Chapter 2

Understanding Errors and Uncertainties in the Physics Laboratory

2.1 Introduction: Measurements, Observations, and Progress in Physics

Physics, like all natural sciences, is a discipline driven by observation. The concepts and methodologies that you learn about in your lectures are not taught because they were first envisioned by famous people, but because they have been observed always to describe the world. For these claims to withstand the test of time (and repeated testing in future scientific work), we must have some idea of how well theory agrees with experiment, or how well measurements agree with each other. Models and theories can be invalidated by conflicting data; making the decision of whether or not to do so requires understanding how strongly data and theory agree or disagree. Measurement, observation, and data analysis are key components of physics, equal with theory and conceptualization.

Despite this intimate relationship, the skills and tools for quantifying the quality of observations are distinct from those used in studying the theoretical concepts. This brief introduction to errors and uncertainty represents a summary of key introductory ideas for understanding the quality of measurement. Of course, a deeper study of statistics would enable a more quantitative background, but the outline here represents what everyone who has studied physics at the introductory level should know.

Based on this overview of uncertainty, you will perhaps better appreciate how we have come to trust scientific measurement and analysis above other forms of knowledge acquisition, **precisely because we can quantify what we know and how well we know it**. Scientific ideas themselves are not sacrosanct; contrarily, they can be replaced with improvements and refinements to understanding. This progress is measurable through analysis of uncertainty in data, measurement, and calculation. Scientific progress is based on skepticism and continued critical analysis; uncertainties guide us in our inquiries.

2.2 Some References

The study of errors and uncertainties is part of the academic field of statistics. The discussion here is only an introduction to the full subject. Some classic references on the subject of error analysis in physics are:

- Philip R. Bevington and D. Keith Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, 1992.
- John R. Taylor, *An Introduction to Error Analysis; The Study of Uncertainties in Physical Measurements*, University Science Books, 1982.
- Glen Cowan, *Statistical Data Analysis*, Oxford Science Publications, 1998

2.3 The Nature of Error and Uncertainty

Error is the difference between an observation and the true value.

$$\text{Error} = \text{observed value - true value}$$

The "observation" can be a direct measurement or it can be the result of a calculation that uses measurements; in which case, the "true" value might also be a calculated result. Note that according to this definition, an observation has an error even if we do not know what the true value is. (Estimating this 'true value' is often the point of doing the experiment!) Therefore, we will actually be analyzing the *uncertainty*, the estimate of the *expected error* in an observation.

Example: Someone asks you, what is the temperature? You look at the thermometer and see that it is 71° F. But, perhaps, the thermometer is mis-calibrated and the actual temperature is 72° F. There is an *error* of -1° F, but you do not know this. What you *can* figure out is the reliability of measuring using your thermometer, giving you the *uncertainty* of your observation. Perhaps this is not too important for casual conversation about the temperature, but knowing this uncertainty would make all the difference in deciding if you need to install a more accurate thermometer for tracking the weather at an airport or for repeating a chemical reaction exactly during large-scale manufacturing.

2.3.1 Sources of Error

No real physical measurement is exactly the same every time it is performed. The uncertainty tells us how closely a second measurement is expected to agree with the first. Errors can arise in several ways, and the uncertainty should help us quantify these errors. In a way the uncertainty provides a convenient 'yardstick' we may use to estimate the error.

- **Systematic error:** Reproducible deviation of an observation that biases the results, arising from procedures, instruments, or ignorance. Each systematic error biases every measurement in the same direction.

- **Random error:** Uncontrollable differences from one trial to another due environment, equipment, or other issues that reduce the repeatability of an observation. They may not actually be random, but deterministic (if you had perfect information): dust, electrical surge, temperature fluctuations, etc. In an ideal experiment, random errors are minimized for precise results. Random errors are sometimes positive and sometimes negative; they are sometimes large but are more often small. In a sufficiently large sample of the measurement population, random errors will average out.

Random errors can be estimated from statistical repetition and systematic errors can be estimated from understanding the techniques and instrumentation used in an observation.

Other contributors to uncertainty are *not* classified as 'experimental error' in the same scientific sense, but still represent difference between measured and 'true' values. The challenges of estimating these uncertainties are somewhat different.

