2

Quick Start

Replicate Error Basics

Our first topic is random error, a subject intimately tied to statistics.

When we make an experimental determination of a quantity, one of the questions we ask about our result is, if someone else came along and did the same measurement on similar equipment, would they get the same value as we did? We would like to think that they would, but there are many slight, random differences between what is done in any two laboratories and in how two similar apparatuses perform, so we accept that a colleague's answer might be a little different from our answer. A non-laboratory example of this would be weighing oneself on the same kind of scale at home and at the gym—these two numbers might differ by a kilogram or two. Even if you weigh yourself repeatedly throughout the day on the same bathroom scale, you may see some variation due to what you have eaten recently, whether you have exercised, or if your clothing is a bit different for each measurement. Quantities that have this characteristic of variability are called *stochastic variables*.

To identify a good value for a measured variable that is subject to a variety of influences, we turn to statistics. If effects are random, statistics tells us the probability distribution of the effect happening (random statistics), and we can rigorously express both a best estimate for the quantity and error limits on the best estimate [5, 38]. The *mean* of replicated measurements expresses the expected value of the measured quantity, and the *variance* of replicated measurements can be used to quantify the effect of random events on the reproducibility of the measurements, allowing us to construct error limits. We discuss these topics now. Additional background on the statistics of stochastic variables may be found in the literature [5, 38].

2.1 Introduction

When we repeatedly determine a quantity from measurements of some sort, the measured numbers often vary a bit, preventing us from knowing that number with absolute precision. Consider, for example, the time it takes to go from your home to your workplace. You may know, roughly, that it takes 30 min, but that number changes a bit from day to day and may vary with what time of day you make the trip, with the type of weather encountered, and with traffic conditions.

To determine a good estimate of the time it takes to make the trip from your home to your workplace, you might measure it several times and take the average of your measurements. Repeated measurements of a stochastic variable are called *replicates*. From replicates we can calculate an average or mean; the mean of a set of replicates is a good estimate of the value of the variable. In replicate analysis we use the following terms:

$$x \quad \text{stochastic or random variable} \tag{2.1}$$

$$x_i \quad \text{an observation of } x \text{ (collectively, the sample set)} \tag{2.2}$$

$$n \quad \text{number of observations in a sample set (sample size)} \tag{2.3}$$

$$\bar{x} \quad \text{mean value of the observations of } x \tag{2.4}$$

$$s^2 \quad \text{variance of the observations of } x \tag{2.5}$$

$$s \quad \text{standard deviation of the observations of } x \tag{2.6}$$

We define these terms in the paragraphs that follow.

Repeated measurements of a quantity such as commuting time may be thought of as observations of a stochastic variable. When we identify a quantity as a stochastic variable, we are saying that the value is subject to influences that are random. These random influences serve both to increase the variable (heavy traffic due to a visiting dignitary slows you down and increases commuting time) and to decrease the variable (leaving earlier in the morning before there is too much traffic decreases your commuting time). Because the influences are random, they average out, leaving a mean value that stays constant throughout the random effects. The definition of the mean of n observations x_i of a stochastic variable x is

$$\text{Sample mean} \qquad \bar{x} \equiv \left(\frac{x_1 + x_2 + \cdots + x_n}{n} \right) \tag{2.7}$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.8}$$

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

This formula is the familiar arithmetic average. In the spreadsheet program Microsoft Excel,[1] the mean of a range of numbers is calculated with the function AVERAGE(range); all Excel functions mentioned in the text are listed for reference in Appendix C. In the MATLAB computing environment,[2] the mean is calculated with the built-in function mean(array); all MATLAB functions or commands mentioned in the text are listed for reference in Appendix D. Appendix D also contains a table comparing equivalent Excel and MATLAB commands.

The list of terms given earlier includes two quantities that assess the variability of replicates: the sample variance s^2 and the sample standard deviation s. The definitions of these are [52]:

$$\text{Sample variance} \quad s^2 \equiv \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \right) \tag{2.9}$$

$$\text{Sample standard deviation} \quad s = \sqrt{s^2} \tag{2.10}$$

Looking at the definition of variance in Equation 2.9, we see that it is a modified average of the squared differences between the individual measurements x_i and the sample mean $\bar{x}$. The use of *squared* differences ensures that both positive and negative deviations count as deviations and do not cancel out when the sum is taken. The sample variance is not quite the average of squared differences – the average of the squared differences would have n in the denominator instead of $(n-1)$ – but this difference is not significant for our purposes. The variance turns out to be a very useful measure of variability of stochastic quantities. The presence of $(n-1)$ in the denominator of the equation defining sample variance (Equation 2.9) rather than n is called for by statistical reasoning.[3] Sample variance and its square root, sample standard deviation, are widely used to express the variability or spread among observations x_i of stochastic variables.

In Excel, the variance of a sample set is calculated with the function VAR.S(range) and the standard deviation of a sample set with STDEV.S(range) or SQRT(VAR.S(range)); in MATLAB, these commands are var(array) and

[1] We used Microsoft Excel 2013, www.microsoft.com, Redmond, WA.
[2] We used MATLAB r2018a, www.mathworks.com/products/matlab.html, Natick, MA.
[3] In short, we use s^2 as an estimate of the population variance σ^2, but when we use n instead of $(n-1)$ in Equation 2.9, the expected value of s^2 is not σ^2 but rather the quantity $\left(\sigma^2 - \sigma^2/n\right)$. Knowing this, and with a little algebra, we arrive at the conclusion that, if we use the definition of sample variance in Equation 2.9, the expected value of sample variance s^2 will be σ^2 as we desire. This is called Bessel's correction; see [38, 52].

Replicated Variable, Y:						Units:		
Measured values $Y_1, Y_2, \ldots, Y_n$		Sample Mean, $\bar{Y}$	Sample Variance, s^2	Sample Standard Deviation, s	Standard Error, $e_s = \frac{s}{\sqrt{n}}$	95% Confidence Interval based on n replicates (*Student's t* distribution)		
Y_1						$n = 1$	n/a	(include units) $\pm$
Y_2						$n = 2$	$\pm 12.7e_s$	
Y_3						$n = 3$	$\pm 4.30e_s$	
Y_4						$n = 4$	$\pm 3.18e_s$	
Y_5						$n = 5$	$\pm 2.78e_s$	
Y_6						$n = 6$	$\pm 2.57e_3$	
Y_7						$n \geq 7$	$\pm 2e_s$	
						∞	$\pm 1.96e_s$	

$$\bar{Y} \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad s^2 \equiv \frac{1}{(n-1)}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Figure 2.1 When n data replicates are available for a measured variable, the average of the measurements is a good estimate for the value of the measured variable, and the sample variance s^2 and sample standard deviation s are associated with the variability of the measurements. As we see later in the chapter, these quantities allow us to determine standard replicate error $e_{s,random} = s/\sqrt{n}$, as well as error limits for the quantity of interest. The replicate error worksheet in Appendix A organizes this calculation.

std(array).[4] In Example 2.1 we show a calculation of sample mean, variance, and standard deviation using spreadsheet software. A worksheet in Appendix A (an excerpt is shown in Figure 2.1) organizes these types of calculations.

Example 2.1: A good estimate of the time to commute from home to workplace. *Over the course of a year, Eun Young takes 10 measurements of her commuting time under all kinds of conditions (Table 2.1). Calculate a good estimate of her time to commute. Calculate also the variance and the standard deviation of the dataset. What is your estimate of Eun Young's typical commuting time? What is your estimate of Eun Young's commuting time tomorrow?*

[4] The sample variance s^2 and the sample standard deviation s are different from the population variance σ^2 and population standard deviation σ, respectively. We calculate the sample variance from n measurements, which is just a sample of the entire population of all possible measurements. We cannot take all possible measurements, so we cannot know the population variance, but s^2 is a good estimate of the population variance in most cases. These issues are considered further in the statistics literature [5, 38].

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Table 2.1. *Data for Example 2.1: ten replicate measurements of commuting time.*

Index, i	Commuting time, min
1	23
2	45
3	32
4	15
5	25
6	28
7	18
8	50
9	26
10	19

Replicated Variable, Y:						**Units:**		
Measured values $Y_1, Y_2, \ldots, Y_n$		Sample Mean, $\bar{Y}$	Sample Variance, s^2	Sample Standard Deviation, s	Standard Error, $e_s = \frac{s}{\sqrt{n}}$	95% Confidence Interval based on n replicates (*Student's t* distribution)		
Y_1	23 min	28 min	131 min	11 min	?	$n = 1$	n/a	(include units) ± ?
Y_2	45					$n = 2$	$\pm 12.7 e_s$	
Y_3	32					$n = 3$	$\pm 4.30 e_s$	
Y_4	15					$n = 4$	$\pm 3.18 e_s$	
Y_5	25					$n = 5$	$\pm 2.78 e_s$	
Y_6	28					$n = 6$	$\pm 2.57 e_3$	
Y_7	...					$n \geq 7$	$\pm 2 e_s$	
						∞	$\pm 1.96 e_s$	

$$\bar{Y} \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$s^2 \equiv \frac{1}{(n-1)} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Figure 2.2 The replicate error worksheet can organize the calculations for the commuting-time example. We have not yet introduced the replicate standard error $e_{s,random}$; see the discussion surrounding Equation 2.26.

