2006-431: HANDS-ON EXPERIMENTAL ERROR! IMPROVING STUDENTS' UNDERSTANDING OF ERROR ANALYSIS

Stephanie Farrell, Rowan University

STEPHANIE FARRELL is Associate Professor of Chemical Engineering at Rowan University. She received her B.S. from the University of Pennsylvania, her MS from Stevens Institute of Technology, and her Ph.D. from New Jersey Institute of Technology. Prior to joining Rowan in September, 1998, she was a faculty member in Chemical Engineering at Louisiana Tech University. Stephanie's current educational research focuses on the role of hands-on experiments in inductive learning.

Hands-on Experimental Error!

Improving Students' Understanding of Error Analysis

Introduction

An understanding of error analysis is crucial for the scientist or engineer who must estimate uncertainties in experimental measurements and reduce them when necessary. Error analysis is a vital part of any experiment; without appropriate error analysis, meaningful conclusions cannot be drawn from the data. Unfortunately, as pointed out by Taylor [1], error analysis is often introduced through handouts containing formulas which students are simply told to use in their laboratory reports. Students fail to grasp the underlying concepts and rationale and to develop the insight which makes error analysis a truly interesting and important part of the laboratory experience.

The motivation for the development of this workshop was a perceived need to improve lower level engineering students' grasp of basic concepts of error analysis. While students at Rowan University were previously introduced to these basic concepts in introductory science courses (and to a limited extend, during lecture periods in the introductory engineering course), error analysis was often neglected entirely in engineering reports unless the details of the requirement were explicitly listed in the assignment. If present, error analysis was often inaccurate and meaningless– a cursory sentence or rote calculation included at the end of the report. For example, human error was frequently cited as a source of error in experimental procedure – with the implication that this is acceptable, legitimate, or unavoidable. In the laboratory, students failed to use techniques to reduce experimental error when necessary. Data were often not reported correctly to reflect uncertainty in measurement, and simple statistical techniques were rarely used to analyze error.

A variety of methods for the introduction of error analysis to lower level engineering students have been described by other educators. Sterrett and Helgeson[2] used parametric computer simulations to introduce error analysis to sophomores in a design course. Reardon[3] introduces linear regression and propagation of error analysis through a hands-on design project in a freshman engineering course. Rubino[4] describes a project-based freshman Engineering Technology course in which one module which introduces students to gross, systematic, and random error via hands-on measurements. The workshop described in this paper comprises a series of hands-on activities in which students conduct a variety of measurements and calculations in a familiar context, allowing experimental error and error analysis to become the primary focus of the investigation without being obscured by new theoretical subject content or extensive report writing.

This workshop was performed during a three-hour laboratory period at the beginning of the semester, prior to conducting any laboratory experiments which introduced new engineering concepts. A dramatic improvement was observed in the treatment of experimental error and analysis in the students' laboratory reports, and this was maintained throughout the semester. The hands-on nature of the workshop and the use of a familiar context in which to present new concepts are thought to be key elements in the success of this project.

After completing the workshop, students should have a solid grasp of the basic concepts of uncertainty in experimental measurements, how to reduce uncertainty in experimental measurements, and the proper representation and analysis of experimental data and uncertainty. These concepts were introduced in a familiar context via measurements and calculations associated with height, body temperature, weight, and pulse rate. These concepts are listed in Table 1 and are described in more detail throughout this paper.

Table 1. Concepts taught in this hands-on workshop

Uncertainty in experimental measurement
Systematic error
Accuracy
Precision
Least count and instrument limit of error
Estimation of uncertainty in repeatable measurements
How to reduce uncertainty in experimental measurements
Reporting and Using Uncertainty
Significant Figures
Accuracy
Precision
Comparison of measured numbers
Fractional uncertainties
Propagation of error
Analysis of Error
Mean average
Standard deviation, standard deviation of the mean, and confidence levels
Rejection of data
Chauvenet's criterion

Introduction to Uncertainty Analysis

The workshop was preceded by a one-hour introcuction to basic concepts related to experimental measurements and error analysis (see Table 1). In addition, students were given a short "pre-lab" assignment which was completed prior to the start of the workshop.

The statistical methods introduced in this workshop are based on the assumption of a normal distribution. Students are briefly introduced to the notion of distributions and the normal distribution as the limiting distribution of results for a measurement subject to many small, random errors. The fact that some of their measurements are not expected to follow a normal distribution is pointed out (for example, the number of heart beats would follow a binomial distribution). Students are reminded that the assumption of a normal distribution is an engineering approximation applied in this workshop. Further treatment of distributions is beyond the scope of this workshop.

