equal to or greater than the final measured volume with the first bar, V_5. This setup enables a direct experimental verification of Archimedes' principle by relating mass variations to displaced liquid volume. The collected data will be analyzed to calculate water density and assess buoyant force's dependence on object properties.

2 Pedagogical Approach and Pitfalls

The group of students must follow the procedure.

- Using the data obtained for water, create a table containing the quantities M and V, along with their respective uncertainties.
- Initially, identify in the table which parameters are obtained directly and indirectly from the experiment.
- In the report, show how the results of the indirect measurements and their respective uncertainties were determined—for example, the error on the Ordinary Least Square parameters estimators.

Figure 1: Experimental setup for determining the density of water using hydrostatic principles. (left) Initial setup: A container filled with water is placed on a digital scale, measuring the total weight of the container and the liquid. right) Modified setup: A metal rod is suspended by an apparatus and partially submerged in the water. The system demonstrates the buoyant force exerted by the liquid on the rod, leading to changes in the scale's reading. By analyzing these variations, the density of the liquid can be experimentally determined using Archimedes' principle. ChatGPT-4o generated both images.

- Use the equations from the theoretical framework to determine the equation of the line $M = aV + b$ and then perform a linear fit using the experimental data to determine the values of the slope and intercept coefficients.
- With slope and intercept values or the linear model, indirectly determine the water density value and calculate the estimate's error.
- Anotate the values of $a \pm \delta a$ for the slope and $b \pm \delta b$ for the intercept in tables in the report.

To avoid pitfalls, the students must be attentive to some caveats in the experimentation procedure.

- The same student must perform the same procedure to reduce errors since each has a different sight, height, or manner of doing the measurements.
- The group of students must be organized and methodical to write down the data as soon as the measurement is performed.
- Watch out for the significant numbers of each measurement on the mass and the volume. In some experiments, the setup might be 'old school' on purpose. For example, using an old scale instead of a precision scale.
- Be careful with the crank when lowering down the metal road. If an angle is formed with the vertical, tangential forces may appear, and an experimental error may

affect the final calculated value for the water density.

- Do not let the metal rod touch the sides of the container for the same reason as the last item. Tangential forces may appear due to the contact of the rod with the recipient wall.
- Be aware of the dimensional analysis. The water density is 0.997 g/ ml at 25 degrees Celsius.

There are some caveats that the students must be attentive to regarding the physical interpretation of the experimental results.

- What model is the best choice to reduce errors if it is either $M \times V$ or $V \times M$? And why is that so?
- How to calculate the linear regression estimator errors that fit the data with the best line.
- What is the physical interpretation of $M_R = M_0 - \rho V_0$.
- Why is the slope calculation in a millimeter paper less accurate than ordinary least squares?
- How to properly propagate errors and how to estimate experimental mistakes.

3 Theoretical Framework

In the scenario where the metal bar is not immersed yet in the water, the scale only reads the reaction of the normal force on the container + liquid system

$$F_0 = M_0 g \tag{3.1}$$

where F_0 is the force acting on the balance, M_0 is the container's mass plus the liquid's mass, and g is the acceleration due to gravity. When a metallic bar is partially inserted into the liquid, forces begin to act on the liquid and the bar. In the static situation, only pressure forces contribute to the resultant force since the force due to the viscosity of the liquid depends on a relative velocity between the bar and the fluid. The sum of the pressure forces that a liquid exerts on a solid is called buoyant force, and it is given by the **Archimede's Principle** [4, 5]:

$$E = \rho V_d g \tag{3.2}$$

where ρ is the density of the liquid, and V_d is the volume of liquid displaced by the solid. In this situation, the buoyant force acts upwards, counteracting to push the metal bar out of the liquid. The reading on the scale is now $M > M_0$ since a Buoyant force is acting on the system. The resultant force is now $F = Mg$. Newton's second law applied to the liquid + container system results in the following expression

$$Mg = E + M_0 g \tag{3.3}$$

which can be read as the following expression

$$E = (M - M_0)g \tag{3.4}$$

Inserting Eq. (3.2) in Eq. (3.4), and noticing that the dislocated volume on the container is given by $V_d = V - V_0$, where V is the metal bar volume immersed on the liquid, results

$$E = (V - V_0)\rho g \tag{3.5}$$

Eq. (3.4) and Eq. (3.5) both represent the buoyancy force, so they must be equal

$$(V - V_0)\rho g = (M - M_0)g \tag{3.6}$$

And we can put Eq. (3.6) in the following manner

$$M = \rho V + (M_0 - \rho V_0) \tag{3.7}$$

or also in the following manner

$$V = \frac{M}{\rho} + \left(V_0 - \frac{M_0}{\rho}\right) \tag{3.8}$$

The theoretical model predicts that the buoyant force does not depend on any property of the solid, only on the volume of the object immersed in the liquid, see Eq. (3.2). Eq. (3.7) and Eq. (3.8) both represent the same model and show a linear relationship between M and V, of the form $y = ax + b$, where a is the slope and b is the linear coefficient. However, one must note that statistically, there are differences between those two models.