- **Mistake, or 'illegitimate errors':** This is an error introduced when an experimenter does something wrong (measures at the wrong time, notes the wrong value). These should be prevented, identified, and corrected, if possible, and ideally they should be completely eliminated. Lab notebooks can help track down mistakes or find procedures causing mistakes.

- **Fluctuations:** Sometimes, the variability in a measurement from its average is not a random error in the same sense as above, but a physical process. Fluctuations can contain information about underlying processes such as thermal dynamics. In quantum mechanics these fluctuations can be real and fundamental. They can be treated using similar statistical methods as random error, but there is not always the desire or the capacity to minimize them. When a quantity fluctuates due to underlying physical processes, perhaps it is best to redefine the quantity that you want to measure. (For example, suppose you tried to measure the energy of a single molecule in air. Due to collisions this number fluctuates all over the place, even if you could identify a means to measure it. So, we invent a new concept, the *temperature*, which is related to the average energy of molecules in a gas. Temperature is something that we can measure, and assign meaningful uncertainties to. Because of physical fluctuations caused by molecular collisions, temperature is a more useful quantity in most cases.)

2.3.2 Accuracy vs. Precision

- **Accuracy:** Accuracy is how closely a measurement comes to the 'true' value. It describes how well we eliminate systematic error and mistakes.

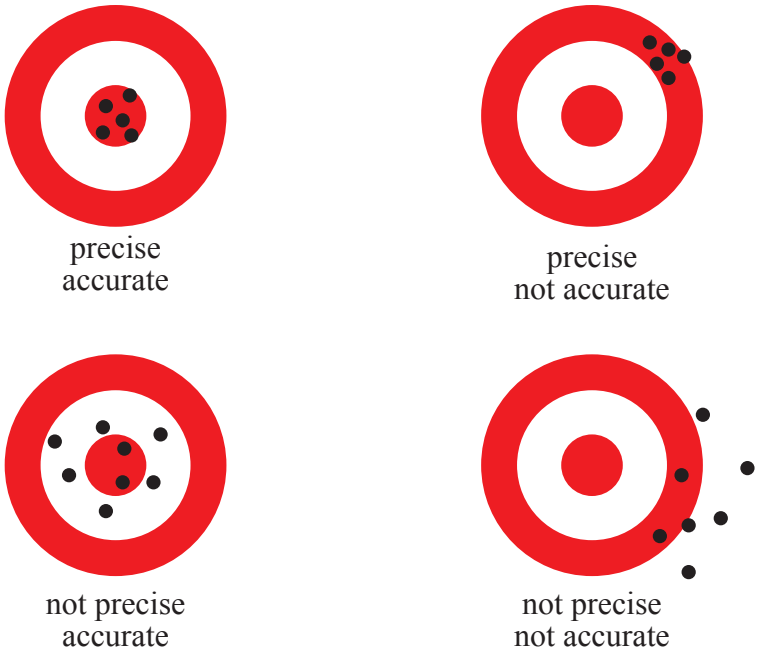

Figure 2.1: *Several independent trials of shooting at a bullseye target illustrate the difference between being accurate and being precise.*

- **Precision:** Precision is how exact a result is determined without referring to the 'true' value. It describes how well we suppress random errors and thus how well a sequence of measurements of the same physical quantity agree with each other.

It is possible to acquire two precise, but inaccurate, measurements using different instruments that do not agree with each other at all. Or, you can have two accurate, but imprecise, measurements that are very different numerically from each other, but statistically cannot be distinguished.

2.3.3 Notation of Uncertainties

There are several ways to write various numbers and uncertainties.

- **Absolute Uncertainty:** The magnitude of the uncertainty of a number in the same units as the result. We use the symbol δx for the uncertainty in x, and express the result as $x \pm \delta x$.
  Example: For $\delta x = 6$ cm and a length L of 2 meters, we would write $L = 2.00 \pm 0.06\,\text{m}$.
- **Relative Uncertainty:** The uncertainty as a fraction of the number.
  Example:

$$\frac{\delta x}{x} = \frac{0.06}{2.00} = 0.03$$

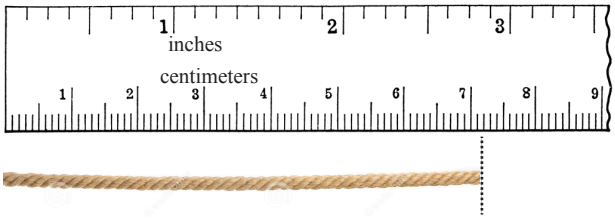

Figure 2.2: *Measuring a string with a ruler. A reasonable measurements from this might be reported as* 7.15 ± 0.05 *cm.*