Solution: The data vary from 15 to 50 min, indicating that some factors significantly influence the time it takes to make this trip. We are asked to calculate the sample mean (Equation 2.8), variance (Equation 2.9), and standard deviation (Equation 2.10) of commuting time, x. Using the Excel functions AVERAGE(), VAR.S(), and STDEV.S(), and the data in Table 2.1, we calculate (see Figures 2.2 and 2.3):

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

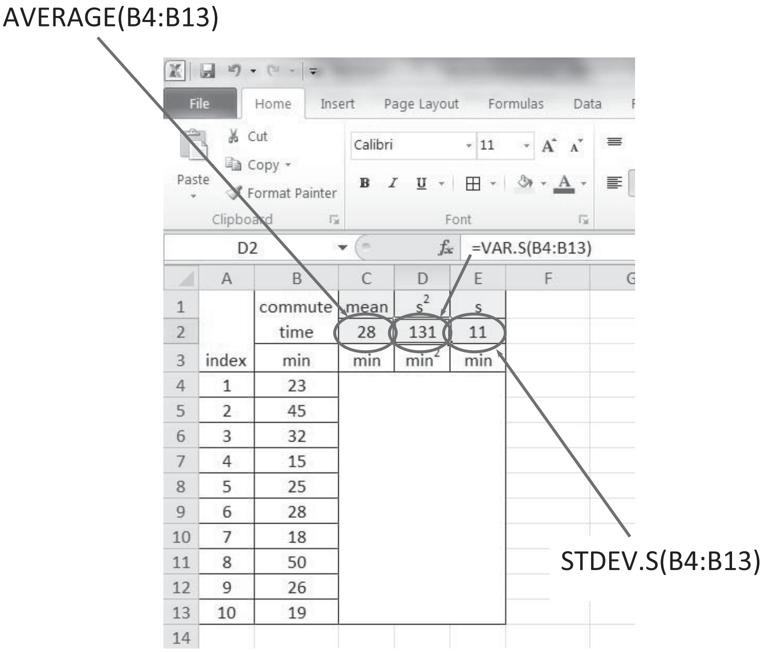

Figure 2.3 We use the spreadsheet software Excel to carry out the calculation of the mean and standard deviation of the data provided in Table 2.1.

$$\bar{x} = 28 \text{ min}$$
$$s^2 = 131 \text{ min}^2$$
$$s = 11 \text{ min}$$

Note the units on the quantities.

The meaning of the statistic $\bar{x}$ or “mean commuting time” is straightforward: on average, for the data available, the commute took 28 min. The other two statistics, s^2 and s, have been calculated with the help of the formulas and Excel, but as yet we do not know the meaning of these numbers. We know them simply as measures of the variability of the data; we address this topic next. Based on the calculations just performed, we estimate that, typically, it would take about 28 min (give or take) for Eun Young to make the commute.

In Example 2.1 we calculated the mean of a sample set, which is also the *expected value* of the variable we are measuring, the commuting time. At any given time, all things being equal, we expect that if Eun Young made her commuting trip, it would take about 28 min.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Perhaps you feel some unease with this prediction. If you were going to make that trip, would you allow exactly 28 min, or would you allow more or less time? After all, one time that Eun Young recorded her commute it took 50 min, and one time she made it in 15 min. Although we expect that, on average, the trip takes about 28 min, intuitively we know that it will take 28 min, give or take some minutes. We do not know how many minutes we should "give and take," however.

To explore this issue of how much to "give and take," consider the following. When we take sets of replicates, we can calculate the mean of the replicates $\bar{x}$ and may use that sample mean as the expected value of the variable x. Imagine we do this entire process – the sampling and the calculating – six times: take n measurements and calculate the means of each set. Will we get the same sample mean for all six sets of n replicates? We explore this question in Example 2.2.

Example 2.2: A second estimate of the mean time to commute from home. *Over the course of a year, Eun Young took 20 measurements of her commuting time under all kinds of conditions. Ten of her observations are shown in Table 2.1 and have already been discussed. Ten other measurements were recorded as well, using the same stopwatch and the same timing protocols; this second sample of* 10 *commuting times is shown in Table 2.2. Calculate a good estimate of Eun Young's time-to-commute from the second set of 10 observations. Calculate also the variance and the standard deviation of the*

Table 2.2. *Data for Example 2.2: ten replicate measurements of commuting time (second dataset).*

Index, i	Commuting time, min
1	27
2	40
3	25
4	22
5	45
6	22
7	28
8	35
9	32
10	41

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

second dataset. Compare your results to those obtained from the first dataset and discuss any differences.

Solution: The procedure for calculating the sample mean, variance, and standard deviation for this second set is identical to that used with the first set. The results are:

$$\bar{x} = 32 \text{ min}$$
$$s^2 = 68 \text{ min}^2$$
$$s = 8 \text{ min}$$

These numbers are different from those we obtained from the first set: the mean was 28 min in the first set, and it is 32 min in the second set. The variance is twice as big in the first set (131 min^2) as in the second set (68 min^2). It seems that taking the second dataset has only made things less clear, rather than more clear.

Although the numbers for the estimates of sample mean and standard deviation differ between the two datasets, this is not automatically cause for concern. Remember that we are sampling a population (all possible commuting times), and observations vary. It is because of this type of variation from sample to sample that the field of statistics has developed good methods (described here) to allow us to draw appropriate conclusions from data samples. The results in both Examples 2.1 and 2.2 are reasonable estimates of the mean commuting time, given the number and reproducibility of measurements used in calculating $\bar{x}$ and s each time. What we lack, as yet, is a way to express any single estimate of the mean commuting time $\bar{x}$ with its error limits. In addition, we would benefit from being able to express and defend how confident we are in the estimate that we calculate.

Before we acquired the second dataset, we thought we knew the average commuting time to be 28 min; then, the second dataset gave an average that was 4 min longer. Example 2.2 showed us that we need to exercise care when interpreting results drawn from a single dataset. If we obtain a second dataset, we may well get (indeed, will most likely get) different values for the mean and the standard deviation.

One positive aspect of the two measurements discussed in Examples 2.1 and 2.2 is that the two calculated means are not too different. If we took a third, fourth, fifth, and sixth sets of data, we expect we would get different numbers for $\bar{x}$ and s each time, but, as we saw earlier, we would expect the means of all sets to be reasonably close to each other.

The near reproducibility of sample means is a phenomenon that is well understood [5, 38]. When we repeatedly measure a quantity that is subject to stochastic variations, we do not get the same value every time. However, if we repeatedly sample the variable and calculate the average for each sample of size n, the values of the sample mean we obtain will be grouped around the true value of the mean (which we call μ) and will be symmetrically distributed around the true value. We can use these facts to develop error limits around the estimate of the sample mean.

We began with the question of determining a good estimate of Eun Young's commuting time, and this question has led us into the topic of how to quantify things that vary a bit each time we observe them. The uncertainty problem is not unique to commuting time: whenever we make an experimental measurement, we encounter stochastic variations and other sources of uncertainty.

Taking stock of the situation thus far:

1. In the course of our work, we often find that there is a stochastic variable x that interests us: in Examples 2.1 and 2.2, the variable is the commuting time; in a future example, it is the measured value of density for a liquid; or it could be any other measured quantity.
2. Due to random effects, individual observations x_i of the variable x are not identical.
3. If we average a set of n observations x_i of the stoichiometric variable x, we can obtain the average value $\bar{x}$ and the standard deviation s of the sample set. The average of the sample set is a good estimate of the value of the variable, and the standard deviation is an indication of the magnitude of stochastic effects observed in the sample set.
4. If we have one such sample of size n, mean $\bar{x}$, and standard deviation s, we would like to estimate how close (the "give and take") the sample mean $\bar{x}$ is to the true value of the mean of the distribution of x. The true value of the mean is given the symbol μ. A mathematical way of expressing our question is to ask, what is a range around a sample mean $\bar{x}$ within which there is a high probability that we will capture μ, the true value of the mean of x?

$$\mu = \text{estimate} \pm \text{(error limits)}$$
$$\mu = \bar{x} \pm \text{(error limits)}$$

5. The wider the error limits placed on a quantity, the higher the probability that we will capture the true value between the limits. Unfortunately, while expanding the error limits increases certainty, it also decreases the usefulness of the answer – it is not so helpful to say that the commuting time is somewhere between zero and a million hours. We need to establish

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

a reasonable middle ground between answers that are highly precise but uncertain (narrow error limits) and imprecise but certain (wide error limits).