4 Data Analysis and Discussion

Consider an experimental setup consisting of a graduated container partially filled with a liquid of density ρ_{liquid} and a metallic bar that can be gradually immersed in the liquid. We define the key variables as follows:

- M_0: Mass reading on the balance when the bar is outside the liquid. It is the mass of the system liquid + container. It can be directly measured using the scale.
- M: Mass reading when the bar is partially immersed. It can be measured experimentally. It is formed by the system liquid + container + partially immersed metal rod.
- V_0: Initial volume of the liquid before submersion of the metal rod. It can be directly measured.
- V: Volume of the liquid after submersion. It is the volume V_0 added by the dislocated volume V_d. It can be directly measured experimentally.
- V_d: Volume of the liquid displaced by the submerged part of the bar. It can only be calculated with the formula $V_d = V - V_0$.

- g: Acceleration due to gravity. It cannot be directly measured but does not play a key role in this experiment; it only appears due to Newton's laws.
- E: Buoyant force acting on the submerged portion of the bar.
- M_R: It is the intercept of the model $M = aV + M_R$, defined by $M_R = M_0 - \rho V_0$. It cannot be measured directly; it is only calculated.

The table below presents a submerged object's measured mass and volume values to study buoyancy and fluid properties in an experimental setup. Each measurement includes its associated uncertainty, denoted as δM_i for mass and δV_i for volume, which accounts for instrumental precision and experimental variations. Additionally, the relative uncertainties, $\frac{\delta M_i}{M_i}$ and $\frac{\delta V_i}{V_i}$, are provided to quantify the accuracy of the measurements.

i	$(M_i \pm \delta M_i)\ g$	$\frac{\delta M_i}{M_i}$	$(V_i \pm \delta V_i)\ ml$	$\frac{\delta V_i}{V_i}$
1	208.12 ± 5.20	0.025	110.00 ± 5.50	0.050
2	217.37 ± 5.45	0.025	120.00 ± 6.00	0.050
3	228.34 ± 5.71	0.025	130.00 ± 6.50	0.050
4	241.61 ± 6.05	0.025	140.00 ± 7.00	0.050
5	251.05 ± 6.28	0.025	150.00 ± 7.50	0.050
6	262.01 ± 6.55	0.025	160.00 ± 8.00	0.050
7	272.14 ± 6.80	0.025	170.00 ± 8.50	0.050
8	278.10 ± 6.95	0.025	180.00 ± 9.00	0.050
9	290.44 ± 7.26	0.025	190.00 ± 9.50	0.050
10	297.54 ± 7.44	0.025	200.00 ± 10.00	0.050

Table 1: Experimental Data Table

The graph in Fig. 2 illustrates a given substance's relationship between mass and volume based on the experimental data summarized in Table 1. The data points, represented by black markers with error bars, correspond to the measured values of mass and volume, including their respective uncertainties. The linear regression analysis performed on these data points yields the best-fit equation:

$$y = (0.9894 \pm 0.0148)x + (101.30 \pm 2.33) \tag{4.1}$$

where the slope and intercept are provided with their associated uncertainties. The slope of the fitted line, approximately 0.9894, suggests a nearly proportional relationship between mass and volume, indicative of a constant density characteristic of the material being analyzed. The uncertainty band, represented by the shaded blue region, accounts for the confidence interval of the regression model, providing a visual representation of the possible deviation from the best-fit line.

From Table 1, it is evident that the relative uncertainties in both mass and volume measurements remain consistent, with $\frac{\delta M_i}{M_i} = 0.025$ and $\frac{\delta V_i}{V_i} = 0.050$ across all data points. This consistency ensures that the error propagation in the regression analysis is well-controlled. The regression computation incorporated the experimental uncertainties to provide a more robust estimation of the parameters.

The strong linearity observed in Figure 2 suggests that the underlying physical relationship adheres closely to the expected behavior of a homogeneous material obeying the principle of mass conservation. The minimal deviation of the experimental data from the regression line further confirms the measurements' reliability and the linear model's suitability for describing the system.

