European Journal of Applied Sciences – Vol. 13, No. 03
Publication Date: June 29, 2025
DOI:10.14738/aivp.1303.19028.
Singh, N., Mamta, Kumar, K., Singh, I., & Arun, P. (2025). Data and Error Analysis for an Undergraduate Experiment. *European Journal of Applied Sciences*, Vol - 13(03). 492-508.

Data and Error Analysis for an Undergraduate Experiment

Nimmi Singh
Department of Physics, S.G.T.B. Khalsa College,
University of Delhi, Delhi 110007, INDIA

Mamta
Department of Physics, S.G.T.B. Khalsa College,
University of Delhi, Delhi 110007, INDIA

Kuldeep Kumar
Department of Physics, S.G.T.B. Khalsa College,
University of Delhi, Delhi 110007, INDIA

Inderpreet Singh
Department of Electronics, S.G.T.B. Khalsa College,
University of Delhi, Delhi 110007, INDIA

P. Arun
Department of Electronics, S.G.T.B. Khalsa College,
University of Delhi, Delhi 110007, INDIA

ABSTRACT

Under-graduate students in the physics lab are expected to report their experimental results with possible errors or uncertainties, however, our experience at under-graduate level classes is that the students do not understand the significance of error analysis in an experiment and carry a lot of misconceptions. In an age, where, students are looking for quick solutions/ reads, the textbooks on error analysis and statistical mathematics seem to be too lengthy, abstract, theoretical and mundane. To address this problem, we decided to put together an article that is a short version of error analysis. The present work develops the standard ideas of error analysis around the data collected from a simple Ohm's law experiment, an experiment done by almost all science students at school level.

Keywords: Data Analysis, Least square fitting, Ohm's Law.

INTRODUCTION

The objective of doing an experiment in Physics or Electronics undergraduate laboratory is to compute some physical parameter using data obtained and verify a theory or physical law. The final step of doing an experiment is data analysis and commenting on the accuracy and precision with which the experiment was done. A common notion among students is that if experiments are done carefully, one will get standard values as listed in books. The student's faith in his/her experiment is also evident from the faith they peg on the absolute error.

$$Absolute\ Error\ =\ |Y_{exp} - Y_{true}| \tag{1}$$

or the relative percentage error

$$\%\text{Error} = \frac{|Y_{exp} - Y_{true}|}{Y_{true}} \times 100\% \tag{2}$$

This definition of error assumes that some exact or true value exists, which is far from true! Usually the students take the value of Y_{true} from some standard Tables or internet. However, true value is never known, i.e., students fail to realize that the values given in books too have been obtained by experiments and hence are error prone as well.

On the onset, it is important for a student to realize that the presence of errors does not mean that the experiment done was wrong or useless. On the other hand, it only means that every time a measurement is made, a different value would be returned. These obtained values are usually found to vary in and around what we can make out as the average value.

In fact an experimental result is subjected to several errors or uncertainties (we actually report uncertainties consider for knowing an error we need to know the true value) usually categorized into systematic errors[1] and statistical or random error[2]. A reduction in systematic errors improves the accuracy[3] of result while a reduction in the random errors improves the precision[4] of result. Reporting of our results with an estimation of uncertainty enables us to not only make judgments about the quality of the experiment but also to comment about the expected result if the experiment is repeated.

This argues well for a comprehensive review of "error and its analysis" broken to the level of first year under-grads. For doing the same, we analyse the results of a simple experiment like Ohms law verification which is usually introduced to a student at the age of fifteen (in India). In the process, we also investigate and comment on the nature of instruments to be selected for the conduction of this experiment. In passing, we also compare the best method of analysing results of such a simple experiment. The reader will note that this selection which they might be prejudiced to think is trivial, is not so simple.

ESTIMATION OF UNCERTAINTY

We now address the question "How to estimate and report the result with error?"

---

[1] Systematic Errors are reproducible inaccuracies that result from some systematic ef fect such as faulty equipment, wrong calibration, or technique. The zero error or incorrect calibration of an instrument are examples of systematic error and can be completely ac counted for if the source is known because they affect all measurements in some well-defined way. These are usually difficult to detect and cannot be analyzed statistically. Reduction in systematic errors improves the accuracy of the result.

[2] Random errors refer to the spread in the values of a physical quantity from one mea surement of a quantity to the next, caused by random fluctuations in the measured value and can be reduced by increasing the number of observations. The source of statistical f luctuations can be the precision limitations of the measurement device, variation in envi ronmental factors, reaction times etc. These can be evaluated through statistical analysis- without reference to the source of errors.

[3] Accuracy is a measure of how close a measurement is to the correct value of the quantity being measured

[4] Precision is a measure of how close a series of measurements are to one another without reference to a true value.

Instrumental Error

Taking measurements in an experiment involves reading of some scale or digital display. In the present work, we restrict our discussions to analog devices. The uncertainty in this measurement is called reading error or instrumental error and refers to the uncertainties caused by the limitations of our measuring equipment. For an analog scale, the instrumental error Δx is usually taken to be equal to or half of the least count of the instrument. Thus, if Δx is the estimate of absolute instrumental error (uncertainty) in the measurement x of a quantity, the value x could have been anywhere in the range [x − Δx, x + Δx].

Usually in an experiment, quantity of interest (x) is not measured directly, but is calculated from other measured quantities. The errors of the direct measurements propagate to the finally calculated value and lead to an uncertainty Δx in the final value of x. How to calculate this uncertainty?