For the purposes of this workshop, the definitions and equations described below were employed.

For N measurements of the quantity x, The best estimate for the value of x is taken to be the mean $\bar{x}$, given by

$$\bar{x} = \frac{\sum x_i}{N}$$

The sample standard deviation σ_x is used to represent the average uncertainty of N measurements, or how much the i^{th} measurement differs from he average $\bar{x}$:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum(x_i - \bar{x})^2}$$

The sample standard deviation represents a 68% probability that a single measurement is within σ_x of the correct value, assuming a normal distribution.

The standard deviation of the mean (SDOM) is the uncertainty associated with the mean as the best estimate of the true value of *x*.

$$SDOM = \frac{\sigma_x}{\sqrt{N}}$$

While the standard deviation is not expected to be affected appreciably by increasing the number of measurements, the standard deviation of the mean does slowly decrease with an increasing number of measurements. Therefore, making repeated measurements has the potential to increase precision.

The controversial subject of measurement rejection is explored during this lecture. Chauvenet's Criterion is introduced as a simple test that can be applied to determine whether data can be legitimately rejected. The basis of this test is the determination of the probability oft a legitimate measurement being as "deviant" as the suspect measurement. When this probability is multiplied by N, the number of measurements, the result is the fraction of legitimate measurements that are expected to be at least as deviant as the suspect measurement. Chauvenet's Criterion states that if this fraction is less than ½, the suspect measurement can be rejected. The choice of ½ is arbitrary, but it is one of several accepted quantitative measures of the "reasonableness" of a result.

Workshop Activities

Height

Each team of two students was provided three rulers: a short ruler (15 cm) with 1 mm divisions, a tape measure with 1 mm divisions, and (unbeknownst to students) a "trick" ruler (25 cm with 1 mm divisions) produced on a copy machine which enlarged the scale by 5% to introduce systematic error. Students noted the least count of each ruler, and then estimated the instrument limit of error (ILE). Since the divisions on the rulers were close together, most students estimated the ILE to be +/- one half of one division. Each student made repeated height measurements for his/her partner by first marking the height on the wall with removable tape, and then measuring the distance between the floor and the mark. Students collected a set of three measurements and a set of ten measurements using each ruler.

For each set of measurements, students prepared a table of their data showing each measurement and the instrument and experimental uncertainties (Table 2). Since the smaller rulers were used several times to obtain a single height measurement, the concept of error propagation was introduced. The propagated uncertainty based on ILE was calculated by direct addition because the uncertainties are not independent. For each data set students calculated the mean, standard deviation, and the average deviation of the mean.

Students observed that while the mean and standard deviation do not change significantly according to the number of measurements made, the SDOM gradually decreases with an increasing number of samples. Therefore the uncertainty be reduced by repeating measurements, but there is a diminishing return and good judgment must be exercised to determine the extent to which a measurement should be repeated. The measurement uncertainty associated with the ILE was compared with the standard deviation. In the case of the short ruler the propagated instrument uncertainty is appreciable, and a more appropriate choice of measuring device would increase the precision of the measurement. When students discovered that the "trick" ruler introduced systematic error, they realized the importance of instrument calibration. Students also observed that the systematic error is not reduced by the number of measurements taken.

Table 2. Measurement of height using three different rulers. Measurements were performed three and ten times, denoted by (3) and (10) respectively.

Ruler	Estimated Instrument Uncertainty (cm)	Uncertainty (propagated)	Number of Measurements	Mean Height (cm)	Standard Deviation (cm)	SDOM (cm)
15 cm	0.05	0.6	3 10	169.77 169.78	0.38 0.36	0.22 0.11
Tape	0.05	0.05	3 10	169.70 169.67	0.05 0.05	0.03 0.02
Trick	0.05	0.35	3 10	161.4 161.4	0.23 0.25	0.13 0.08

Body Temperature

Students were provided with five different models of fever thermometers purchased from a local pharmacy. An adequate supply of plastic hygiene covers was also available. Students obtained three measurements of body temperature using each thermometer. Measurments were taken orally following procedures described in the instrument documentation, with the thermometer in the same location in the mouth each time.