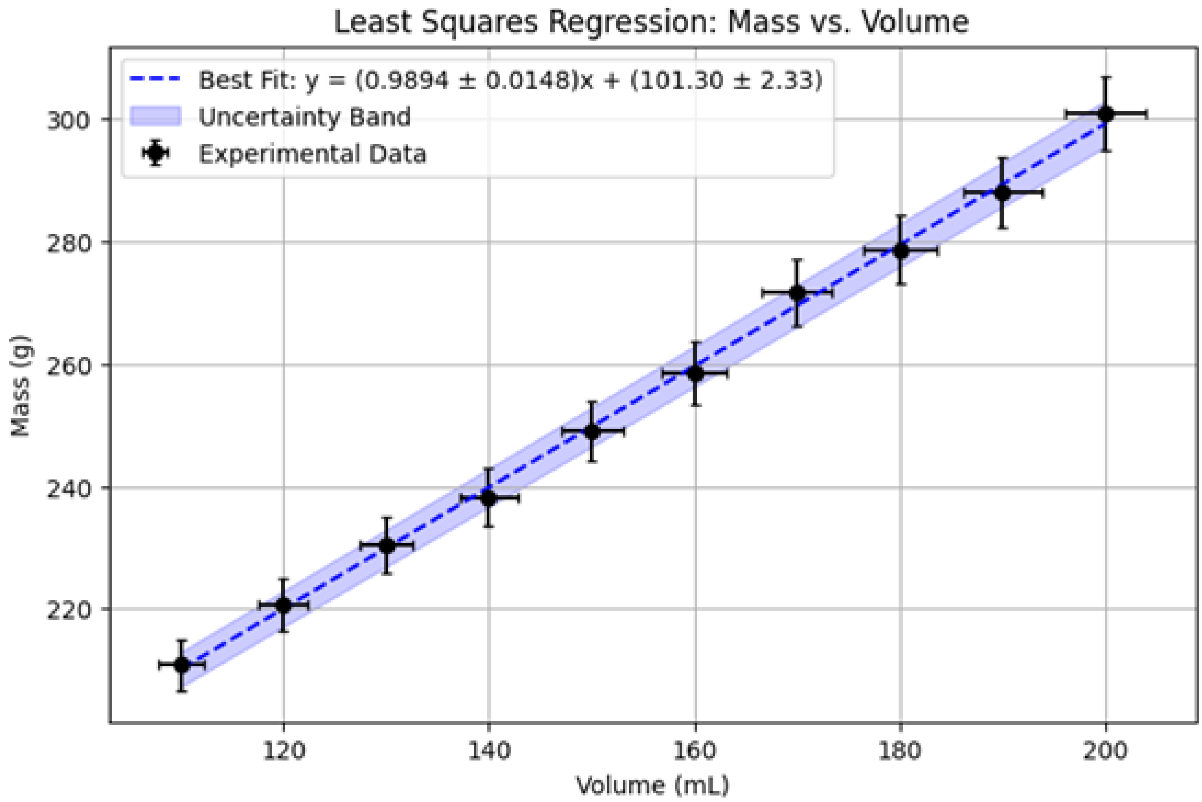

Figure 2: Least Squares Regression: Mass vs. Volume. The plot shows experimental data (black markers with error bars) and a linear regression best- fit line (dashed blue). The equation of the best-fit line is given as $y = (0.9894 \pm 0.0148)x + (101.30 \pm 2.33)$, where the uncertainties in the slope and intercept are provided. The shaded blue region represents the uncertainty band around the regression line, indicating the confidence interval.

To estimate the density of water from the experimental data, students can determine the slope of the mass-volume relationship using a simple graphical method on millimeter paper. A straight-line approximation can be drawn through the data points by plotting the mass M against the volume V. The slope of this line, which corresponds to the density, can be estimated using the fundamental definition from differential calculus:

$$a = \frac{\Delta M}{\Delta V} = \frac{M_2 - M_1}{V_2 - V_1} \tag{4.2}$$

where (V_1, M_1) and (V_2, M_2) are two points chosen from the experimental data. For instance, selecting the points $(V_1 = 110.0, M_1 = 210.81)$ and $(V_2 = 200.0, M_2 = 300.78)$

from the experimental table, we compute the slope as:

$$a = \frac{300.78 - 210.81}{200.0 - 110.0} = \frac{89.97}{90.0} = 0.9997\,\mathrm{g/mL}. \tag{4.3}$$

While this method estimates the density, it is susceptible to the specific points chosen. Ideally, the best-fit line for the data should be obtained through an Ordinary Least Squares regression, which minimizes the sum of squared residuals, given by:

$$e_i = M_i - \hat{a}V_i - \hat{b}, \tag{4.4}$$

Where $\hat{a}$ and $\hat{b}$ are the slope and intercept of the best-fit line, the best estimator parameters for the best line selected by the Ordinary Least Square method. This approach minimizes the overall error across all data points rather than relying on just two chosen points. In contrast, manually selecting points introduces significant variability, as minor fluctuations in measurement values can lead to disproportionately large errors in the estimated slope.

By employing Ordinary Least Square regression, students can more accurately determine the density of water while accounting for the inherent uncertainties in experimental data. Though useful for a rough approximation, the graphical method is prone to errors that statistical regression techniques can systematically reduce.