Let us take the example of Ohms law, where resistance (R) of a resistor is to be determined. It involves measuring the values of V and I followed by which the value of resistance is evaluated using

$$R = \frac{V}{I} \tag{3}$$

i.e. it is a derived quantity because the measured voltage (V) across the resistor and current (I) flowing through it is used to determine the resistance (R = V/I). The error in R due to the instrumental error is given as

$$ln(R) = ln(V\ ) + ln(I) \tag{4}$$

The negative sign that should have appeared before *ln(I)*, is replaced with a positive sign to evaluate the maximum possible error. Hence

$$\Delta R = \pm R\ \left(\left|\frac{\Delta V}{V}\right| + \left|\frac{\Delta I}{I}\right|\right) \tag{5}$$

The instrumental error computed this way is also called “log error”. Since, the equation evaluates the maximum possible error, this equation overestimates the error. A better estimate of the uncertainty is given by law of propagation of error. Using eqn (5) and taking the square of this expression, we have

$$\left(\frac{\Delta R}{R}\right)^2 = \left(\frac{\Delta V}{V} + \frac{\Delta I}{I}\right)^2$$

$$\left(\frac{\Delta R}{R}\right)^2 = \left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta I}{I}\right)^2 + 2\left(\frac{\Delta V}{V}\right)\left(\frac{\Delta I}{I}\right)$$

Since the third term would be very small (approaching zero), we can write Taking the square of this expression, we have

$$\left(\frac{\Delta R}{R}\right) = \sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta I}{I}\right)^2} \tag{6}$$

$$\Delta R = \pm R\left[\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta I}{I}\right)^2\right]^{1/2} \tag{7}$$

The best estimates for ΔI and ΔV are the least counts of the ammeter and voltmeter, respectively. From equations (5), it is clear that if the measured values of current and voltage are large as compared to the least count of the instruments, the error would be small. This should make it obvious to the students that while selecting an instrument he/she should select it such as to get near maximum deflections. Based on these concerns, listed above, instrumental error can be minimized. Although, it should be mentioned here that with improving technology the instrumental errors have been constantly decreasing but they are still significant in under-graduate labs as the best available instruments are not provided to students.

Statistical Error

Consider a student who repeats an experiment repeatedly under identical conditions with the same instruments. Even with utmost precautions his/her results would never be the same. In fact, the results would be distributed around some mean value. The spread in the values of a physical observable comes from either practical limitations of any measuring instrument or the change in surrounding conditions like temperature, pressure etc. Thus, the spread in values is also a measure of uncertainty in measurement. If a quantity x is measured several times (say N), it is assumed that the best estimate of the measured value is the average value[5] (μ) and the variation or spread among the measured values (quantified by the standard deviation[6], σ) gives an idea of the uncertainty in the measurement. To be precise, the estimate for the uncertainty in the mean value[7] is given by

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \tag{10}$$

and the value of quantity x is reported as $\mu \pm \sigma_{\bar{x}}$. The relation (10) should convince the students that higher is the value of 'N', smaller is the uncertainty in mean value.

---

[5] Mean (or average value) of N number of observations xi, i = 1, 2, ... N is defined by

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{8}$$

[6] The standard deviation 'σ' is given by

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{\infty}(x_i - \mu)^2} \tag{9}$$

[7] This is sometimes wrongly called random error of the mean. It is a measure of the uncertainty of the mean due to random effects. The exact value of the error in the mean arising from these effects cannot be known

The Central Limit Theorem[8] (CLT) assures that the mean of a large number of random variables approximately follows a normal distribution. Thus, the random errors (or the deviation of measurement from true value) are Gaussian[9] in nature and their probability distribution function is given by

$$\mathcal{P}(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{11}$$

where 'μ' is the average value and 'σ' is called the standard deviation. Larger is the value of σ, larger will be the spread of values showing a lack of precision in conduction of the experiment. If the errors are only due to random effects and there is no systematic error in the experiment, then the errors will have a normal distribution with μ = 0.

Rather than proving CLT mathematically (which can be found in any statistics book by interested students), we verify it by simulation using a python program. Consider the experiment of rolling N dice together k number of times. The number on each dice is a discrete random variable that can take a value 1 to 6. A roll of a single true dice when thrown will give a number among 1-6 with the probability of getting any of these numbers being $\frac{1}{6}$. Thus, rolling of single dice produces a uniform distribution with mean μ = 3.5 and standard deviation σ = 1.7078 as shown in the histogram of top most panel of Figure 1. However, as the number of dice increases, the histogram (distribution from simulation) matches the Gaussian curve computed using Equation (11) with mean 3.5 and standard deviation 1.7078/ N with N = 1,10 and 30. In the present work, we took 30 sets of I − V characteristics with 10 data points in each set. Here V is the independent variable while I is the dependent variable. Analysis can then be done by either of the following methods

(i) determine the value of R for each set and thereafter calculate the average resistance (R_{av}).

(ii) fit data $(V_i, \bar{I}_i)$ using the method of least squares (which would be explained briefly in section 3.2.1), where $\bar{I}_i$ is the average value of thirty current values measured for Vi.

THE OHM'S LAW EXPERIMENT

Ohms Law is a simple experiment that is introduced at high school level. Being linear in nature, we felt it appropriate to use data of Ohms law to explain the concepts of error analysis. Figure (2) shows the circuit used for making measurements.

Direct Evaluation by R=V/I

Students commonly use what is called the direct method, which is indeed the simplest way of determining the value of resistance.

[8] The Central Limit Theorem states that average (μ) of n independent and identically distributed random variables (X1, X2, X3, ⋯, Xn) each taken from a distribution with mean μ and standard deviation σ follows a distribution that is approximately normally with mean μ and and S.D. σ/√n for large enough n (usually n ≥ 30 is considered to be large enough). This is true for any distribution followed by variables Xi.

[9] 10If a random variable X follows a Gaussian or Normal distribution given by probability density function (11), and the probability that X has a value in the range [x, x + dx] is given by P(x)dx

Table 1: The least count and full scale deflections of the instruments used for the first set of observations.

	voltmeter	ammeter
Least Count	0.1 V	1 mA
Full deflection	5 V	50 mA

Table 2: The least count and full scale deflections of the instruments used for the second set of observations.

	voltmeter	ammeter
Least Count	0.05 V	0.5 mA
Full deflection	3 V	25 mA

First Set:

We have conducted the experiment twice using analog meters with different least counts. The least counts and full deflections of these analog meters are listed in Table 1 and Table 2.