We must be very careful to say when our uncertainty is relative

$$x = 2.00\,\text{m} \pm 0.03(\text{rel}) = 2.00\,\text{m} \pm 3\% \tag{2.1}$$

- **Percent Difference:** The difference between measurements can sometimes be represented as a percentage difference. The percentage difference between two values x_1 and x_2 can be written as the difference divided by the average times 100% or:

$$\text{Percentage difference} = 200\% \times \frac{x_1 - x_2}{x_1 + x_2} \tag{2.2}$$

2.3.4 Estimating Uncertainties

2.3.5 Level of Uncertainty

How do you actually estimate an uncertainty? First, you must settle on what the quantity δx actually means. If a value is given as $x \pm \delta x$, what does the range $\pm\delta x$ mean? This is called the *level of confidence* of the results.

Note that the true value of a measurement does *not* need to fall in the range of the uncertainty. This is a common misconception! In fact, for a 68% confidence interval, the actual value does not fall in the range 32% of the time. In results of final measurements, it is common to use 68%, 95%, and 99% confidence levels for reporting data, but unfortunately, most practicing scientists are not always clear what they mean with an uncertainty. **In this course, we will follow the convention of 68% confidence intervals, so that δx represents one standard deviation of the data set.** Section 2.3.8 discusses the standard deviation of a data set.

2.3.6 Reading Instrumentation

Measurement accuracy is limited by the tools used to measure. In a car, for example, the speed divisions on a speedometer may be only every 5 mph, or the digital readout of the odometer may only read down to tenths of a mile. To estimate instrumentation accuracy, assume that the uncertainty is one half of the smallest division that can be unambiguously read from the device. **Instrumentation accuracy must be recorded during laboratory**

measurements. In many cases, instrument manufacturers publish specification sheets that detail their instrument's errors more thoroughly. In the absence of malfunction, these specifications are reliable; however, 'one half of the smallest division' might not be very reliable if the instrument has not been calibrated recently.

2.3.7 Experimental precision

Even on perfect instruments, if you measure the same quantity several times, you would obtain several different results. For example, if you measure the length of your bed with a ruler several times, you might find a slightly different number each time. The bed and/or the ruler could have expanded or contracted due to a change in temperature or a slightly different amount of tension. These unavoidable uncertainties are always present to some degree in an observation. Even if you understand their origin, the randomness cannot always be controlled. We can use statistical methods to quantify and to understand these random uncertainties.

2.3.8 Quantifying Uncertainties

Here we note some mathematical considerations of dealing with random data. These results follow from the central limit theorem in statistics and analysis of the normal distribution. This analysis is beyond the scope of this course; however, the distillation of these studies are the point of this discussion.

2.3.9 Mean, Standard Deviation, and Standard Error

The mean: Suppose we collect a set of measurements of the *same* quantity x, and we label them by an integer index i: $\{x_i\} = (x_1, x_2, \ldots x_N)$. What value do we report from this set of identical measurements? We want the **mean**, μ, of the population from which such a data set was randomly drawn. We can approximate μ with the *sample mean* or average of this particular set of N data points:

$$\bar{x} = \frac{1}{N}\sum_i x_i \tag{2.3}$$

Of course, this is not the true mean of the population, because we only measured a small subset of the total population. But it is our best guess and, statistically, it is an *unbiased predictor* of the true mean.

The standard deviation: How precisely do we know the value of x? To answer this question of statistical uncertainty based on the data set $\{x_i\}$, we consider the squared deviations from the sample mean $\bar{x}$. The **sample variance** s_x^2 is the sum of the squared deviations divided by the 'degrees of freedom' (DOF). For N measurements the DOF for variance is $N - 1$. (The origin of the $N - 1$ is a subtle point in statistics. Ask if you are

interested.) The **sample standard deviation**, s_x, is the square root of the sample variance of the measurements of x.

$$s_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\text{DOF}}} \tag{2.4}$$

The sample standard deviation is our best 'unbiased estimate' of the true statistical standard deviation σ_x of the population from which the measurements were randomly drawn; thus it is what we use for a 68% confidence interval for *one* measurement.