When selecting a model to determine the density of water, one must consider the mathematical implications of choosing either the $M \times V$ model, where mass is expressed as a function of volume, or the $V \times M$ model, where volume is described as a function of mass. The choice significantly affects the density estimation accuracy due to differences in how errors propagate. For the $M \times V$ model, the relationship is given by:

$$M = \rho V + (M_0 - \rho V_0), \tag{4.5}$$

where the slope of the regression line directly corresponds to the water density, ρ. The uncertainty in the estimated slope, σ_a, is given by:

$$\sigma_a = \frac{\sigma}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}, \tag{4.6}$$

Where σ is the standard deviation of the residuals, ϵ_i follows a normal distribution and can be estimated from the data table. This model allows for a straightforward calculation of ρ, as the slope of the linear fit directly gives it. On the other hand, the $V \times M$ model follows the equation:

$$V = \frac{1}{\rho}M + \left(V_0 - \frac{M_0}{\rho}\right). \tag{4.7}$$

Here, the slope of the regression line is $a = \frac{1}{\rho}$, meaning that the density must be obtained by inverting the slope:

$$\rho = \frac{1}{a}. \tag{4.8}$$

However, this inversion introduces a more complex error propagation, increasing the uncertainty in ρ. The standard error in the density estimate becomes:

$$\sigma_\rho = \left|\frac{d\rho}{da}\right| \sigma_a = \frac{\sigma_a}{a^2}. \tag{4.9}$$

Since the error is magnified by the inverse square of the slope, the $V \times M$ model leads to a more significant estimate error in ρ compared to the $M \times V$ model. This increased uncertainty makes precise density determination less desirable. Thus, from a statistical standpoint, the best approach is to use the $M \times V$ model, where the density is directly obtained from the slope, minimizing the error propagation and providing a more reliable estimate of ρ.

5 Pedagogical Discussion

This experiment is a fundamental exercise in experimental physics, teaching students the essential skills required to derive, measure, and analyze physical quantities that cannot be directly observed with the available apparatus. Determining water density exemplifies how direct measurements of mass and volume can be used to estimate an unknown parameter through mathematical modeling and data analysis. This process is crucial for students to develop a deeper understanding of physical laws and how to translate observed phenomena into quantitative models.

A key learning outcome of this experiment is the necessity of deriving mathematical equations that describe physical reality. Students must establish a theoretical framework that links mass and volume to density, collect corresponding data points, and use regression techniques to estimate the model parameters. This structured approach fosters a better comprehension of how scientific models are built and refined, reinforcing that physical observables are often not directly measurable but must be inferred from empirical data.

Beyond theoretical modeling, students are introduced to statistical methods for treating experimental data. Implementing Ordinary Least Squares regression is a crucial part of the learning process, as it allows for parameter estimation by minimizing residual errors. Many students struggle to grasp the significance of this method despite its historical relevance dating back to Gauss and its continued application in modern data analysis. By working through this experiment, students gain firsthand experience applying regression to actual data, appreciating its importance in ensuring accurate and reliable estimations.

Moreover, this experiment highlights the necessity of computational methods in modern physics and engineering. Real-world datasets often contain missing values, errors, or inconsistencies, making manual data handling impractical. Encouraging students to use programming tools such as Python for data analysis fosters a computational mindset, equipping them with indispensable skills in today's data-driven scientific landscape. With advancements in machine learning and neural networks, data estimation and gap-filling techniques have become more sophisticated, and students must be aware of these evolving methodologies.

Another critical pedagogical aspect of this experiment is the emphasis on graphical representation. In an era where data literacy is increasingly essential, students must learn to interpret and construct meaningful visualizations. Many struggle with reading tables or understanding simple linear relationships between independent and dependent variables. Using graphing techniques, students develop a more precise intuition for how one physical quantity influences another within a mathematical model. These skills are vital in physics and a broad range of STEM disciplines, where data visualization is a key component of decision-making and communication.

Finally, an often-overlooked yet fundamental skill in experimental physics is the ability to write a structured scientific report. Communicating findings clearly, formally, and technically is essential for students pursuing careers in STEM fields. The ability to articulate experimental objectives, describe methodologies, analyze results, and present conclusions coherently and professionally is just as important as conducting the experiment itself. By emphasizing the scientific method in their writing, students refine their ability to document and convey their findings effectively, preparing them for future research and technical work.

In conclusion, this experiment provides a comprehensive learning experience that integrates theoretical modeling, statistical data treatment, computational tools, graphical literacy, and scientific communication. Students develop a well-rounded skill set that prepares them for more complex challenges in physics, engineering, and data science by engaging with these elements. Encouraging a rigorous approach to experimental analysis enhances their understanding of physical principles and cultivates critical thinking and problem-solving abilities essential for any scientific career.

Conclusion

This study analyzed an undergraduate physics laboratory experiment designed to determine the density of water using fundamental measurement techniques and regression analysis. The experimental setup, which includes a precision scale, a graduated container filled with water, and a suspended metal rod, allows students to develop critical skills in experimental physics. Throughout the experiment, students are challenged to derive theoretical models that link observable variables, such as mass and volume, to the physical quantity of interest—water density.

One of the main difficulties students encounter is understanding the process of model linearization and applying linear regression via ordinary least squares methods. Many struggle with interpreting the slope and intercept of the fitted linear model, which are essential for extracting meaningful physical parameters. Additionally, simple but often overlooked physical phenomena, such as frictional forces between the metal rod and the container's surface, can introduce systematic errors, affecting the results. Furthermore, students often face conceptual challenges in selecting the appropriate model—whether to analyze mass as a function of volume ($M \times V$) or mass ($V \times M$). This choice directly influences the error propagation and the reliability of the final density calculation.