We have taken important steps toward our goal of quantifying uncertainty. To understand and communicate quantities determined from measurements, we classify them as stochastic variables and apply the methods of statistics. We use statistical reasoning based on sampling to obtain a good estimate of the measured quantity, and statistics also allows us to obtain error limits and the associated probability of capturing the true value of the quantity within the error limits. Our goal in this chapter is to explain the basics of how all this is accomplished.[5]

2.2 Data Sampling

We have a single purpose, which is to identify appropriate error limits for a quantity we measure. Quantities we measure are continuous stochastic variables.

$$\begin{array}{c}\text{Goal: to determine a}\\ \text{plausible value}\\ \text{of a measured quantity}\\ \text{in the form}\end{array} \qquad \text{Answer} = (\text{value}) \pm (\text{error limits}) \tag{2.11}$$

We seek to address this purpose by taking samples of the stochastic variable of interest. To explain the role of sampling in uncertainty analysis, we begin with a discussion of the mathematics of continuous stochastic variables. Taking an experimental data point is, in a statistical sense, "making an observation of" or "sampling" a *continuous stochastic variable*.

A key tool that characterizes a continuous stochastic variable is its *probability density function* (*pdf*). The pdf of a stochastic variable is a function that encodes the nature of the variable – what values the variable takes on and with what frequency or probability. A stochastic variable has its own inherent, underlying pdf called the population distribution. As we discuss in this section (see Equation 2.12), the probability of a continuous stochastic variable taking on a value within a range is expressed as an integral of its pdf across that range.

[5] Once the basics are established, we finish the consideration of Eun Young's commuting time in Example 2.7, where we determine the error range for mean commuting time based on the 95% confidence level.

We *sample* stochastic variables to learn about their probability density functions. When we sample a stochastic variable and calculate a statistic such as the sample mean or the sample variance, the value obtained for the chosen statistic likely will be different every time we draw a sample, as we saw in the previous section. The chosen statistic (i.e., the sample mean $\bar{x}$ or the sample variance s^2) is itself a continuous stochastic variable, and it has its own probability density function separate from that of the underlying distribution associated with x. It turns out that the pdf of the statistic "sample mean" allows us to quantify the probabilities we need to establish sensible error limits for measured data.

The pdf of the statistic *sample mean* allows us to quantify the probabilities we need to establish sensible error limits for measured data.

We discuss this in Section 2.2.3.

Finally, the probability density function of sample means is quite reasonably taken to be a well-known distribution called the *Student's t distribution*. We discuss why this is the case and show how Excel and MATLAB can facilitate determinations of error limits with the Student's t distribution.

To recap, this section contains (1) an introduction to the topic of continuous stochastic variables and their pdfs; (2) information on how to determine probabilities from probability density functions; and (3) discussion of determining error limits for stochastic variables using sampling and the Student's t distribution. These topics advance our goal of learning to quantify uncertainty in experimental measurements. For some readers, it may serve your purposes to skip ahead to Section 2.3, which shows how the methods discussed here are applied to practical problems. After reviewing the examples, a reader who has skipped ahead may wish to return here to explore why, when, and how these methods work.

2.2.1 Continuous Stochastic Variables

When flipping a coin, what is the probability (Pr) that the result comes up heads? This is a first question in the study of probability. The answer is $\Pr = \frac{1}{2}$, since there are two possible states of the system – heads and tails – and each is equally likely. Thus, we expect that half the time, on average, the result of a coin toss will yield heads and half the time the result will be tails. The outcome of a coin toss is a discrete random variable. When a stochastic variable is

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

discrete, meaning the variable has a finite set of outcomes, it is straightforward to calculate probabilities: probability of an outcome equals the number of ways the outcome may be produced divided by the total number of possible outcomes.

When measuring commuting time (Example 2.1), what is the probability that it will take Eun Young 29.6241 min to make her commute? This probability is not at all obvious. In addition, perhaps we can agree that it would be highly unlikely that she would ever take exactly 29.6241 min to make her commute. Commuting time is a *continuous* variable, and this type of variable requires a different type of probability question.

Getting a useful answer
requires first identifying
a useful question.

What is the probability that it will take Eun Young between 20 and 30 min to make her commute? This probability is also not obvious, but it seems like a more appropriate question for a variable such as commuting time. Based on the data we have seen in Examples 2.1 and 2.2, it seems likely that the probability would be pretty high that Eun Young's commute would take between 20 and 30 min. We guess that this probability (Pr) is greater than 0.5 and perhaps as high as Pr $=$ 0.7 or 0.8 (80% chance). We know the probability would not be Pr $=$ 1 (100%), since when we measured the commuting time in Example 2.1 one trip took 50 min. At this point it is not so clear how to be more rigorous in estimating these probabilities.

In the preceding discussion, we explored the difference between establishing probabilities with discreet stochastic variables, such as the outcome of a coin toss, and with continuous stochastic variables, such as commuting time. For discrete stochastic variables, we establish probabilities by counting up all the possible outcomes and calculating the number of ways of achieving each outcome. For continuous stochastic variables, we cannot follow that procedure. We cannot count the number of possible commuting times between 20 and 30 min and we cannot count the number of ways of having Eun Young take 29.6241 min to make her commute. Continuous stochastic variables require a different approach than that used for discrete variables. The approach we use, based on calculus, is to define a pdf and to calculate probabilities by integrating the pdf across a range of values. We discuss how and why this works in the next section.

2.2.2 Probability Density Functions

Commuting time is a continuous stochastic variable. Quantities such as commuting time are *continuous* because, in contrast to discrete quantities (the number of people in a room, for example), values of commuting time are not limited to integer values but can take on any decimal number. The continuous nature of experimental variables affects how we quantify the likelihood of observing different values of the variable.

When sampling a continuous stochastic variable, the probability of observing any *specific* outcome is very small, basically zero [38]. For example, in the commuting-time measurement, if we ask about the likelihood of observing a commuting time of exactly 29.6241 min, the answer is zero. Any other precise value is highly unlikely as well; we can even say that between 20 and 30 min there are an infinite number of unlikely values.

Yet our experience tells us that observing some commuting time in this interval is likely. Individual values are unlikely, but when we aggregate over intervals (integrate), the probability becomes finite. The choice of interval matters as well. The interval between 20 and 30 min captures much of the data we know about for Eun Young's commuting time, but this interval is special: not all intervals are equally likely to contain observed commuting times. For example, it is unlikely to observe a commuting time between 100 and 110 min. Thus, probability changes when we ask about different intervals. In addition, the breadth of the interval changes the probability. At one extreme, if we choose a very narrow interval, we find the probability is zero. If we broaden our interval, we are more likely to capture observed commuting times. If we use an extremely broad interval, the probability of observing a value of commuting time becomes nearly certain. If we choose finite-sized limits throughout the domain of possible values, we obtain different, finite probabilities.

An effective approach to the challenge of calculating probabilities for continuous variables is to think of probability in terms of how it adds up over various intervals. We define the *probability density function* $f(x)$ to calculate the probability Pr of the variable x taking on values in an interval between limits a and b (Figure 2.4):

Definition of f:
probability is expressed
as an integral of a probability
density function (pdf)
(continuous stochastic variable)

$$\boxed{\Pr[a \leq x \leq b] \equiv \int_a^b f(x')dx'} \tag{2.12}$$

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

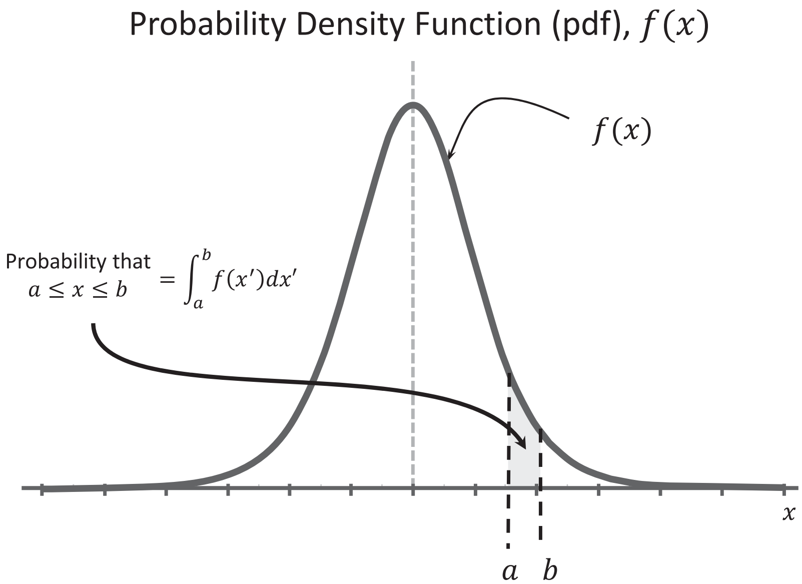

Figure 2.4 For continuous probability distributions, we cannot evaluate the probability of observing a particular value such as an average commuting time of 29.6241 min, but we can calculate the probability that the variable of interest is bracketed in a range of values. (What is the probability that the commuting time is between 28 and 30 min?) The probability is calculated as the area under the curve of the probability density function, between the two values that form the range of interest.