Students noted the least count for each thermometer, and estimated the instrument limit of error. It was assumed that the number of significant figures in digital readings was the same as the number of figures displayed, and that the uncertainty in the final digit was +/-1. The conventional alcohol thermometer had divisions of 0.2°F and the estimated ILE was generally accepted as +/- 0.1°F. Students calculated the mean, standard deviation, and the standard deviation of the mean for each set of data.

As shown in Table 3 there is significant variation in the mean temperature values obtained from the different digital thermometers. Students compare results obtained from the different thermometers and are asked to make conclusions regarding the precision of the measurements, the published accuracy, and their measurement technique. One observation is that the mean temperatures obtained with the different thermometers are not within 0.2 °F, indicating that there is either a flaw in the measurement technique or that the accuracy is not as stated. The conventional thermometer provides the most repeatable results, with three consecutive measurements yielding the same temperature of 97.9°F +/- 0.1°F.

Table 3. Measurement of body temperature using five fever thermometers. The pubslihed accuracy for each thermometer was 0.2°F in the range measured, and the ILE was 0.1 °F for each instrument.

Thermometer	Mean Temperature (°F)	Standard Deviation (°F)	SDOM (°F)
CVS Digital	98.0	0.26	0.15
Vick's Digital	98.4	0.15	0.09
CVS Quick Read, Regular Mode	97.8	0.12	0.07
CVS Quick Read, Quick Mode	98.17	0.23	0.13
Conventional alcohol thermometer	97.9	0	0

Heart rate

Students were provided with a digital stop watch with a digital display showing 0.01 second divisions. Prior to the start of the experiment, students were asked to decide between two viable methods for measuring pulse rate and time: the first method involves counting the beats in a specific time interval; the second entails measuring the time required for a specified number of heart beats. This is illustrated by trials using both methods with a metronome (http://www.metronomeonline.com/). Students quickly become aware of the difficulty inherent in timing an exact time and counting the number of events during that time period. They choose the latter method as their preferred technique. They typically estimate the uncertainty in their count to be +/- 0.5 beats. In addition, students practice their timing technique and estimate the uncertainty of timing due to their response time using the stop watch. This is done using a stopwatch to time ten second intervals displayed on an online timing device. With practice, students could typically measure within +/- 0.15 s of the "true" time. An insightful student recognizes that the timing device can give a misleading impression of accuracy -- limitations in timing technique introduce a significant uncertainty that exceeds both the published 0.01% accuracy of the device and the ILE of +/- 0.01 s.

Students obtain the pulse rate first by measuring the time for 20 heart beats (method 1), and then for 80 heart beats (method 2). Each measurement is repeated three times. The pulse rate (beats per minute) can then be calculated easily by dividing the number of beats by the time and converting seconds to minutes. The data and experimental uncertainty are displayed in Table 4.

Students then determine the propagated instrument uncertainty for each set of data, considering uncertainties in both timing (+/- 0.15s, determined in the independent experiment) and counting (+/-0.5, estimated). The difference in the propagated uncertainty associated with the two measurement techniques is quite significant (3.5% for method 1 vs. 0.9% for method 2). As expected, the standard deviation and SDOM for method 1 weres also greater than those for method two: according to Table 4 reveals that method 1 yields a result of 76.96 +/- 1.1 beats per minute. The more precise method 2 yields a result of 76.03 +/- 0.2 beats per minute.

Students realize that the uncertainty is greatly reduced by measuring the time for a larger number of heart beats, but that this technique presents the obvious disadvantage that more time is required to perform the measurement. When asked which technique is better, students quickly realize that the answer to this depends on the level of certainty required – a general consideration to bear in mind with the choice of measurement technique for any experiment.

Table 4. Determination of heart rate

Number of Beats	Mean Time (s)	Heart Rate (1/s)	Propagated Uncertainty (%)	Standard Deviation (1/s)	Standard Deviation (%)	SDOM (1/s)	SDOM(%)
20	15.61	76.95	3.5	2.7	3.5	1.1	1.4
80	63.13	76.03	0.9	0.4	0.5	0.2	0.2

The measurement of heart rate provides a good opportunity for discussion of distributions. The number of heart beats (a count) is expected to follow a binomial distribution, but would be well-represented by a normal distribution as the limiting case for a large number of samples. Timing measurements have a single-sided error in which the tail of the distribution would appear in the positive direction only (since stopping the watch is triggered by an event which must occur first, the error is due to response time and will always be positive.)