The standard error: If we do not care about the standard deviation of a measurement but, rather, how well we can rely on a calculated average value, $\bar{x}$, then we should use the **standard error** or standard deviation of the mean $s_{\bar{x}}$. This is found by dividing the sample standard deviation by $\sqrt{N}$:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}} \tag{2.5}$$

2.3.10 Reporting Data

Under normal circumstances, the best estimate of a measured value x predicted from a set of measurements $\{x_i\}$ is given by $x = \bar{x} \pm s_{\bar{x}}$. If we perform N more measurements of x, average them, and find their standard error, these two averages will agree within their standard errors 68% of the time.

2.3.11 Error Propagation

One of the more important rules to remember is that the measurements we make have a range of uncertainty so that any calculations using those measurements also must have a commensurate range of uncertainty. After all the result of the calculation will be different for each number we should choose within range of our measurement.

We need to learn how to propagate uncertainty through a calculation that depends on several uncertain quantities. Final results of a calculation clearly depend on these uncertainties, and it is here where we begin to understand how. Suppose that you have two quantities x and y, each with an uncertainty δx and δy, respectively. What is the uncertainty of the quantity $x + y$ or xy? Practically, this is very common in analyzing experiments and statistical analysis provides the answers disclosed below.

For this course we will operate with a set of rules for uncertainty propagation. It is best not to round off uncertainties until the final result to prevent accumulation of rounding errors. Let x and y be measurements with uncertainty δx and δy and let c be a number with negligible uncertainty. We assume that the errors in x and y are *uncorrelated*; when one value has an error, it is no more likely that the other value's error has any particular value or trend. We use our measurements as described below to calculate z and the propagated uncertainty in this result.

- **Multiplication by an exact number:** If $z = c\,x$, then

$$\delta z = c\,\delta x \tag{2.6}$$

- **Addition or subtraction by an exact number:** If $z = c + x$, then

$$\delta z = \delta x \tag{2.7}$$

- **Addition or subtraction:** If $z = x \pm y$, then

$$\delta z = \sqrt{(\delta x)^2 + (\delta y)^2} \tag{2.8}$$

- **Multiplication or division:** If $z = xy$ or $z = \frac{x}{y}$, then

$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2} \tag{2.9}$$

- **Power:** If $z = x^c$, then

$$\frac{\delta z}{z} = c\,\frac{\delta x}{x} \tag{2.10}$$

The important pattern in these rules is that when you combine multiple uncertainties, you do not add them directly, but rather you square them, add, and then take the square root. The reason for this is intuitive: if one error is randomly positive, the other one is sometimes negative, which *reduces* the total error. Therefore, it is incorrect to estimate the combination of two uncertainties as their sum since this overestimates the average size of the *combined* error.

As one example consider $z = \frac{x^3}{y}$. Before we can use the division formula, Equation (2.9), we must use the power formula to get the error in the numerator. Let $w = x^3$ and use Equation (2.10) to find that $\delta w = 3w\,\frac{\delta x}{x} = 3x^2\delta x$. Next, Equation (2.9) gives our result

$$\delta z = z\sqrt{\left(\frac{\delta y}{y}\right)^2 + \left(\frac{\delta w}{w}\right)^2} = z\sqrt{\left(\frac{\delta y}{y}\right)^2 + \left(\frac{3\,\delta x}{x}\right)^2}.$$

The exponent makes the numerator three times more sensitive to uncertainty than the denominator.

2.3.12 Optional Reading – Advanced Topic: The General Formula for Error Propagation

There is a general calculus formula encapsulating the rules for uncertainty propagation for a general function of several variables $\big(A = f(x, y, z)\big)$. Derivation of this is beyond the scope

of this course. If you continue in your study of science, you will become more familiar with equations like this.

$$\delta A = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 (\delta x)^2 + \left(\frac{\partial f}{\partial y}\right)^2 (\delta y)^2 + \left(\frac{\partial f}{\partial z}\right)^2 (\delta z)^2} \tag{2.11}$$

2.3.13 Significant Figures

The *significant figures* of a number are the digits in its representation that contribute to the precision of the number. In practice, we assume that all digits used to write a number are significant. Therefore, completely uncertain digits should not be used in writing a number and results should be rounded to the appropriate significant figure. For example, you should not express your height as 70.056 inches if your uncertainty is ± 0.1 inch. It would more appropriately be written as 70.1 inches. Uncertainties specified using only significant digits are always ± 5 times a power of 10; the least significant displayed digit was the result of rounding up or down by as much as 0.5 of that digit. Usually we know our uncertainty to be something close to this but yet different. Further, results of simple calculations should not increase the number of significant digits. Calculations transform our knowledge; they do not increase our knowledge. The rounding should be performed at the final step of a calculation to prevent rounding errors at intermediate steps from propagating through your work but one or two extra digits suffice to prevent this.