The quantity $f(x')dx'$ is the probability that x takes on a value between x' and $x' + dx'$. The integral represents the result of adding up all the probabilities of observations between a and b. If we let $a = -\infty$ and $b = \infty$, then the probability is 1. For all other intervals $[a,b]$, the probability is less than 1 and is calculated by integrating the pdf $f(x)$ between a and b.[6]

In the remaining sections and chapters of this text, Equation 2.12 is the key tool for determining error limits for experimental data. We seek error limits $[a,b]$ on the expected value of x so that the probability is high that the true value of a quantity we measure is found in that range. We now explain how that error-limit range $[a,b]$ is found.

To build familiarity with probability density functions and probability calculations, it is helpful at this point to work out a specific probability example

[6] We use the notation x' in the integral to make clear that x' is a dummy variable of integration that disappears once the definite integral is carried out. This distinguishes temporary variable x' from the variable x, which has meaning outside of the integral.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

for a case when the pdf is known. After the example, we turn to the question of how one determines a pdf for a system of interest.

Example 2.3: Likelihood of duration of commuting time. *If we know the pdf of George's commuting time is the function given in Equation 2.13 (plotted in Figure 2.5), what is the probability that his commute takes between 25 and 35 min? What is the probability that the commute will take more than 35 min?*

Probability density function (pdf) of George's commuting time

$$f(x) = Ae^{-\frac{(x-B)^2}{C}} \tag{2.13}$$

where $A = (1/\sqrt{50\pi})$ *min*$^{-1}$, $B = 29$ *min, and* $C = 50$ *min*^{2}.

Solution: The definition of pdf in Equation 2.12 (repeated here) allows us to calculate probabilities for continuous stochastic variables if the pdf is known, as it is in the case of George's commuting time.

$$\Pr[a \leq x \leq b] \equiv \int_a^b f(x')dx' \qquad \text{(Equation 2.12)}$$

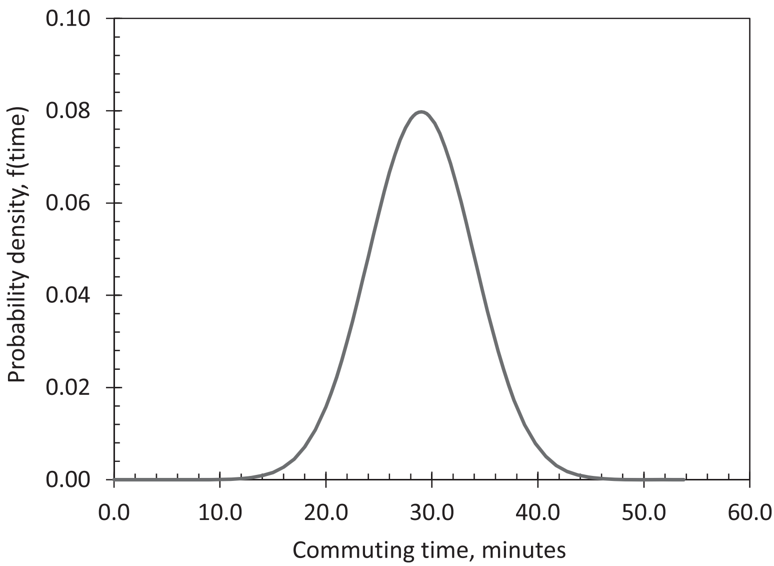

Figure 2.5 The probability density function that characterizes George's commute. Having the pdf makes probability calculations straightforward; later we see how to make probability estimates for variables without knowing the underlying pdf.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

For the question posed in this example, we obtain the requested probability by integrating the pdf (Equation 2.13) between 25 and 35 min.

$$\Pr\left[25 \le x \le 35 \text{ min}\right] = \int_{25}^{35} A e^{-\frac{(x'-B)^2}{C}} dx' \qquad (2.14)$$

We have everything we need to finish the problem; the rest is mathematics.

The integral in Equation 2.14 is a deceptively simple one, and it does not have a closed-form solution. This integral is sufficiently common in mathematics that it has been defined as a function all its own, called the *error function*, erf (u). Like $\sin u$ and $\ln u$, the function erf u comes preprogrammed in mathematical software.

$$\text{Error function (defined):} \quad \boxed{\text{erf}\,(u) \equiv \frac{2}{\sqrt{\pi}} \int e^{-u^2} du} \qquad (2.15)$$

$$\int e^{-u^2} du = \frac{\sqrt{\pi}}{2} \text{erf}\,(u) \qquad (2.16)$$

We carry out the integral in Equation 2.14 in terms of the error function, which we subsequently evaluate in Excel [in both MATLAB and Excel the command for the error function is ERF()]:[7]

$$\begin{aligned}
\Pr\left[25 \le x \le 35 \text{ min}\right] &= \int_{25}^{35} A e^{-\frac{(x'-B)^2}{C}} dx' = A\sqrt{C} \int_{25}^{35} e^{-\left(\frac{(x'-B)}{\sqrt{C}}\right)^2} \left(\frac{1}{\sqrt{C}} dx'\right) \\
&= \left(A\sqrt{C}\right)\left(\frac{\sqrt{\pi}}{2}\right) \text{erf}\left(\frac{x'-B}{\sqrt{C}}\right)\Bigg|_{x'=25}^{x'=35} \\
&= \frac{1}{2}\left[\text{erf}\left(\frac{(35-29)}{\sqrt{50}}\right) - \text{erf}\left(\frac{(25-29)}{\sqrt{50}}\right)\right] \\
&= \frac{1}{2}(0.76986 - (-0.57629)) \\
&= 0.673075 = 67\%
\end{aligned}$$

We obtain the result that about two thirds of the time George needs between 25 and 35 min to make his commute.

For the second question, to determine how often the commute will be more than 35 min, we integrate the pdf from 35 to ∞; the answer is that there is a 12% probability of a commute 35 min or longer (this calculation is left to the reader; Problem 2.5).

[7] Hint: In the calculation shown, let $u \equiv \frac{(x'-B)}{\sqrt{C}}$ and thus $du = \frac{dx'}{\sqrt{C}}$.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

As we saw in Example 2.3, it is straightforward to calculate probabilities for continuous stochastic variables when the pdf of the underlying distribution is known. With this ability, we can answer some interesting questions about the variable. The problem now becomes, how do we determine the probability density function for a continuous stochastic variable of interest? For Eun Young's commuting time, for instance (Examples 2.1 and 2.2), how do we obtain $f(x)$?

Determining a pdf is a modeling question. As with throwing a die, if we know the details of a process and can reason out when different outcomes occur, we can, in principle, reason out the pdf. To do this, we research the process and sort out what affects the variable, and we build a mathematical model.

For complicated processes such as commuting time, a large number of factors impact the duration of the commute – the weather, ongoing road construction, accidents, seasonal events. Unfortunately, there are too many factors to allow us to model this process accurately. This is the case with many stochastic variables. Because of complexity, the most accurate way to determine the pdf of a real quantity turns out to be to measure it, rather than to model it. If we have patience and resources, a reasonably accurate version of the pdf is straightforward to measure: we make a very large number of observations over a wide range of conditions and cast the results in the form of a pdf.

Although measuring the pdf is straightforward if we are patient and well financed, measuring a pdf is rarely easy. Measuring the pdf for commuting time is a substantial project: to accurately determine the pdf, we must ask Eun Young to time her commute for years under a variety of conditions. Before embarking on this measurement, it would be reasonable to ask, are we justified in making this effort? If we just want to know how to plan a future commute, can we do something useful that is less time-consuming than measuring the pdf?