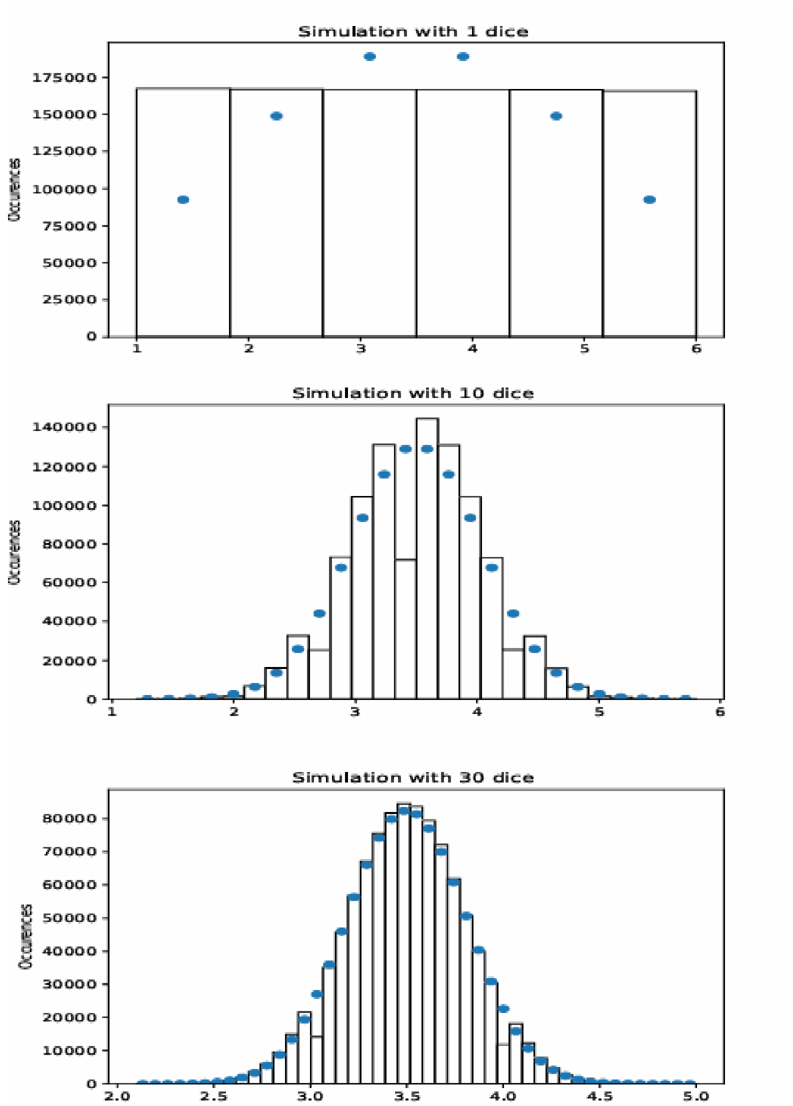

Figure 1: Results of simulation (using pseudo random number generator in a Python program) of rolling together N dice k = 10^6 times. x−axis denotes the mean of numbers obtained on N dice (which will be a number between 1 and 6) and y−axis gives the number of occurrences of each of these in the k throws. Each histogram is the distribution of the mean of N numbers drawn from the uniform distribution U (1,6) while the scatter points are the expected values (the Gaussian limit for the distribution expected from from CLT) i.e. values computed using Equation (11) with mean 3.5 and standard deviation 1.7078/ $\sqrt{N}$ with N = 1,10 and 30.

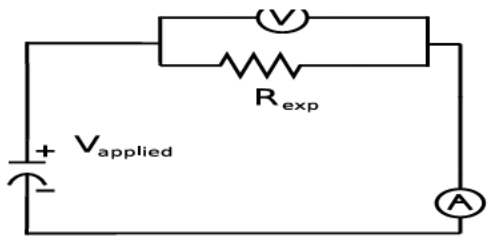

Figure 2: Circuit diagram of the set up used to perform Ohm's law experiment.

Table 3: Lists the thirty set of observations made. The first row in this table confirms that zero current flows when no voltage is applied. Thus, there is no zero error in our instruments. We now use the data in the first column of Table 3 and use the direct method (given in (3)) to compute the reistance (see Table 6).

V	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	I_{20}	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{27}	I_{28}	I_{29}	I_{30}
0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.5	5	5	6	5	5	5	5	5	5	6	5	6	5	5	5	6	5	5	5	6	5	5	5	5	5	5	5	5	5	5
1.0	11	10	11	11	10	10	10	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
1.5	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	17	16	16	16	16	16
2.0	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
2.5	27	28	27	27	28	27	27	27	27	27	27	27	28	28	28	28	28	28	28	28	28	27	27	27	27	27	27	27	27	27
3.0	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
3.5	39	38	38	38	38	39	38	39	38	38	38	38	38	39	39	39	38	38	39	39	39	38	38	39	38	38	38	39	38	39
4.0	44	44	44	44	44	44	44	44	44	44	44	43	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44
4.5	50	49	50	49	49	50	50	50	49	50	50	50	50	50	49	50	50	50	50	50	50	49	49	49	49	49	50	50	50	50

Table 4: The values of R and fractional uncertainty $\frac{\Delta R}{R}$ are listed for the first set of observations of Table 3

V (volt)	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
I (mA)	0	5	11	16	22	27	32	39	44	50
RΩ	_	100	90.91	93.75	90.91	92.59	93.75	89.74	90.91	90.00
$\frac{\Delta R}{R}$ **(eq 5)**	–	0.40	0.19	0.13	0.096	0.077	0.065	0.058	0.048	0.042
$\frac{\Delta R}{R}$ **(eq 7)**	–	0.28	0.14	0.091	0.068	0.054	0.046	0.038	0.034	0.030

It is clear from the values in Table (4) that higher are the values of V and ΔR I, lower is the fractional reading uncertainty, $\frac{\Delta R}{R}$. Further, specifically, the R instrumental uncertainty is less than 10% for the readings in range ≥ 50% of the full-scale deflection in the measuring instruments. For our case, this means that it is prudent to only consider the readings for V ⩾ 2.0volts. This is in line with the discussion in the section 1, where it was mentioned that if the measured values of current and voltage are large as compared to the least count of the instruments, the uncertainty would be small.