This article addresses these challenges by providing theoretical and computational guidance for students to gain deeper insight into the physical interpretations of their experimental results. Python scripts are provided to fit the linear model and visualize the experimental data, reinforcing the importance of integrating computational tools into experimental physics education. The findings highlight the necessity of pedagogical approaches that bridge theoretical concepts with hands-on experimental work, ultimately fostering a more robust understanding of data analysis and physical modeling in undergraduate science and engineering curricula.

A Ordinary Least Squares

Given a dataset of N data points (x_i, y_i), the objective is to determine a linear model

$$y_i = ax_i + b \tag{A.1}$$

where a is the slope, and b is the intercept, which are the free parameters of this model. For this, we will derive the closed formulas using the Least Squares Method, which minimizes the sum of squared residuals e_i

$$e_i = y_i - ax_i - b\,, \tag{A.2}$$

So, we determine the function

$$S(a, b) = \sum_{i=1}^{N} (y_i - ax_i - b)^2\,. \tag{A.3}$$

To find a and b, we must optimize the function $S(a, b)$ concerning its parameters, computing the partial derivatives of S and setting them to zero.

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{N} 2(y_i - ax_i - b)(-x_i) = 0\,, \tag{A.4}$$

which results in the following expression

$$\sum_{i=1}^{N} x_i y_i = a \sum_{i=1}^{N} x_i^2 + b \sum_{i=1}^{N} x_i\,, \tag{A.5}$$

now performing the partial derivatives concerning b

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{N} 2(y_i - ax_i - b)(-1) = 0. \tag{A.6}$$

the above expression results in the following

$$\sum y_i = a \sum x_i + Nb\,. \tag{A.7}$$

For simplicity, we can use the following notation

$$S_{xy} = S_{yx} = \sum_{i=1}^{N} x_i y_i\,, \tag{A.8}$$

$$S_{xx} = \sum_{i=1}^{N} x_i^2 \,, \tag{A.9}$$

$$S_x = \sum_{i=1}^{N} x_i \,, \tag{A.10}$$

$$S_y = \sum_{i=1}^{N} y_i \tag{A.11}$$

So equations Eq. (A.5) and Eq. (A.7) can be put in the form

$$S_{xy} = aS_{xx} + bS_x \tag{A.12}$$

$$S_y = aS_x + Nb \tag{A.13}$$

Or in matrix form

$$\begin{pmatrix} S_{xx} & S_x \\ S_x & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix} \tag{A.14}$$

Solving for the system of equations given by Eq. (A.5) and Eq. (A.7) for a and b results in the following expression for the estimator $\hat{a}$ (slope)

$$\hat{a} = \frac{NS_{xy} - S_x S_y}{NS_{xx} - S_x^2} = \frac{\sum_{i=1}^{N}(x_i - \hat{x})(y_i - \hat{y})}{\sum_{i=1}^{N}(x_i - \hat{x})^2}. \tag{A.15}$$

Remembering that from the linear model

$$b = \hat{y} - a\hat{x} \tag{A.16}$$

where $\hat{y}$ and $\hat{x}$ are the average estimators given by

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} y_i = \frac{S_y}{N} \,, \tag{A.17}$$

$$\hat{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{S_x}{N} \,, \tag{A.18}$$

Substituting Eq. (A.17), Eq. (A.18) and Eq. (A.15) into Eq. (A.16) results in

$$\hat{b} = \frac{S_y S_{xx} - S_{xy} S_x}{NS_{xx} - S_x^2} = \hat{y} - \frac{\sum_{i=1}^{N}(x_i - \hat{x})(y_i - \hat{y})}{\sum_{i=1}^{N}(x_i - \hat{x})^2} \hat{x} \tag{A.19}$$

Notice that

$$\begin{aligned}
\sum_{i=1}^{N}(x_i - \hat{x})^2 &= \sum_{i=1}^{N} x_i^2 - 2\hat{x}\sum_{i=1}^{N} x_i + \sum_{i=1}^{N}\hat{x}^2 \\
&= \sum_{i=1}^{N} x_i^2 - 2N\hat{x}^2 + N\hat{x}^2 \\
&= \sum_{i=1}^{N} x_i^2 - N\hat{x} \\
&= \sum_{i=1}^{N} x_i^2 - \frac{1}{N}\sum_{i=1}^{N} x_i
\end{aligned} \tag{A.20}$$

B Error in Estimator a

A Step-by-Step Derivation of Variance and Standard Error of $\hat{b}$. To fully understand the derivation of the Variance and standard error of $\hat{b}$, we need to go deeper into the mathematics. Recall the Least Squares Estimate for $\hat{b}$. The least squares estimate of the slope in a simple linear regression model is given by

$$\hat{a} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\,, \tag{B.1}$$

where $\bar{x} = \frac{1}{n}\sum x_i$ is the mean of x, and $\bar{y} = \frac{1}{n}\sum y_i$ is the mean of y. This formula tells us that $\hat{b}$ is a linear function of y_i, which allows us to compute its Variance. Express $\hat{a}$ in Terms of the Error terms. The regression model assumes:

$$y_i = a + bx_i + \varepsilon_i \tag{B.2}$$

where ε_i are independent, normally distributed errors with mean zero and Variance σ^2

$$E[\varepsilon_i] = 0, \quad \mathrm{Var}(\varepsilon_i) = \sigma^2. \tag{B.3}$$