For casual concerns about one person's commuting time, there is little justification for undertaking the difficult and complex task of measuring the pdf. Many realistic questions we might ask about Eun Young's commuting time could be addressed by taking a guess at the probable time, then adding some extra time to the estimate to protect against circumstances that cause the commute to be on the longer side. For questions relating to commuting time, we probably do not really need to know the best value and its error limits – a worst-case estimate is sufficient (see, for example, Problem 2.30). We characterize this approach as relying on an expert's opinion and incorporating a safety factor; the effort to determine the pdf is not warranted.

Although worst-case thinking has its place in decision making, for scientific and engineering work a worst-case estimate of a quantity of interest is often insufficient. In science, our measurements are usually part of a broader project in which we hope to make discoveries, learn scientific truths, or build reliable devices, processes, or models. We may measure density, for example, as part of the calibration of a differential-pressure meter. The accuracy of the calibrated meter depends substantially on the accuracy of the density measurement for the fluid used during calibration, and we need to know the extent to which we can rely on our measurements. Many technological applications of measurements are like this – dependent on instrument and measurement accuracy. Not infrequently, accuracy can be a matter of life and death (for example, when building structures, designing safety processes, and manufacturing health-related devices) or can be what determines the success/failure of a project (cost estimates, instrument sizing, investments). In such cases we cannot use a worst-case value and talk ourselves out of the need to establish credible error limits on our numbers.

The objections raised here seem to argue that we have no choice but to measure the pdf for stochastic variables that we study for scientific and engineering purposes. This sounds like a great deal of work (and it is), but there is some good news that will considerably reduce this burden.

First, it turns out that we can quite often reasonably assume that the stochastic effects in measurements are *normally distributed*; that is, their pdf has the shape of the normal distribution (Figure 2.6):

Normal probability distribution (pdf)

$$\boxed{f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \tag{2.17}$$

The normal distribution (see the literature [38] and Section E.2 in the appendix and for more details) is a symmetric distribution with two parameters: a mean μ, which specifies the location of the center of the distribution, and the standard deviation σ, which specifies the spread of the distribution. With the assumption that the random effects are normally distributed (that is, they follow Equation 2.17), we reduce the problem of determining the pdf to determining the two parameters μ and σ.

Second, often the questions we have about a stochastic variable can be answered by examining a *sample* of the values of the stochastic variable. We introduce sampling to get around the difficulty of determining the underlying pdf of the variable. The approach is this: we take data replicates – three, four, or

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

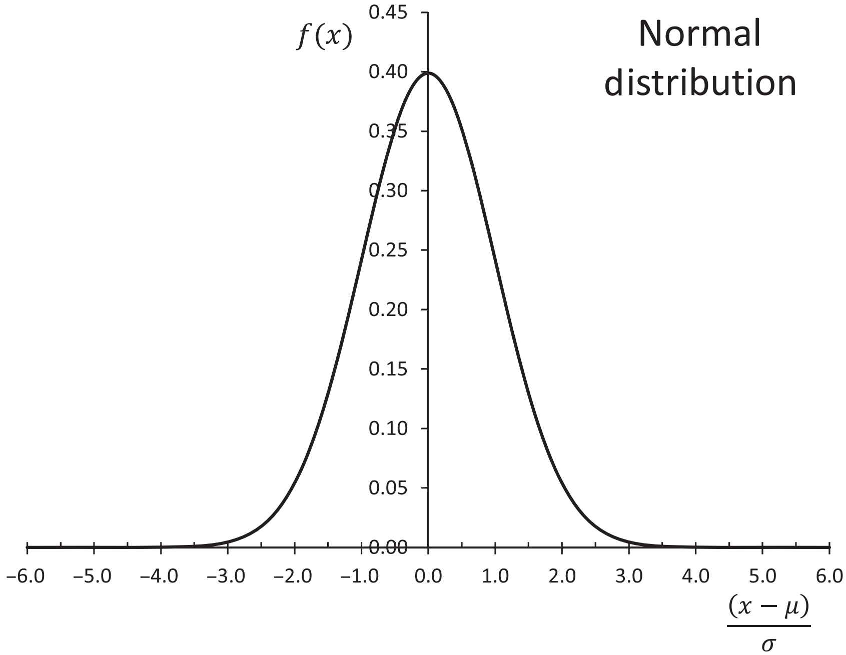

Figure 2.6 The normal distribution, famous as the bell curve, is a symmetric probability density distribution characterized by two parameters, μ, which is its center, and σ, which characterizes its spread. The pdf has been plotted here versus a dimensionless version of x, translated to the mean $(x - \mu)$ and scaled by the standard deviation of the distribution. When $\mu = 0$ and $\sigma = 1$, this is called the standard normal distribution. For the normal distribution, 68% of the probability is located within $\pm\sigma$ of the mean, and 95% of the probability is located within $\pm 1.96\sigma$ or $\approx \pm 2\sigma$ of the mean.

more observations – and we ask, what is a good value (including error limits) of the stochastic variable based on this sample set? The sample is not a perfect representation of the variable, but thanks to considerable study, the field of statistics can tell us a great deal about how the characteristics of samples are related to the characteristics of the true distribution – all without us having to know in detail the underlying pdf of the variable. We can put this statistical knowledge to good use when determining error limits.

In the next section we show how we use a probability density function called the *sampling distribution of the sample mean* $\bar{x}$ to determine a good value (including error limits) for a stochastic variable based on finite data samples.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

2.2.3 Sampling the Normal: Student's *t* Distribution

We return to our purpose, which is to identify error limits for a quantity we measure.

$$\begin{array}{c}\text{Goal: to determine a}\\ \text{plausible value}\\ \text{of a measured quantity}\\ \text{in the form}\end{array} \qquad \text{Answer} = (\text{value}) \pm (\text{error limits}) \tag{2.18}$$

We seek to address this purpose by taking samples of the measured quantity, which is a stochastic variable. We take replicate measurements (a sample of size n) of the quantity, from which we wish to estimate a good value of the measured quantity. To determine the appropriate error range about that value, we also ask: what is the size of the error range we need to choose (Figure 2.7) to create an interval that has a good chance (we define a "good chance" as 95% probability) of including the true value of the variable?

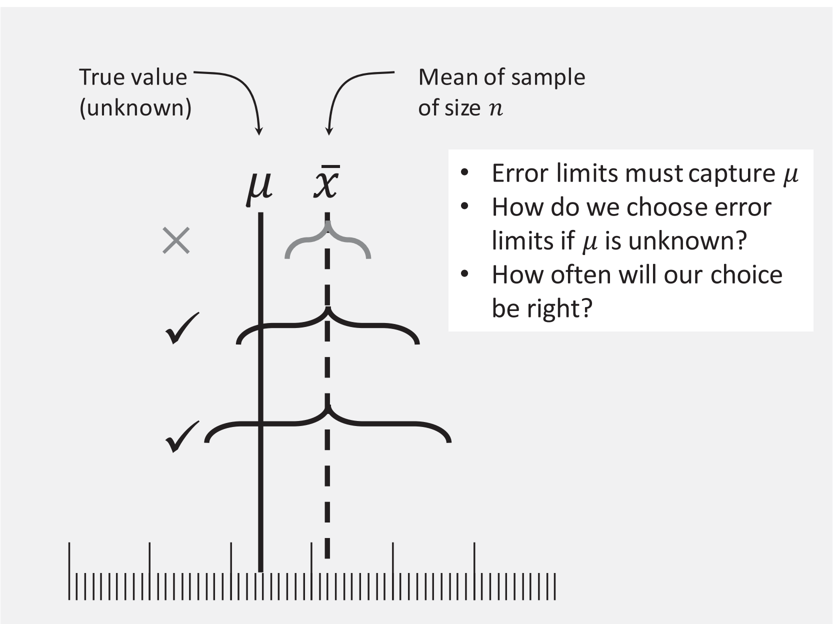

Figure 2.7 We measure $\bar{x}$ and seek to determine error limits that capture the true value of the variable, μ. We do not know μ, however. Whatever we choose, we also need to be able to say quantitatively how likely we believe it is that the mean $\pm$ the error limits will capture the true value of the variable.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Restated goal: to determine error limits on an estimated value that, with 95% confidence, captures the true value of x	Answer = (value) $\pm$ (error limits)

(2.19)

The value of writing our goal this way is that, through sampling, we can address the restated goal.

As we discuss here, we can say a great deal about samples of a variable without knowing the details of the variable's underlying distribution. For the quantity we are measuring, we first agree to assume that its underlying pdf is a normal distribution of unknown mean μ and unknown standard deviation σ (Equation 2.17).[8] Second, we take a sample of n measurements and calculate $\bar{x}$ and s for the sample set. Finally, we pose our questions. First, based on the sample set, what is a good estimate of the mean μ of the underlying distribution? Second, what $\pm$ error limits should we apply to the good estimate so that, with 95% confidence, the true value μ is captured in the range (value) $\pm$ (error limits)?