Weight

As part of their take-home assignment, students measure their weight on two scales in the University's Recreation Center. Each scale is a 350 lb Detecto balance-beam style scale. (Data are provided for students to use as an alternative if they are not comfortable sharing their own weight with classmates or the instructor. Students are still requested to measure and record their own weight, in order to receive the educational benefit of the hands-on activity; we would have no way to enforce this, but so far no student has opted to use the alternative data.)

Table 5. Measurement of body weight using two different scales.

Scale	Instrument Uncertainty (lb)	Mean Mass (lb)	Standard Deviation (lb)	SDOM (lb)
Scale A	2 oz	149.5	0.07	0.04
Scale B	2 oz	153.7	0.07	0.04

After noting the least count of each scale (4 oz.) and estimating the instrument limit of error (+/-2 oz.), students make ten successive weight measurements on each scale. They calculate the mean, standard deviation, and standard deviation of the mean of each data set as shown in Table 5. Comparison of the results obtained from the two different scales reveals a significant difference between the two average masses, although the repeatability is good as indicated by the low values of standard deviation and SDOM. One innovative team performed a calibration check on each scale using three 50 lb weights from the gymnasium, and determined that one of the scales was introducing a systematic error of about 2.8%. The team proposed a calibration correction to account for the systematic error associated with the less accurate scale.

Assignment

Finally, students are asked to select appropriate devices from a scientific catalog for each type of measurement made during the experiment. They must consider appropriate accuracy and repeatability as well as cost. They are asked to present three feasible choices for each instrument, and comment on how the price varies with accuracy and precision, and then to identify the most appropriate device for their application.

Assessment

Prior to the development of this workshop, students were introduced basic concepts of experimental error and error analysis in introductory science courses. This topic was also covered in the introductory engineering course at our university, in a lecture format with many in-class examples. In the author's experience, students could perform calculations that were specifically requested, but lacked the insight required to offer meaningful analysis in a laboratory report. Error analysis was often neglected entirely in engineering reports unless specifically required in the assignment. If present, error analysis was often inaccurate and meaningless– a cursory sentence or rote calculation included at the end of the report. Students lacked the ability to identify uncertainty in experimental measurements, to reduce experimental error using proper measurement techniques, to represent data properly, and to analyze experimental error in a meaningful way. The purpose of this workshop was for students to develop a real understanding of these concepts and to use them in future laboratory work and data reporting.

Student performance in the workshop was encouraging: 90% of the students demonstrated the ability to suggest appropriate techniques to reduce uncertainty in their measurements, identify the uncertainty in experimental measurements, represent data appropriately, and provide a meaningful analysis of uncertainty in their experiments. After the workshop, students performed eight additional experiments throughout the semester. The reporting assignments did not include a detailed description of error analysis required in the reports; students were simply reminded to include appropriate error analysis and discussion of experimental uncertainty. Over three years, 87.5% of the laboratory reports submitted contained a reasonable analysis and discussion of experimental uncertainty. A "reasonable analysis" included mention of instrument accuracy, ILE, standard deviation and SDOM of measured values, propagation of error, and rejection of data where appropriate; it did not include assertions such as "human error" accounting for discrepancies in results. Three-quarters of the teams also included a discussion of experimental uncertainty in their final oral presentations.

Conclusions

"Tell me and I forget. Show me and I may remember. Involve me and I understand." This quote has been attributed to scientist and statesman Benjamin Franklin and explains the improved effectiveness of this workshop over prior attempts to teach the same concepts using more traditional teaching techniques . Its hands-on nature as well as the introduction of new concepts in a familiar context are thought to be key elements in the success of this workshop.

Acknowledgments

Funding for this project was obtained from the National Science Foundation Course, Curriculum, and Laboratory Improvement Program (NSF DUE #0088437).

[1] Taylor, John R., An Introduction to Error Analysis, 2nd edition, University Science Books, Sausalito, CA, p. xv.

[2] Sterrett, J.D. and R.J. Helgeson, Proceedings of the 1999 Conference of the American Society of Engineering Education, Session 3225, 1999.

[3] Reardon, K., Proceedings of the 2001 Proceedings of the 1999 Conference of the American Society of Engineering Education, Session 3553, 2001.

[4] Rubino, F.Jl, Proceedings of the 1998 Conference of the American Society of Engineering Education, Session 3547, 1998.