Zeros are also considered significant figures. If you write a number as 1,200, we assume there are four significant digits. If you only mean to have two or three, then it is best to use scientific notation: 1.2×10^3 or 1.20×10^3. Leading zeros are not considered significant: 0.55 has just two significant figures.

There are some guidelines for tracking significant figures throughout mathematical manipulation. This is useful as a general method to keep track of the precision of a number so as not to carry around extra digits of information, but you should generally be using more formal error estimates from Sections 2.3.4 and 2.3.8 for reporting numbers and calculations in the physics lab.

- **Addition and Subtraction:** The result is known to the decimal place of the *least* precise input number.

  Example: $45.37 + 10 = 55$, not 55.37 or 55.4
  Why? $\delta = \sqrt{0.005^2 + 0.5^2} = 0.5$

  Where we used the sum formula Equation (2.7).

- **Multiplication and Division:** The result is known to as many significant figures as are in the least precise input number.

  Example: $45.4 \times 0.25 = 11$, not 11.4
  Why? $\delta = 11\sqrt{\left(\frac{0.05}{45}\right)^2 + \left(\frac{0.005}{0.25}\right)^2} = 0.2 > 0.05$

Where we used the product formula Equation (2.9).

Example: If you measure a value on a two-digit digital meter to be 1.0 and another value to be 3.0, it is *incorrect* to say that the ratio of these measurements is 0.3333333, even if that is what your calculator screen shows you. The two values are measurements; they are not exact numbers with infinite precision. Since they each have two significant digits, the correct number to write down is 0.33.

For this lab, you should use proper significant figures for all reported numbers. We will generally follow a rule for significant figures in reported numbers: **calculate your uncertainty to two significant figures, if possible, using the approach in Sections 2.3.4 and 2.3.8, and then use the same level of precision in the reported error and measurement.** This is a rough guideline, and there are times when it is more appropriate to report more or fewer digits in the uncertainty. However, it is always true that the result must be rounded to the same significant figures as the uncertainty. The uncertainty tells us how well we know our measurement.

2.4 How to Plot Data in the Lab

Plotting data correctly in physics lab is somewhat more involved than just drawing points on graph paper. First, you must choose appropriate axes and scales. The axes must be scaled so that the data points are spread out from one side of the page to the other. **Axes must always be labeled** with physical quantity plotted and the data's units. Then, plot your data points on the graph. Importantly, **you must add your error bars to your data points**. Often, we only draw error bars in the vertical direction since these tend to dominate, but there are cases where it is appropriate to have both horizontal and vertical error bars. **In this course, we use one standard deviation (standard error if appropriate for the data point) for the error bar**. This means that 68% of the time the 'true' value should fall within the error bar range.

Do *not* connect your data points by lines. Rather, fit your data points to a model (often a line), and then add the best-fit model curve to the figure. The line, representing your theoretical model, is the best fit to the data collected in the experiment. Because the error bars represent just one standard deviation, it is fairly common for a data point to fall more than an error bar away from the fit line. *This is OK!* Your error bars are probably too large if the line goes through all of them!

The fitting parameters are usually important to our experiment as measured values. These measured parameters and other observations help us determine whether the fitting model agrees or disagrees with our data. If they agree, then some of the fitting parameters might yield measurements of physical constants.

More information can be found in Appendix B on page 155 and in Appendix C on page 157.

2.4.1 Fitting Data

Fully understanding this section is not required for Physics 136. You will use least-squares fitting in the laboratory, but we will not discuss the mathematical justifications of curve fitting data. Potential physics and science majors are suggested to internalize this material; it will become an increasingly important topic in upper division laboratory courses and research.