These questions deal with the sampling process. The sampling process introduces a new continuous stochastic variable, the sample mean $\bar{x}$. We saw in our commuting-time examples earlier in the chapter that when we draw a sample and calculate the sample mean $\bar{x}$, the observed mean varies from sample to sample. It is intuitive, perhaps, that the values of $\bar{x}$ will be in the neighborhood of the true value of x. This can be rigorously shown to be true [5, 38]. Formally, we say that when a random sample of size n is drawn from a population of stochastic variable x, the expected value of the sample mean $\bar{x}$ is equal to μ, the true value of the mean of the underlying distribution of x.[9] In terms of our goal, this result from statistics tells us that sample mean $\bar{x}$ is a good estimated value for μ.

Expected value $E()$ of sample mean $\bar{x}$ (any distribution)	$E(\bar{x}) = \mu$	(2.20)

Taking $\bar{x}$ as our estimate of a good value of x, we next ask about the probability that an individual observation of sample mean $\bar{x}$ will be close or not close to μ, the mean of the underlying distribution of x. To answer this question, we start by quantifying "close" or "not close" to the true value by defining the *deviation* as the difference between the observed sample mean and the true value of the mean of x.

[8] We can relax this assumption later.
[9] See the literature discussion of the central limit theorem [15, 38].

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$$\text{Deviation between the observed sample mean } \bar{x} \text{ and } \mu \text{, the true mean of the stochastic variable } x \qquad \text{deviation} \equiv (\bar{x} - \mu) \tag{2.21}$$

The deviation defined in Equation 2.21 is also a continuous stochastic variable. Since the normal distribution is symmetric, the deviations are equally likely to be positive and negative, and overall the expected (mean) value of the deviation $(\bar{x} - \mu)$ is zero. As with all continuous stochastic variables, the probability of observing any specific value of the deviation is zero, but if we have the pdf for the deviation, we can calculate the probability that the deviation takes on values between two limits (using Equation 2.12). For example, for a measured average commuting time of $\bar{x} = 23.3$ min, if we knew the pdf $f(\bar{x} - \mu)$ of the deviation defined in Equation 2.21, we could answer the question, what is the probability that the deviation of our measured mean from the true mean, in either direction, is at most 5.0 min?

$$\text{Maximum deviation} \qquad |(\bar{x} - \mu)_{max}| = 5.0 \text{ min} \tag{2.22}$$

$$\text{Probability that } (\bar{x} - \mu) \text{ is in the interval } [-5.0, 5.0] \qquad \Pr[-5.0 \leq (\bar{x} - \mu) \leq 5.0] \tag{2.23}$$

$$\Pr[-5.0 \leq (\bar{x} - \mu) \leq 5.0] = \int_{-5.0}^{5.0} f(\bar{x}' - \mu) d(\bar{x}' - \mu) \tag{2.24}$$

This question about the deviation is the same as asking, what is the probability that μ, the true value of x, lies in the range 23.3 ± 5.0 min? By focusing on deviations, the error-limits problem now becomes the question of determining the pdf $f(\bar{x} - \mu)$ of the deviation between the sample mean $\bar{x}$ and the true mean μ. This formulation has the advantage that to determine the pdf of the deviation we do not need to know μ, the true mean value of x.

> By focusing on the pdf of *deviation* $(\bar{x} - \mu)$, rather than of $\bar{x}$, we avoid having to know μ, the true mean value of x.

To recap, in a previous section we established the pdf's role as the tool needed to calculate the probability that a stochastic variable takes on a value within a range (Equation 2.12). When the pdf of underlying experimental errors is not known, we customarily assume that the underlying distribution is a normal distribution (this is a good assumption for many experimental errors), and we sample the distribution (obtain n replicates). We use the mean of the sample set $\bar{x}$ to determine a good value of x. To determine error ranges for our good value of x, we then seek the pdf of a new continuous stochastic variable, $(\bar{x} - \mu)$, the deviation of the sample mean from the true value of the mean. Concentrating on $(\bar{x} - \mu)$ means we avoid having to know the value of μ.

The pdf we seek, $f(\bar{x} - \mu)$, the pdf of the deviation between a sample mean and the true mean of a normally distributed population of unknown standard deviation, has been determined [16, 34], and we present its mathematical form next (it is Student's t distribution with $(n - 1)$ degrees of freedom, Equation 2.29). The derivation of $f(\bar{x} - \mu)$ is sufficiently complex that we do not present it here. It is useful, however, to sketch out the properties of the distribution by collecting our expectations for the distribution. This exercise helps us understand the answer from the literature, which we use to construct error limits.

Characteristics of the pdf of the deviation of a sample mean from the true mean $(\bar{x} - \mu)$:

1. For all sample sets, it seems reasonable that the most likely value of $\bar{x}$ is μ and thus that the most likely value of the deviation $(\bar{x} - \mu)$ is zero.
2. Since errors are random, positive and negative deviations are equally likely, and thus the probability density function of $(\bar{x} - \mu)$ is expected to be symmetric around zero.
3. We do not expect observed deviations $(\bar{x} - \mu)$ to be very large; thus the probability that $(\bar{x} - \mu)$ is large-positive or large-negative is very small.
4. If the standard deviation of a sample set s is large, this suggests that random variations are large, and the probability density of the deviation $(\bar{x} - \mu)$ will be more spread out.
5. If the standard deviation of a sample set s is small, this suggests that stochastic variations are small, and the probability density of the deviation $(\bar{x} - \mu)$ will be more tightly grouped near the maximum of the pdf, which is at $(\bar{x} - \mu) = 0$.
6. If the sample size n is small, we know less about the variable x, and thus the probability density of the deviation $(\bar{x} - \mu)$ will be more spread out.
7. If the sample size n is large, we know more about the variable x, and thus the probability density of the deviation $(\bar{x} - \mu)$ will be grouped closer to the maximum of the pdf, which is at $(\bar{x} - \mu) = 0$.

The probability density function of the deviation of the sample mean from the true mean has been worked out for the case when the underlying distribution is a normal distribution of unknown standard deviation [5, 16, 34, 38]. In agreement with our list of the characteristics of this pdf, the distribution depends on the sample standard deviation s and the sample size n. To write the distribution compactly, it is expressed in terms of a dimensionless *scaled deviation*, t:

$$\text{Scaled deviation} \qquad \boxed{t \equiv \frac{(\bar{x} - \mu)}{s/\sqrt{n}}} \tag{2.25}$$

where s and n are the sample standard deviation and sample size, respectively. The deviation $(\bar{x} - \mu)$ is scaled by the replicate standard error $s/\sqrt{n}$, producing the unitless stochastic variable t.

$$\begin{array}{c}\text{Standard error}\\ \text{of replicates}\end{array} \qquad \boxed{e_{s,random} \equiv \frac{s}{\sqrt{n}}} \tag{2.26}$$

The quantity $s/\sqrt{n}$ is also called the standard random error. For individual observations of the mean $\bar{x}$ of a sample set of size n, t represents the number of standard errors the mean lies from the true value of the mean μ (Figure 2.8).

The quantity $s/\sqrt{n}$ in Equation 2.25 has meaning. It appears during consideration of the properties of the underlying distribution of x, the distribution we are sampling. When the standard deviation of the underlying normal distribution is known, its sampling distribution is also a normal distribution, and we can easily show (see Appendix E, Section E.4) that the standard deviation of the sample mean is $\sigma/\sqrt{n}$.

$$\begin{array}{c}\text{Normal}\\ \text{population}\\ (\text{known } \sigma)\end{array} \quad \begin{array}{c}\text{Standard}\\ \text{deviation}\\ \text{of the mean}\end{array} = \frac{\sigma}{\sqrt{n}} \tag{2.27}$$

When the standard deviation of the underlying distribution not known, the standard deviation of the mean is estimated by substituting s for σ, where s is the sample standard deviation.

$$\begin{array}{c}\text{Sampling normal}\\ \text{with unknown}\\ \text{standard deviation}\end{array} \quad \begin{array}{c}\text{Standard}\\ \text{deviation}\\ \text{of the mean}\\ (\text{estimate})\end{array} = \frac{s}{\sqrt{n}} \tag{2.28}$$

The pdf of the sampling distribution of an underlying normal distribution of unknown standard deviation is not a normal distribution. The problem of determining the statistics of sampling an underlying normal distribution of unknown standard deviation was worked out in the late 1800s by William Sealy