Substituting this into our equation for $\hat{a}$, leads to

$$\hat{a} = \frac{\sum(x_i - \bar{x})(b + ax_i + \varepsilon_i - \bar{y})}{\sum(x_i - \bar{x})^2}. \tag{B.4}$$

Expanding $\bar{y} = a + b\bar{x} + \bar{\varepsilon}$:

$$\hat{a} = \frac{\sum(x_i - \bar{x})(b + ax_i + \varepsilon_i - (b + a\bar{x} + \bar{\varepsilon}))}{\sum(x_i - \bar{x})^2}. \tag{B.5}$$

Since $\sum(x_i - \bar{x})\bar{\varepsilon} = 0$, simplifying gives

$$\hat{a} = a + \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}. \tag{B.6}$$

Compute the Variance of $\hat{a}$ by taking the Variance of both sides

$$\mathrm{Var}(\hat{a}) = \mathrm{Var}\left(\frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}\right). \tag{B.7}$$

Since the errors ε_i are independent and have variance σ^2

$$\mathrm{Var}\left(\sum(x_i - \bar{x})\varepsilon_i\right) = \sum(x_i - \bar{x})^2 \mathrm{Var}(\varepsilon_i) = \sigma^2 \sum(x_i - \bar{x})^2. \tag{B.8}$$

Since variance scales by $1/k^2$ when dividing by a constant k, we get:

$$\mathrm{Var}(\hat{a}) = \frac{\sigma^2 \sum(x_i - \bar{x})^2}{(\sum(x_i - \bar{x})^2)^2} = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}. \tag{B.9}$$

Thus, the error on the estimator $\hat{a}$ is given by

$$\sigma_{\hat{a}} = \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}}. \tag{B.10}$$

C Error in Estimator b

To derive the Variance and standard error of the estimator $\hat{b}$ in the regression equation

$$y = ax + b + \varepsilon, \tag{C.1}$$

we start with its least squares estimate:

$$\hat{b} = \bar{y} - \hat{a}\bar{x}. \tag{C.2}$$

Substituting $\hat{a}$,

$$\hat{b} = \bar{y} - \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\bar{x}. \tag{C.3}$$

Expressing in terms of the true model,

$$\bar{y} = a\bar{x} + b + \bar{\varepsilon}. \tag{C.4}$$

Thus,

$$\hat{b} = a\bar{x} + b + \bar{\varepsilon} - \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}\bar{x}. \tag{C.5}$$

Simplifying,

$$\hat{b} = b + \bar{\varepsilon} - \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}\bar{x}. \tag{C.6}$$

Taking variances,

$$\mathrm{Var}(\hat{b}) = \mathrm{Var}\left(\bar{\varepsilon} - \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}\bar{x}\right). \tag{C.7}$$

Since

$$\mathrm{Var}(\bar{\varepsilon}) = \frac{\sigma^2}{n} \tag{C.8}$$

and

$$\mathrm{Var}\left(\frac{\sum(x_i-\bar{x})\varepsilon_i}{\sum(x_i-\bar{x})^2}\right)=\frac{\sigma^2}{\sum(x_i-\bar{x})^2}, \tag{C.9}$$

we use the property

$$\mathrm{Var}(A+B)=\mathrm{Var}(A)+\mathrm{Var}(B)+2\mathrm{Cov}(A,B) \tag{C.10}$$

and the known covariance result

$$\mathrm{Cov}\left(\bar{\varepsilon},\frac{\sum(x_i-\bar{x})\varepsilon_i}{\sum(x_i-\bar{x})^2}\right)=-\frac{\sigma^2\bar{x}}{\sum(x_i-\bar{x})^2}. \tag{C.11}$$

Thus,

$$\mathrm{Var}(\hat{b})=\frac{\sigma^2}{n}+\frac{\sigma^2\bar{x}^2}{\sum(x_i-\bar{x})^2}. \tag{C.12}$$

Taking the square root,

$$\sigma_{\hat{b}}=\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)}. \tag{C.13}$$

Simplifying even further results in the following expression

$$\sigma_{\hat{b}}=\sqrt{\frac{\sum_{i=1}^{N}x_i^2}{\sum(x_i-\bar{x})^2}}\sigma. \tag{C.14}$$

D Python Requirements

Requirements for the functions to work. Libraries to import.

keywordstyle
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

E Python Code: Generates Synthetic Data

The function `generate_experimental_data` is designed to simulate experimental measurements of mass and volume for an object submerged in a fluid. It generates synthetic data that mimics real-world measurements, incorporating a linear mass-to-volume relationship with controlled uncertainty and random noise. These values are then computed using the linear relation

$$M=\rho V+(M_0-\rho V_0)+\text{noise} \tag{E.1}$$

where ρ is the density of the fluid, M_0 and V_0 are reference mass and volume values, and a **random noise component** (uniformly distributed between `-noise` and `noise`) is added to simulate small variations in the experimental data.