From the values in Table 4, we can calculate the mean value of resistance using eqn (8)

$$\bar{R} = \frac{1}{9}\sum_{i=1}^{9} R_i \approx 92.5\Omega \tag{12}$$

We compute the uncertainty in each R using eqn (7) and then the average uncertainty ($\Delta\bar{R}$). We report the result of the first-set using direct evaluation of resistance as

$$R_{result} = \bar{R} \pm \Delta\bar{R} \text{ (eq (5))}$$

$$R_{result} = 92.5 \pm 11.6\Omega \tag{13}$$

The value of the resistance being measured lays somewhere between 81 Ω and 104 Ω. That is, there is an uncertainity of ±11 Ω, which is the uncertainty in our experiment. However, remember this uncertainty given by eq (5) is an over-estimation. To overcome this problem of over-estimation, Taylor [] has given a better method for estimating error. Based on the rules of propagation of errors, he showed that the errors added in quadrature (eq (7)).

The result using this for our first-set would be reported as

$$R_{result} = \bar{R} \pm \Delta\bar{R} \text{ (eq (7))}$$

$$R_{result} = 92.5 \pm 8.2\Omega \tag{14}$$

i.e. the value of resistance will lay somewhere between 84 − 100 Ω. As stated, eqn (5) reports the maximum possible uncertainty and gives a larger margin of uncertainty, while the margin given by eqn (7) will always be less than the maximum possible uncertainty.

Second Set:

Table 6: The values of R and fractional uncertainty are listed for the second set of observations of Table 5V (volts)

V (V)	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.25	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
0.50	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.5	4.0	4.0	4.0
0.75	6.0	6.0	6.0	6.0	6.0	6.0	6.5	6.0	6.5	6.5	6.5	6.5	6.5	6.5	6.5
1.00	8.5	8.0	8.0	8.5	8.5	8.5	8.5	8.5	8.0	8.5	8.5	8.5	9.0	8.5	8.5
1.25	10.5	10.5	10.5	10.5	11.0	11.0	10.5	11.0	11.0	10.5	10.5	10.5	11.0	10.5	10.5
1.50	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.5	13.0	13.0	13.0
1.75	15.0	15.0	15.0	15.0	15.0	15.5	15.0	15.0	15.0	15.5	15.0	15.0	15.0	15.0	15.5
2.00	17.0	17.0	17.0	17.0	17.0	17.0	17.0	17.0	17.0	17.5	17.0	17.0	17.0	17.0	17.0
2.25	19.0	18.5	19.0	18.5	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.0	18.5	19.0	19.0
2.50	20.5	20.5	20.5	20.5	21.0	21.0	21.0	21.0	21.0	20.5	20.5	21.0	20.5	20.5	21.0
2.75	23.5	23.0	23.0	22.5	23.5	23.5	23.5	23.5	23.5	23.5	23.0	23.5	23.5	23.5	24.0

V (V)	I_{16}	I_{17}	I_{18}	I_{19}	I_{20}	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{27}	I_{28}	I_{29}	I_{30}
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.25	2.0	2.5	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
0.50	4.0	4.5	4.5	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
0.75	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.5
1.00	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5
1.25	10.5	11.0	11.0	10.5	11.0	11.0	10.5	11.0	11.0	11.0	11.0	11.0	11.0	10.5	11.0
1.50	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0
1.75	15.0	15.0	15.5	15.0	15.0	15.5	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
2.00	17.0	17.0	17.5	17.0	17.0	17.0	17.0	17.0	17.0	17.5	17.0	17.5	17.5	17.0	17.5
2.25	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.0	19.5	19.0	19.0	19.0
2.50	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0
2.75	23.0	23.5	23.5	23.5	23.5	23.5	23.5	23.0	23.5	23.0	23.0	23.0	23.5	23.0	23.0

A cursory study of current listed for a given voltage in Table 3 shows very small of no variations. In fact in some cases all thirty columns have the same current values. An immediate reasoning for the lack of scattering in data could be associated with the least count of the ammeter used (Δ I = 1 mA). One could reason more scattering would be present in the data if a device of lower least count is used. This is indeed the case seen when an ammeter with least count of 0.02 mA was used (Table 5). One would also wonder, "will we get a result with less uncertainty using instruments with lower least counts?"

To answer this question, seek result of our resistance with our new set of instruments. The result of the second-set done by direct evaluation of resistance, using (eq (7)) works out to be

$$R_{result} = \bar{R} \pm \Delta\bar{R} \text{ (eq (7))}$$

$$R_{result} = 120 \pm 10.6\Omega \quad (15)$$

i.e. the value of resistance is between 109 − 130 Ω. Comparing the results given in eqn 14 and eqn 15, it would seem that the selection of instruments with lower least count has not helped (atleast for the direct method). While the $\bar{R}$ has changed and the uncertainity in measurement $\Delta\bar{R}$ has not improved or at worst has only improved marginally.