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

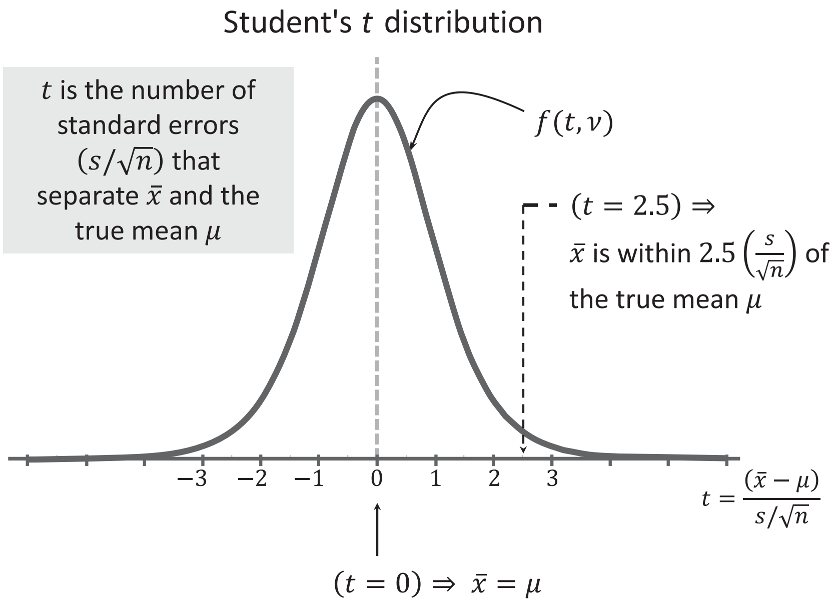

Figure 2.8 The scaled variable t indicates how many standard errors $s/\sqrt{n}$ separate the observed sample mean $\bar{x}$ and the true mean μ. The Student's t probability density distribution allows us to calculate the probability of observing values of t within a chosen range.

Gosset, an analyst for the Guinness Brewery in Ireland, and the distribution is named for the pseudonym under which he published ("Student") [15, 16].

Student's t distribution pdf of the scaled deviation t for various degrees of freedom ν

$$\boxed{f(t,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}} \tag{2.29}$$

The Student's t distribution is the pdf of t, the scaled deviation of the sample mean from the true mean; it is written as a function of t and ν, where ν is called the degrees of freedom (Figure 2.9).[10] $\Gamma()$ is a standard mathematical function called the gamma function [54]. The Student's t probability distribution function $f(t,\nu)$ is plotted in Figure 2.9 for various values of ν, and we see that, as expected, it is symmetric, peaked at the center, and with very little density in its tails at low and high values of t. For the problem of sampling means

[10] Degrees of freedom are important when we are estimating parameters – it is important to avoid overspecifying a problem [5, 38].

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

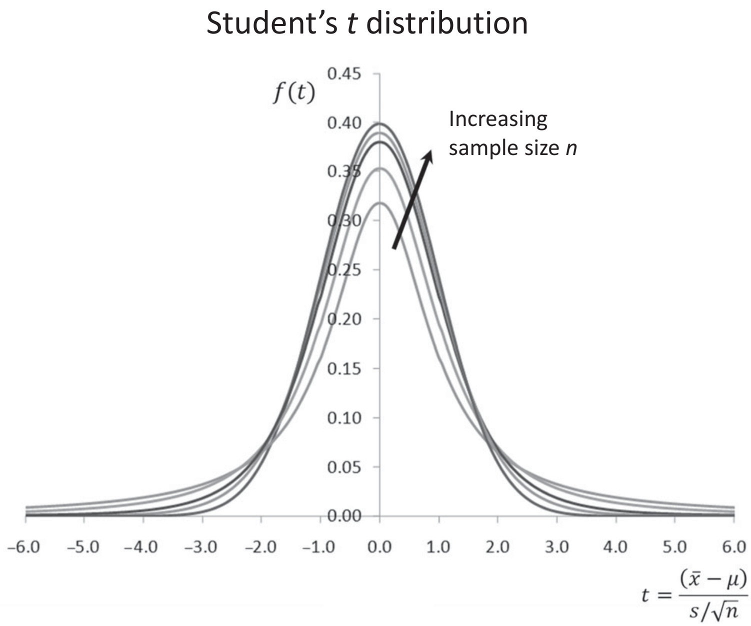

Figure 2.9 The Student's t distribution $f(t, \nu)$ describes how the value of the sample mean is distributed for repeated sampling of a normally distributed stochastic variable of unknown standard deviation σ. The Student's t probability distribution depends on the sample size n, and $f(t, \nu)$ gives the probability density as a function of the scaled deviation $t = \frac{(\bar{x}-\mu)}{s/\sqrt{n}}$. As n approaches infinity, the Student's t distribution approaches the standard normal distribution. Lines shown are for different sample sizes, expressed in terms of degrees of freedom $\nu = (n - 1)$, where the values of ν in the figure are $\nu = 1, 2, 5, 10,$ and ∞.

for samples of size n, the applicable sampling distribution is the Student's t distribution with $\nu = (n - 1)$ degrees of freedom.

Recall our purpose: we seek to determine error limits on measurements of a stochastic variable x. The mean $\bar{x}$ of a set of replicates of the variable is a good estimate of the variable (Equation 2.20). To quantify variability in the sample set, we seek the sampling distribution of the mean of the sample set (Equation 2.29), which is based on the scaled deviation t. The use of deviation avoids the need to know the true mean μ, which we do not know. The scaled deviation t (Equation 2.25) is scaled by an estimate of the standard deviation of the sample mean. This estimate is based on the more straightforward case when the standard deviation is known, the case when the underlying distribution is the normal distribution. Not knowing the standard deviation of the underlying

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

distribution changes and complicates the pdf of the sampling distribution. When the standard deviation of the underlying distribution is not known, the pdf of the sampling distribution of the mean is the Student's t distribution with $(n-1)$ degrees of freedom [16, 34]. With the Student's t distribution pdf and the methods of creating error limits discussed in Section 2.2, we are now able to determine error limits with known levels of confidence.

The formula for the Student's t distribution in Equation 2.29 as well as integrals of $f(t,\nu)$ are programmed into Excel and MATLAB, making the Student's t distribution very easy to use in error-limits calculations (several examples are provided in the next section). Excel is used most of the time; see Table D.1 for equivalent MATLAB commands). In the next section we discuss how to apply the Student's t distribution to the problem of determining appropriate error limits for data replicates.

2.3 Replicate Error Limits

2.3.1 Basic Error Limits

This chapter is about quantifying replicate error, a type of *random* uncertainty present in our data. Replicate error shows up when we make repeated measurements of a stochastic variable – we do not get the same number every time. As discussed in the previous sections, random statistics teaches us that the average of repeated measurements is a good estimate of the true value of the variable. Statistics also guides us as to how to write error limits for this estimated value. The essential tool for writing limits due to random error is the Student's t distribution, introduced in Section 2.2. This distribution allows us to quantify likely variability of the data based on the properties of a sample of the variable. In this section, we show how to use the Student's t distribution to answer some very practical questions about uncertainty in experimental data.

We present two introductory examples, the first one addressing the question of a warrantied average value for a commodity and the second using the Student's t distribution to assess different choices for the number of significant figures to report for an experimental result. These two initial examples lead to the definition of the most common form of the error limit, the 95% confidence interval, which we discuss in depth in Section 2.3.2.

Example 2.4: Probability with continuous stochastic variables: guaranteeing the mean stick length. *A shop sells sticks intended to be* 6 *cm long, and the vendor claims that the average stick length of his stock is between* 5.5 *cm and* 6.5 *cm. To assess this guarantee, we measure the lengths of* 25 *sticks and* *find that the lengths vary a bit. We calculate the mean length to be* $\bar{x} = 6.0$ *cm, and the standard deviation of the sample set is* $s = 0.70$ *cm. Based on these data, what is the probability that the true average stick length in the vendor's stock is between 5.5 and 6.5 cm, as claimed?*

Solution: This is a question about the mean of the distribution of stick lengths. The best estimate of average stick length is the sample mean, $\bar{x} = 6.0$ cm, which for our sample has just the value that the vendor hoped it would have. The lengths vary, however. Based on the sample of stick lengths obtained, how confident should the vendor be that the average stick length is between 5.5 and 6.5 cm? In terms of probability, based on the sample, what is the probability that the true population-average stick length is between 5.5 and 6.5 cm?