In experiments one must often test whether a theory describes a set of observations. This is a statistical question, and the uncertainties in data must be taken into account to compare theory and data correctly. In addition, the process of 'curve fitting' might provide estimates of parameters in the model and the uncertainty in these parameter estimations. These parameters tailor the model to your particular set of data and to the apparatus that produced the data.

Curve fitting is intimately tied with error analysis through statistics, although the mathematical basis for the procedure is beyond the scope of this introductory course. This final section outlines the concepts of curve fitting and determining the 'goodness of fit'. Understanding these concepts will provide deeper insight into experimental science and the testing of theoretical models. **We will use curve fitting in the lab, but a full derivation and statistical justification for the process will not be provided in this course**. The references in Sec. 2.2, Wikipedia, advanced lab courses, or statistics textbooks will all provide a more detailed explanation of data fitting.

2.4.2 Least-Squares and Chi-Squared Curve Fitting

For curve fitting, we make several measurements of a pair in a data set $\{x_i, y_i\}$, labelled by an integer index i. We want to understand how well a theoretical curve predicts y given x. Typically, we also want to extract some parameters $\{a_j\}$ in a model function f, such as $y = f(x, \{a_j\})$. The method to do this is based on the 'method of maximum likelihood', which we do not cover in detail here. In words, we want to find the set of theoretical parameters $\{a_j\}$ for a curve that maximizes the probability of obtaining the observed data set $\{y_i\}$ from $\{x_i\}$. This will be our best guess of the theory curve, and the result will be an estimate of $\{a_j\}$. Here, the index j is a different label than i: j labels the fit parameters $\{a_j\}$ and i labels the data points $\{x_i, y_i\}$.

Let's assume that there is no uncertainty in x, and we have N measurements of y, each with some uncertainty. We have a set of N values $y_{\text{exp},i}$. We also have N predictions from the theory curve $y_{\text{pred},i}$. These predictions may depend on unknown parameters $\{a_j\}$. For the least squares method, we ignore any error bars δy_i and choose the parameters $\{a_j\}$ to minimize the function

$$\sum_{i=1}^{N} (y_{\text{exp},i} - y_{\text{pred},i})^2 \tag{2.12}$$

This maximizes the probability of obtaining the measured data if the theory were correct.

Of course, it is possible that each measurement is not equally well known. Each data point could have its own uncertainty δy_i. In this case, we would minimize the variance with

each term weighted by how well we know it. This is called **Chi-squared fitting** and it is the standard data fitting method used in physics laboratories. We seek to minimize some function:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_{\text{exp},i} - y_{\text{pred},i})^2}{\delta y_i^2} \tag{2.13}$$

We allow the parameters $\{a_j\}$ to vary to give the smallest χ^2. This can be done in some cases exactly from the data, and sometimes it needs an iterative procedure to find the best fit solution. There are also procedures to extract the variance in the parameters $\{\delta a_j\}$ from the method. Typically, software packages are used for this analysis.

2.4.3 Optional Reading – Advanced Topic: Chi-Square 'Goodness of Fit'

Suppose that we have a curve with *given* parameters $\{a_j\}$. How well does it fit some experimental data? There is a statistical test to determine the *goodness of fit*. We calculate the χ^2 statistic as above, and then divide by the degrees of freedom. If the parameters are obtained from a fit, then the degrees of freedom is $N - M$ where M is the number of fit parameters. Basically, we use our N measurements to calculate the M parameters and the DOF tells us how much 'information' we have left in our data set for optimizing the fit.

Think of it this way: If we had N independent fit parameters (unknowns) and N exact data points (equations), we could in principle solve for all N unknowns. Instead we have only M fit parameters so the problem is over-specified. On the other hand our data is also uncertain so that we can use this over-specification to average out much of our uncertainties. This is quite similar to making a measurement several times and then to average them to get a better predictor for the 'true' value.

We calculate the **reduced** χ^2 by dividing by the degrees of freedom. This statistic is then compared to tables of numbers or using software to see how good the model fits the data. If the uncertainties are correct, and the data is truly distributed as expected, and the theory curve truly describes the underlying process, then the reduced χ^2 should be about 1. If it is larger, then there is too much variance and either the fit is bad or the uncertainties are underestimated. If it is significantly less than 1, then the uncertainties may be overestimated or the data might be insensitive to the variation predicted by the theory curve. The chi-square value should be near 1 for a properly fit theory curve correctly describing the physical process with reasonable uncertainty estimates.