The uncertainty in measurement is indeed related to instrumental error, however instrumental error is just a component. Clearly, while in our second set we have decreased the instrumental error, the remaining contributing factors have increased and hence apparently the uncertainty in measurement remains unaltered. The first suspect would be what we described as the statistical error. Though the direct method is simple and intutively easily understood by the students, the method assumes the validity of Ohms Law (a homogeneous and linear relationship between current and voltage) and does not exactly verify it. Hence, we shall reject this method. In the next section, we look into a method which does away with this assumption and is more appropriate for verification of Ohms law and experimental evaluation of the resistance.

Regression Method

The best way to analyse experimental data of Ohms law is to plot the I − V characteristics. The value of Vapplied was varied in steps equal to or in integral multiple of least count of the voltmeter. This was done to get precise readings on voltage across the resistance (without needle pointing between two graduations). It is done to ensure that instrumental error in voltage measurements (the independent variable) are minimal and all errors are forced on the current measurement (the dependent variable). In the case of Ohms law, if regression line is written as I = mV + c, then the slope (m) will give $m = \frac{1}{R}$ and c (the intercept) is expected to be zero. However, experimentally c never turns out to be zero because of the uncertainties in the measurements. A goodness of fit test (like correlation coefficient, chi square etc.) can be employed on our data to verify if the data is consistent with the linear nature of Ohms Law and if the resistance is indeed ohmic.

Regression of Single Data Set:

The first step in analysing any experimental data set, (x_i, y_i) is to make a scatter plot and observe if the observations really show a linear behaviour or not (at least approximately) and whether there is any outlier[10]. In general, all the data points do not lie on a line. The best fit line or the

[10] Sometimes a single bad measurement can have a large effect on the LSF values of slope and intercept. In such cases, either one should repeat that measurement or if not possible, better to omit such an outlier from the fit.

regression line written as y = mx + c is obtained by the method of least square, wherein the values of m and c correspond to the values that minimize the sum of squares of the deviations of the measured values yi from the predicted values, $y(x_i) = mx_i + c$. Mathematically, the method tries to minimize [1] the sum of squares of observations

$$\sum[y_i - y\,(x_i)]^2 = \sum[y_i - (mx_i + c)]^2 \tag{16}$$

resulting in the intercept (c) and slope (m) of the best fit line to be

$$c = \frac{1}{\Delta}\left[\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i\right]$$
$$m = \frac{1}{\Delta}\left[N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i\right] \tag{17}$$

where the denominator in both expressions is given as

$$\Delta = N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2 \tag{18}$$

First Set:

Figure 3 shows the I_1 vs V plot for the readings in first column of Table 3 along with the best fit line. Visually, the scatter plot show data points follow a linear trend, though all the points do not fall on the line.

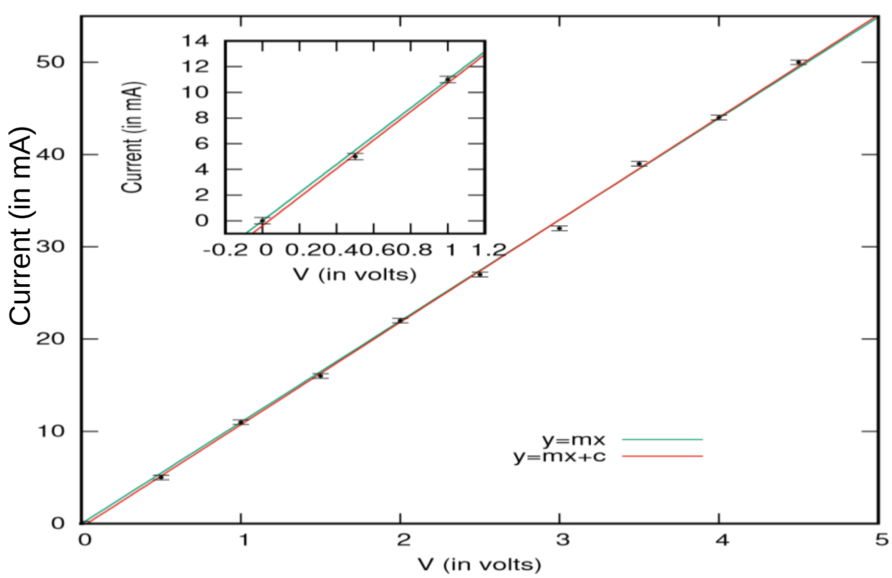

Figure 3: I-V plot for the first column of readings (Table 3). The graph displays are the observation points along with error bars. The lengths of error bars are equal to the common y–

error (σ_y = 0.23mA) obtained from equation (19). The red line is the least square fit line y = mx + c. The green line corresponds to the least square fit line corresponding to the model without intercept, i.e. y = mx. Inset shows enlarged region of graph.

To comment on how well do the data points fall on the trend line, a figure of merit is usually used called the Pearson's regression coefficient (R^2) [2]. It tells us how good the linear fitting to the data is. In our case the value obtained is R^2 = 0.999, which corresponds to a very good fit. The points are the observation points along with the error bars. The length of the error bars are equal to the common y– error obtained from equation[11]

$$\sigma_y^2 = \frac{1}{N-2}\sum_{i=1}^{N}[y_i - (c + mx_i)]^2 \tag{19}$$

σ_y in this case works out to be = 0.23mA. This uncertainity propogates to the values of slope and intercept, determined while fitting the regression line. These uncertainities are given as

$$\sigma_c^2 = \frac{\sigma_y^2}{\Delta}\sum_{i=1}^{N} x_i^2 \ and \ \sigma_m^2 = \frac{N\sigma_y^2}{\Delta} \tag{20}$$

Calculations on our data give

$$I = (m \pm \sigma_m)V + (c \pm \sigma_c)$$
$$I(inmA) = (11.1 \pm 0.1)V + (-0.4 \pm 0.3) \tag{21}$$

From these regression parameters now we can determine the value of resistance and uncertainty in it. For this we again fall back on Taylor's [3] discussion on propagation of errors. This is based on the derivative, i.e.

$$\sigma_R = \left|\frac{\partial R}{\partial m}\right| \sigma_m \tag{22}$$

where one needs to know the functional dependence of *R* on *m*. From relation R=1/m, we have

$$\sigma_R = \frac{\sigma_m}{m^2} \tag{23}$$

The error can now be reported as

$$R = \frac{1}{m} \pm \sigma_R$$

$$R_{y=mx+c} = (91.0 \pm 0.8)\Omega \tag{24}$$

[11] Here N-2 factor in the denominator signifies the number of degree of freedom. Since we have determined two paramteres (m and c), the degree of freedom is N-2.