The function then constructs a pandas data frame to store and return the simulated data. This Dataframe provides a structured representation of the synthetic experimental data, which can be used to test data analysis techniques, validate theoretical models, or simulate laboratory conditions without requiring real-world measurements.

keywordstyle
```
def generate_experimental_data(n_points=10,M0=300,V0=100,rho=1,V_min
    =110,V_max=
200,err_M=0.025,err_V=0.10,noise=5):
    # Generate volume values evenly spaced between V_min and V_max.
    V_values = np.linspace(V_min, V_max, n_points)
    # Calculate the mass values from the linear relation plus noise.
    M_values = rho * V_values + (M0 - rho * V0) + np.random.uniform(-
        noise,
noise, n_points)
    # Create the data frame, including the measurement uncertainties.
    df = pd.DataFrame({
        'M': M_values,
        'sigma_M': M_values * err_M,
        'V': V_values,
        'sigma_V': V_values * err_V
    })
    return df
```

F Python Code: Fits Experimental Data

The function `calculate_fit` performs a linear regression analysis to estimate the relationship between mass and volume from an experimental dataset. It determines the best-fit parameters for a linear model, where mass is expressed as a volume function. It also computes statistical uncertainties associated with the regression parameters. It begins by extracting volume and mass values from the input data frame, determining the total number of observations, and then computing the summations required for the regression.

The slope and intercept of the best-fit line are computed using the least squares method. The slope represents the rate of change of mass concerning volume and is related to the fluid's density. The intercept accounts for initial conditions in the measurement process, meaning the values of (V_0, M_0).

Once the regression parameters are obtained, the function calculates predicted mass values using the fitted model. The residual standard error is computed to quantify how well the fitted line describes the observed data. This metric represents the typical deviation of the measured values from the regression model.

The function calculates the sum of squared differences from the mean volume to estimate the uncertainty in the regression parameters. The uncertainty in the slope is obtained by normalizing the residual standard error concerning this sum. The uncertainty in the intercept is derived by considering the number of data points and the mean volume value.

Finally, the function returns a structured output containing the estimated slope and intercept, their respective uncertainties, the predicted mass values, and the residual standard error. This output is provided as a dictionary and data frame, ensuring compatibility with further data analysis. The function enables the precise determination of fluid density and provides a statistical assessment of the measurements' reliability.

keywordstyle

```
def calculate_fit(df, M0=300, V0=100, rho=1):
    # Extract the independent (x) and dependent (y) values.
    V_values = df['V'].values
    M_values = df['M'].values
    N = len(V_values)
    # Calculate necessary summations.
    sum_x = np.sum(V_values)
    sum_y = np.sum(M_values)
    sum_x2 = np.sum(V_values ** 2)
    sum_xy = np.sum(V_values * M_values)
    # Denominator for the slope and intercept formulas.
    D = N * sum_x2 - sum_x ** 2
    # Calculate slope (a) and intercept (b).
    a = (N * sum_xy - sum_x * sum_y) / D
    #
    b = (sum_y - a * sum_x) / N
    # Predicted Mass values using the regression line.
    y_pred = a * V_values + b
    # Residual standard error.
    sigma_y = np.sqrt(np.sum((M_values - y_pred) ** 2) / (N - 2))
    # Compute Sxx for calculating uncertainties.
    Sxx = sum_x2 - (sum_x ** 2) / N
    #
    sigma_a = sigma_y / np.sqrt(Sxx)
    # Standard error for the intercept.
    x_bar = sum_x / N
    #
    sigma_b = sigma_y * np.sqrt(1 / N + (x_bar ** 2) / Sxx)
    #
    fit_params = {
        'M0': M0,
        'V0': V0,
        'rho': rho,
        'a': a,
        'sigma_a': sigma_a,
        'b': b,
        'sigma_b': sigma_b,
        'y_pred': y_pred,
        'sigma_y': sigma_y,
    }
    tb = pd.DataFrame([fit_params])
    return tb, fit_params
```

G Python Code: Plot Regression Function

The function `plot_regression` is designed to visualize the results of a linear regression model applied to a dataset containing mass (`M`) and volume (`V`) measurements. It takes as input a pandas DataFrame `df` containing the experimental data and a dictionary `fit_` `results` with regression parameters and predictions. The function first extracts the volume and mass values and their respective uncertainties, if available. It then retrieves the regression coefficients a and b along with their standard errors σ_a and σ_b. The predicted values y_{pred} are sorted for a smooth visualization of the regression line. Using error propagation, the function computes the uncertainty in the predicted values and generates an upper and lower confidence band. The data is plotted as scatter points, with error bars if uncertainties are available. The best-fit line is overlaid along with the uncertainty band to highlight the confidence interval around the regression model. Finally, appropriate labels, a grid, and a legend are added for clarity before displaying the plot.