2.4.4 Example: Linear Regression

The optimization procedure is most often performed for a fit to a linear function $y = mx + b$. Here, the two parameters that we wish to estimate are m and b, given a data set $\{x_i, y_i\}$

with N pairs. The best fit results can be derived using calculus:

$$m = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \tag{2.14}$$

$$b = \frac{\sum x_i^2 \sum_i y_i - \sum x_i \sum x_i y_i}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \tag{2.15}$$

Assuming the uncertainties in y dominate and are all equal to δy, the variances can be written:

$$\sigma_m^2 = \frac{N}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}(\delta y)^2 \tag{2.16}$$

$$\sigma_b^2 = \frac{\sum_i x_i^2}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}(\delta y)^2 \tag{2.17}$$

The uncertainties of the parameter estimators are found from the square root of the variances. The goodness of fit can be found from Equation (2.13). A proper statistical determination of parameter uncertainties from real data in linear and nonlinear regression is a subtle topic that is beyond the scope of this course.

2.4.5 Strategy for Testing a Model

Experiments are designed and performed for the purpose of testing whether we can rely upon a particular hypothesis. Suppose we wish to check whether $w = f(x, y, z)$ is a natural truth. Then we must arrange to measure $x \pm \delta x$, $y \pm \delta y$, and $z \pm \delta z$ so that we can *predict* the corresponding value of $w_\mathrm{p} = f(x, y, z)$. We can also utilize what we have learned in Section 2.3.11 to determine how the uncertainties δx, δy, and δz lead to the uncertainty in our prediction δw_p. Together, then we know that $w_\mathrm{p} \pm \delta w_\mathrm{p}$ if our model is correct.

We cannot quit there, however; just because the model made a prediction does not necessarily mean it is correct. We must also arrange to *measure* $w_\mathrm{m} \pm \delta w_\mathrm{m}$ while our apparatus is 'set' to x, y, and z. If the hypothesis is correct and our apparatus faithfully represents f, then this measurement must match f's prediction.

2.4.6 A Comparison of Measurements

If we make two measurements of (hopefully) the same quantity, both measurements contains the inherent uncertainty of the tool(s) used to make the measurements. Then how can we decide whether the two measurements agree? Statistics allow us to answer this question.

First, we form a *null hypothesis* by subtracting the two (hopefully equal) measurements

$$\Delta = |w_\mathrm{p} - w_\mathrm{m}| . \tag{2.18}$$

Even if these two numbers were taken from the same population, this difference has expected error, σ, due to δw_p and δw_m. We can find this error using our propagation formula for difference (Equation (2.8))

$$\sigma = \sqrt{(\delta w_\mathrm{p})^2 + (\delta w_\mathrm{m})^2}\,. \tag{2.19}$$

Since δw_p already contains δx, δy, and δz, σ contains all of our measurement uncertainties and we cannot reasonably expect $\Delta < \sigma$. Since our σ has 68% confidence level, we would actually expect $\Delta > \sigma$ to happen 32% of the time, $\Delta > 2\,\sigma$ to happen 4.6% of the time, and $\Delta > 3\,\sigma$ to happen 0.3% of the time. Larger disagreements are less and less likely.

Because of these statistics, we tend to consider when $\Delta < 2\,\sigma$ that the two numbers agree, $w_\mathrm{p} = w_\mathrm{m}$, and that any minor discrepancy is more likely to be due to randomness in the experiment and 'other sources of error'. After all, claiming inequality would make us wrong about once in twenty times; this is far too often for most scientists. On the other hand, 3 in 1000 times is not too bad; we especially like these odds if they persist as we try to reconcile the difference. Therefore we consider $\Delta > 3\,\sigma$ to indicate that the two numbers disagree.

A disagreement *could* mean that the data contradicts the theory being tested, but it could also just mean that one or more assumptions are not valid for the experiment; perhaps we should revisit these. Disagreement could mean that we have underestimated our errors (or even have overlooked some altogether); closer study of this possibility will be needed. Disagreement could just mean that this one time the improbable happened. These possibilities should specifically be mentioned in your Analysis according to which is most likely, but further investigation will await another publication.

One illegitimate source of disagreement that plagues students far too often is simple math mistakes. When your data doesn't agree and it isn't pretty obvious why, repeat your calculations to make sure you get the same answer twice.