There is a high relative error in the intercept (σ_c). This is usually the case when the intercept is ideally supposed to be zero from the theoretical model.

As shown in Figure 3, least curve fit data should always be accompanied with error bars. Since the process of least square fitting assumes that the uncertainties in y variable (in this case I) are Gaussian, then 68% (about 2/3) of the points should lie on the fitted line within their error bars, and only one point in 20 should be more than twice its error bar off the line while no points can be more than three error bars off. If too many points lie off the line, it means that the uncertainties are underestimated or there is a systematic effect (and hence do not follow the Gaussian distribution) while in case too few points are outside the error bars it means that the uncertainties are either overestimated or they are not really Gaussian.

From Figure 3, we see that three points (out of 10) are outside the error bars. If error bars corresponding to the least count of milli-ammeter is used only one data point lines off the line (these error bars have not be used in figure for brevity). Clearly, the least-count of the instruments give an overestimation of uncertainty in values. On the other hand the uncertainties obtained from equation (19) are better estimates. Figure 3 shows two least square fit lines – the red line corresponding to y = mx + c and the green line for y = mx, where we force the intercept to be zero based on Ohms law. The value of resistance obtained from the fitted line y = mx is given by

$$R = \frac{1}{m} \pm \sigma_R$$
$$R_{y=mx} = (91.0 \pm\ 0.5)\Omega \tag{25}$$

Table 7: Summary of results obtained by fitting the data given in Table 3 using first column to the models y = mx + c and y = mx. σm is the uncertainty in calculated slope. σc is the uncertainty in calculated intercept. $e_i = y_i - y_{pred}$ is the ith residual.

Model	R (in Ω) (in Ω)	$\left\|\frac{\sigma_m}{m}\right\|$	$\left\|\frac{\sigma_c}{c}\right\|$	$\sum_{i=1}^{10} e_i^2$ in (mA)^2
$y = mx + c$	90.1 ± 0.8	$\frac{0.1}{11.1}$	$\frac{0.3}{0.4}$	1.81
$y = mx$	91.0 ± 0.5	$\frac{0.06}{11.0}$	-	2.23

This is consistent with the value given in equation (24) obtained by fitting the model y = mx + c. From the results listed in Table 7, it is clear that analysis using Least square fitting is better than the direct method, considering that the direct method aprior assumes the validity of linear relation between voltage and current.

Second Set:

We now analyze the data in the first column of Table 5 which was obtained with voltmeter and ammeter of higher precision. The resistance was computed using the method of least squares and the results obtained are summarized in Table 8.

Table 8: Comparison of results obtained by fitting the data given in first column of Table 5 to the models y = mx + c and y = mx. σm is the uncertainty in calculated slope. σc is the uncertainty in calculated intercept. $e_i = y_i - y_{pred}$ is the ith residual.

Model	R (in Ω)	$\left\lvert \frac{\sigma_m}{m} \right\rvert$	$\left\lvert \frac{\sigma_c}{c} \right\rvert$	$\sum_{i=1}^{10} e_i^2$
	(in Ω)	(in mA)	Coefficient	in (mA)^2
$y = mx + c$	117.8 ± 1.3	$\frac{0.09}{8.49}$	$\frac{0.15}{0.09}$	0.775
$y = mx$	118.4 ± 0.7	$\frac{0.05}{8.44}$	-	0.802

From the values of slope and intercept of fig 4, we have obtained the resistance and the uncertainity associated with our measurement. Results are compiled in Table 8. Comparing the results listed in Table 7 and Table 8 (focus on residual of errors, $\sum_{i=1}^{10} e_i^2$), we can relate that instruments with small least count, give lower residual errors. This suggests that the precision of the readings improves on using more precise instruments. Also, the deviation from the trend line reduces with more precise instruments. The deviation or variance is proportional to the residual error. A decrease in variance invariably gives a lower residual error.

In the next section, we proceed to look into the advantage of repeating an experiment many times and analyse for computing results.