keywordstyle

```
def plot_regression(df, fit_results):
    # Extract data from the DataFrame.
    V_values = df['V'].values
    M_values = df['M'].values
    # Check if the DataFrame contains error columns.
    V_err = df['sigma_V'].values if 'sigma_V' in df.columns else None
    M_err = df['sigma_M'].values if 'sigma_M' in df.columns else None
    # Extract the regression fit results.
    y_pred = fit_results['y_pred']
    a = fit_results['a']
    b = fit_results['b']
    sigma_a = fit_results['sigma_a']
    sigma_b = fit_results['sigma_b']
    # Sort the Volume values and corresponding predictions for a smooth line.
    sort_idx = np.argsort(V_values)
    V_sorted = V_values[sort_idx]
    y_pred_sorted = y_pred[sort_idx]
    # Compute the uncertainty in the predicted y-values using error propagation:
    uncertainty = np.sqrt((V_sorted * sigma_a)**2 + sigma_b**2)
    y_upper = y_pred_sorted + uncertainty
    y_lower = y_pred_sorted - uncertainty
    # define image
    plt.figure(figsize=(8, 5))
    # Plot experimental data with error bars if uncertainties are provided.
    if V_err is not None or M_err is not None:
        plt.errorbar(V_values, M_values, xerr=V_err, yerr=M_err,
                     fmt='o', color='black', capsize=2, label="Experimental
Data")
    else:
        plt.scatter(V_values, M_values, color='r', label="Experimental Data")
    # Plot the best-fit regression line.
    plt.plot(V_sorted, y_pred_sorted, linestyle='--', color='b',
             label=f"Best Fit: y = ({a:.2f} $\pm$ {sigma_a:.2f})x + ({b:.2f} $
\pm$ {sigma_b:.2f})")
    # Add the uncertainty band around the regression line.
    plt.fill_between(V_sorted, y_lower, y_upper, color='blue', alpha
```

```
        =0.2,
                        label='Uncertainty Band')
    plt.xlabel("Volume (mL)")
    plt.ylabel("Mass (g)")
    plt.title("Least Squares Regression: Mass vs. Volume")
    plt.grid(True)
    plt.legend()
    plt.show()
```

H Python Code: Format Table Function

The function `format_table` formats a pandas DataFrame containing mass (`M`) and volume (`V`) measurements, along with their respective uncertainties (δM and δV). The function returns a new DataFrame where the values are formatted as strings using LaTeX notation, making them suitable for direct inclusion in a LaTeX-rendered table. Each row of the formatted table represents a measurement M_i or V_i along with its uncertainty, expressed in the form $M_i \pm \delta M_i$ (in grams) and $V_i \pm \delta V_i$ (in milliliters). Additionally, the function computes the relative uncertainties, given by $\delta M_i / M_i$ and $\delta V_i / V_i$, providing a normalized uncertainty measure. The values are formatted with three decimal places for consistency. By structuring the data this way, the function facilitates the generation of properly formatted tables for scientific reports and publications where uncertainty representation is crucial.

keywordstyle

```
def format_table(df):
    return pd.DataFrame({
        "$(M_i \\pm \\delta M_i) \\ g$": [f"{M:.3f} \\pm {sigma_M:.3f}"
            for M,
sigma_M in zip(df["M"], df["sigma_M"])],
        "$\\frac{\\delta M_i}{M_i}$": [f"{(sigma_M / M):.3f}" for M,
            sigma_M
in zip(df["M"], df["sigma_M"])],
        "$(V_i \\pm \\delta V_i) \\ ml$": [f"{V:.3f} \\pm {sigma_V:.3f}"
             for
V, sigma_V in zip(df["V"], df["sigma_V"])],
        "$\\frac{\\delta V_i}{V_i}$": [f"{(sigma_V / V):.3f}" for V,
            sigma_V
in zip(df["V"], df["sigma_V"])]
    })
```

References

[1] Tutorial de Física Experimental II, 2024. IF-UFRJ. Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. http://fisexp2.if.ufrj.br/Fisexp2_Apostila2024.pdf.

[2] Folha do Relatório 1: Estudo Experimental da Força de Empuxo, 2024. Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. http://fisexp2.if.ufrj.br/Relatorios/Relatorio_1.pdf.

[3] J. H. Vuolo, *Fundamentos da Teoria de Erros*. São Paulo, Brazil: Editora Edgar Blücher Ltda., 1992.

[4] D. Halliday, R. Resnick, and K. S. Krane, *Physics, Volume 2*. Hoboken, NJ: John Wiley & Sons, 2010. ISBN: 978-0-470-56159-0.

[5] H. M. Nussenzveig, *Física Básica - Volume 2*. São Paulo, Brazil: Editora Edgar Blücher, Unknown.

[6] F. M. White, *Fluid Mechanics* (8th ed.). McGraw-Hill, 2016. ISBN: 9780073398273.

[7] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Volume 1*. Basic Books, 2011. ISBN: 9780465024933.