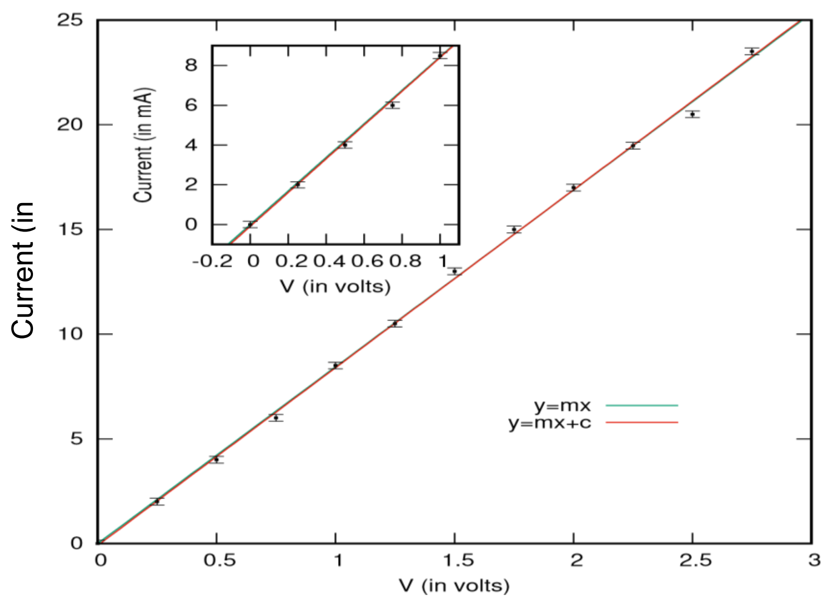

Figure 4: I-V plot for the first column of readings (Table 5). The graph displays the observation points along with error bars. The lengths of error bars are equal to the common y– error ($\sigma_y \approx$ 0.27mA) obtained from equation (19). The red line is the least square fit line y = mx + c. The green line corresponds to the least square fit line corresponding to the model without intercept, i.e. y = mx. Inset shows enlarged region of graph.

Analysis using Regression Line for Multiple Data Set

For statistical data analysis it is advisable to take large number of observations.

Hence, we took 30 set of readings for voltages and corresponding currents, thus resulting in observations (V_i, I_{ij}), with i = 1, 2,..., 10 and j = 1, 2,..., 30. All observations were taken under the same conditions.

There are two strategies to analyse this data:

- Fit data of each column (V, I_j), find the resistance from the slope of each fitting and then take the mean of all resistances. The uncertainty in final result is then determined from the spread of resistances.
- The other approach is to use the mean currents

$$\bar{I}_i = \frac{1}{30}\sum_{j=1}^{30} I_{ij} \tag{26}$$

and perform least square fitting with data $(V_i, \bar{I}_i)$.

The second approach is more popular and less time consuming. Both methods give results consistent within experimental uncertainities.

First Set:

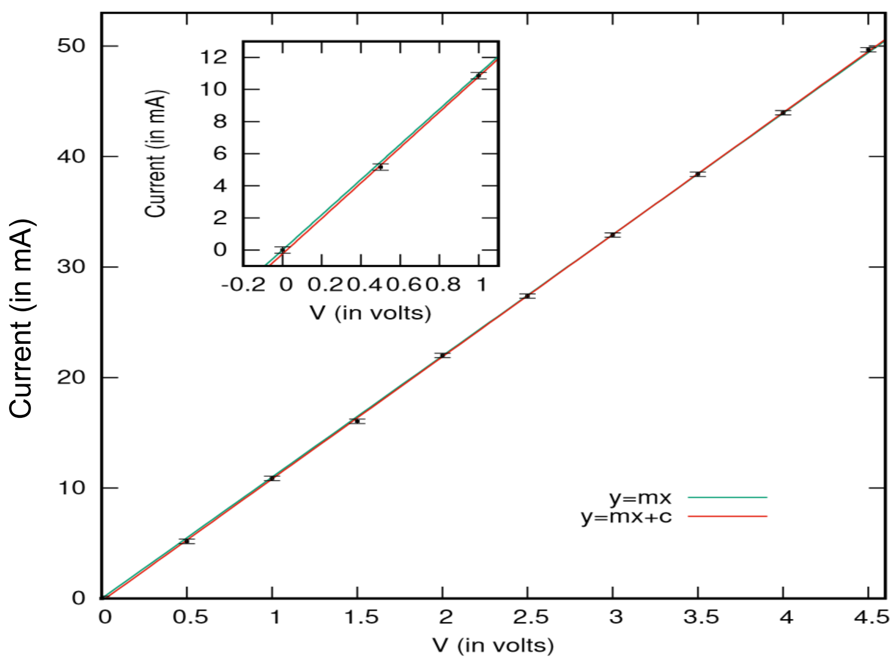

Figure 5: I-V plot for the $(V_i, \bar{I}_i)$ data for Table 3. The points are the observation points along with error bars. The lengths of error bars are equal to the common y− error ($\sigma_y \approx 0.2$mA) obtained from equation (19). The red line is the least square fit line y = mx + c. The green line corresponds to the least square fit line corresponding to the model without intercept, i.e. y = mx. Inset shows enlarged region of graph.

The least square fitting is preformed using the mean currents i.e. $(V_i, \bar{I}_i)$ of Table 3 (see fig 5). However, the error analysis is done using σ_i (instead of σ_y), where σi is the standard deviation in mean of I_i, i.e.

$$\sigma_i^2 = \frac{1}{29}\sum_{j=1}^{30}(I_{ij} - \bar{I}_i)^2 \tag{27}$$

and the uncertainity in the ith mean current is

$$\sigma_{\bar{I}_i} = \frac{\sigma_i}{\sqrt{30}} \tag{28}$$

Table 9: Comparison of results obtained by fitting the average current calculated using the data given in Table 3 to the models y = mx + c and y = mx.

Model	R (in Ω) (in Ω)	$\left\lvert\frac{\sigma_m}{m}\right\rvert$	$\left\lvert\frac{\sigma_c}{c}\right\rvert$	$\sum_{i=1}^{10} e_i^2$ in (mA)^2
$y = mx + c$	90.4 ± 0.3	$\frac{0.04}{11.06}$	$\frac{0.1}{0.2}$	0.22
$y = mx$	91.1 ± 0.2	$\frac{0.02}{10.99}$	-	0.39

We shall use σ_i in further calculations of σ_m and σ_c instead of σ_y as in the preceeding section. The equation of the fitted line, y=mx+c to $(V_i, \bar{I}_i)$ data is

$$I\ (\text{mA}) = (11.06 \pm 0.04)\ V + (-0.2 \pm 0.1)$$

and the corresponding value of resistance obtained from slope is

$$(R_{\bar{I}})_{y=mx+c} = (90.4\ \pm\ 0.3)\Omega \tag{29}$$

Similarly, the result on doing a linear curve fit, y=mx for the data gives

$$I\ (mA) = (10.99 \pm 0.02)\ V \tag{30}$$

The resistance value via this calculation is

$$(R_{\bar{I}})_{y=mx} = (91.1\ \pm\ 0.2)\Omega \tag{31}$$

Results are summarized in Table 9.

Second Set:

Similar analysis was performed for the second data set (Table 5). Fig 6 shows the data fitting done for points $(V_i, \bar{I}_i)$. The results are summerized in Table 10.

Table 10: Comparison of results obtained by fitting the average current calculated using the data given in Table 5 to the models y = mx + c and y = mx.

Model	R (in Ω) (in Ω)	$\left\lvert\frac{\sigma_m}{m}\right\rvert$	$\left\lvert\frac{\sigma_c}{c}\right\rvert$	$\sum_{i=1}^{10} e_i^2$ in (mA)^2
$y = mx + c$	117.8 ± 0.9	$\frac{0.06}{8.48}$	$\frac{0.01}{0.1}$	0.374
$y = mx$	117.9 ± 0.4	$\frac{0.03}{8.48}$	-	0.374

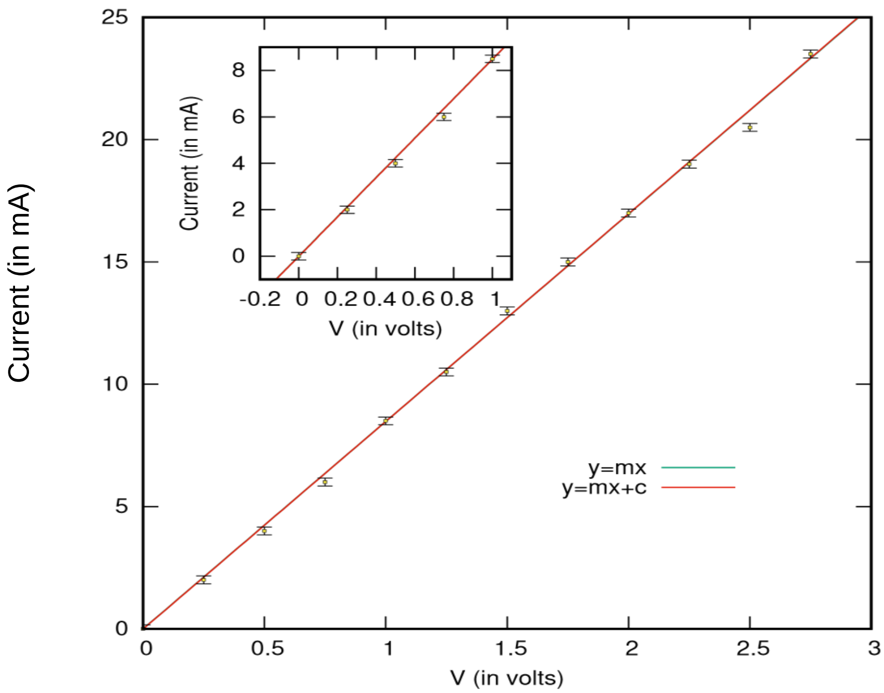

Figure 6: I-V plot for the first column of readings (Table 5). The graph displays the observation points along with error bars. The lengths of error bars are equal to the common y– error ($\sigma_y \approx$ 0.27mA) obtained from equation (19). The red line is the least square fit line y = mx + c. The green line corresponds to the least square fit line corresponding to the model without intercept, i.e. y = mx. Inset shows enlarged region of graph.

Comparing the results of Table 7, 8, 9 and 10. The residual in error decreases when evaluation is done with taking the average of 30 readings. The variance in 'y-axis' decreases when the statistical average is used to evaluate the trend line. This results from the random errors of less precise instruments averaging out over repeated measurements. This is a restatement of the Law of Large Numbers [4] that state, "as the numbers of data points increases, the mean converges to the true mean". Thus, the precision of the instruments used is not a factor when statistical average is used, emphasising the appropriateness of using regression method of multiple sets of reading.

However, a pertinent question still remains, whether y=mx+c or y=mx should be used for analysis. Again comparing the results of Table 7, 8, 9 and 10 and stated in the preceeding section, we find that the value of intercept returned is always less than the least count of the instrument used. Thus, the confidence on the value is low and can be ignored. Implying that the data falls on the y=mx (I=V/R) trend, hence proving Ohms law.

SUMMARY

This work investigates using experimental data of an Ohms law experiment, the best way to analyse it. We have listed the advantages and limitations of direct method, regression of "one shot" experiment and regression with multiple sets of readings. The direct method involving calculation of V/I was rejected on the ground that it a prior assumes the validity of Ohms law and second is less error prone if reading above 50% of full scale deflection is taken into consideration. In th case of regression of "one shot" experiment, to minimize the uncertainity in an experimentally estimated value, an experimentalist must carefully select the measuring devices to be used. This does not practically help since a decrease in least count of an instrument invariably is accompanied with a decrease in (full deflection) range of the instrument. It is clear from the results, to analyse an experiment's data, it is best to repeat the experiment numerous times. The lower variance along with lower residual error clearly justifies this conclusion. This, also makes the result fairly independent of the precision of the instruments used. We hope that the method adopted and presentation of results would be of use to instructors and students alike in understanding the choice involved in analysing an experiment's data.

References

[1] P. C. Hansen, Victor Pereyra and Godela Scherer, "Least Squares Data Fitting with Applications", John Hopkins University Press, Baltimore (USA) 2013.

[2] Mildred L. Patten, Michelle Newhart, "Understanding Research Methods", Routledge 5th Ed, USA 2017.

[3] John R. Taylor, "An Introduction to Error Analysis", University Science Books, CA (USA) 1982.

[4] Hossein Pishro-Nik, "Introduction to Probability, Statistics, and Random Processes", Kappa Research LLC, USA 2014.