We can ask this question in terms of deviation: what is the probability that the maximum deviation of the sample mean $\bar{x}$ from the true mean of the population of stick lengths will be 0.5 cm?

$$\begin{array}{c}\text{Maximum deviation}\\ \text{of the mean from the true}\end{array} \qquad |(\bar{x} - \mu)_{max}| = 0.5 \text{ cm} \tag{2.30}$$

The Student's t distribution with $(n - 1)$ degrees of freedom expresses the probability of observing deviations of various magnitudes, if we have a sample from the population. The Student's t distribution is the pdf for the stochastic variable t (Equation 2.25). As we discussed in Section 2.2, the pdf of a continuous stochastic variable allows us to calculate the probability that the stochastic variable will take on a value within a range; thus, we are able to use the Student's t distribution to calculate the probability that a value of t lies within some range.

To translate our question about mean stick length into a question about values of t, we examine the definition of t:

$$\text{Scaled deviation } t \qquad t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}} \tag{2.31}$$

The quantities in the definition of t refer to three properties of a sample of a stochastic variable (sample mean $\bar{x}$, sample standard deviation s, and number of observations in the sample n) and one property of the underlying population that has been sampled (μ, the mean of the population). For the current discussion of stick lengths, we do not know the population mean μ, but we know all three properties of a sample and we know the target maximum deviation $(\bar{x} - \mu)_{max}$. Thus, we can calculate a value of t_{limit} associated with the target deviation. The question we are seeking to answer asks about the mean stick length deviation being no more than ± 0.5 cm. Thus, for t_{limit}

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

calculated from the maximum deviation of 0.5 cm, the range of t from $-t_{limit}$ to $+t_{limit}$ exactly expresses our mean-stick-length question: the probability that t falls between $-t_{limit}$ and $+t_{limit}$ ($\nu = n - 1$) is the same probability that the population mean stick length μ will be between 5.5 cm and 6.5 cm, as claimed by the vendor.

Our first step, then, is to calculate t_{limit}, the scaled deviation that represents the maximum deviation of mean stick length.

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}} \tag{2.32}$$

Scaled maximum deviation

$$t_{limit} = \frac{\text{(maximum deviation)}}{s/\sqrt{n}} \tag{2.33}$$

$$= \frac{(0.5\text{ cm})}{0.70\text{ cm}/\sqrt{25}}$$

$$= 3.57142 \quad \text{(unitless, extra digits shown)}$$

The sampling distribution of the mean is the Student's t distribution with $(n - 1)$ degrees of freedom. The probability that $-t_{limit} \leq t \leq t_{limit}$ is given by the integral in Equation 2.12, with the pdf function being the pdf of the Student's t distribution and with the limits given by $\pm t_{limit}$ and for $\nu = (n - 1) = 24$ (Figure 2.10).

Probability that scaled deviation t is between $-t_{limit}$ and t_{limit}

$$\boxed{\Pr = \int_{-t_{limit}}^{+t_{limit}} f(t'; n-1)dt'} \tag{2.34}$$

$$= \int_{-3.57142}^{3.57142} f(t'; 24)dt' \tag{2.35}$$

This integral over the pdf of the Student's t distribution, and hence the probability we seek, is readily calculated in Excel, as we now discuss.

Carrying out integrals of the Student's t probability density distribution is a very common calculation in statistics; Excel has built-in functions that evaluate integrals of $f(t, \nu)$ for the Student's t distribution, and we discuss now how to use these to evaluate Equation 2.35. The Excel function T.DIST.2T($t_{limit}, n-1$) is called the two-tailed cumulative probability distribution of the Student's t distribution. For the Student's t distribution, the function T.DIST.2T($t_{limit}, n-1$) gives the area underneath the two probability-density tails that are outside the interval in which we are interested (Figure 2.10):

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

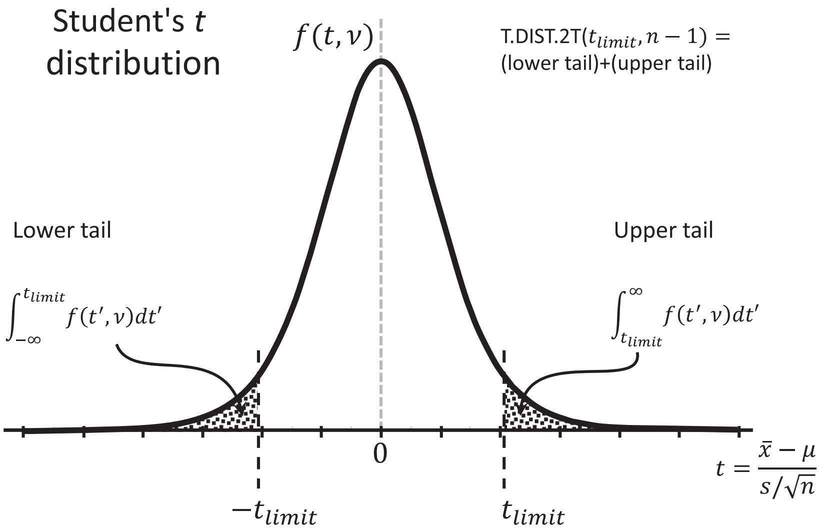

Figure 2.10 The integral under the Student's t distribution pdf from $-\infty$ to $+\infty$ is 1; if we integrate from $-t_{limit}$ to t_{limit} we leave behind two tails, each with the same amount of probability (since the distribution is symmetric). The area under the curve between $\pm t_{limit}$ is just the area in the two tails subtracted from the total area, which is 1.

$$\text{T.DIST.2T}(t_{limit}, n-1) \equiv \int_{-\infty}^{-t_{limit}} f(t'; n-1)dt' + \int_{t_{limit}}^{\infty} f(t'; n-1)dt' \tag{2.36}$$

Thus, the probability we seek in Equation 2.35 is the difference between the total area under the curve, which is equal to 1, ($\int_{-\infty}^{\infty} f(t', n-1)dt' = 1$), and the value yielded by the Excel two-tailed function.

$$\begin{array}{c}\text{Probability that}\\ \text{scaled deviation } t \text{ is}\\ \text{between } -t_{limit} \text{ and } t_{limit}\end{array} \equiv \Pr[-t_{limit} \le t \le t_{limit}] = \int_{-t_{limit}}^{+t_{limit}} f(t'; n-1)dt'$$

$$\boxed{\Pr[-t_{limit} \le t \le t_{limit}] = 1 - \text{T.DIST.2T}(t_{limit}, n-1)} \tag{2.37}$$

The availability of the function T.DIST.2T() in Excel makes calculating the probability we seek a matter of entering a simple formula into Excel (see

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Appendix D for the equivalent MATLAB command.). For the sample of stick length $t_{limit} = 3.57142$ and $\nu = (n-1) = 24$:

$$\begin{array}{c} \text{Probability that} \\ t \text{ is between} \\ -t_{limit} \text{ and } t_{limit} \end{array} = 1 - \text{T.DIST.2T}(3.57142, 24)$$

$$= 0.998456 = 99.8\% \tag{2.38}$$

Based on the sample obtained ($n = 25; \bar{x} = 6.0$ cm; $s = 0.70$ cm) and the target maximum deviation of 0.5 cm, we are 99.8% confident that the true mean stick length of the vendor's supply is in the range 6.0 ± 0.5 cm. The vendor has correctly characterized his collection of sticks.

Note that although the true mean stick length is well characterized by the sample mean, the sample standard deviation is relatively large ($s = 0.70$ cm). Customers should expect the lengths of the sticks to vary.[11] If customers need to have precise stick lengths, they need to consider both the mean and the standard deviation of the vendor's stock.

Example 2.5 applies the Student's t distribution to the problem of determining the number of significant figures to associate with a measurement of density. Appendix C contains a list of Excel functions that are useful for error-limit calculations.

Example 2.5: Play with Student's *t*: sig figs on measured fluid density. *A team of engineering students obtained 10 replicate determinations of the density of a 20 wt% aqueous sugar solution:* $\rho_i (g/cm^3) = 1.0843$, 1.06837, 1.07047, 1.0635, 1.09398, 1.0879, 1.07873, 1.05692, 1.07584, 1.07587 *(extra digits from the calculator are reported to avoid downstream round-off error). Given the variability of the measurements, how may significant figures should we report in our answer?*

Solution: The value of the mean density that we calculate from the students' data is $\bar{x} = \bar{\rho} = 1.075588$ g/cm^3, but the measurements ranged from 1.05692 to 1.09398 g/cm^3, and the standard deviation of the sample set is $s = 0.011297$ g/cm^3 (extra digits shown). When calculating the average density from the data, the computer gives 32 digits, but clearly we cannot justify that degree of precision in our reported answer, given the variability of the data.

[11] In fact, for this vendor's stock, the probability of actually having the length of a stick chosen at random fall between 5.5 and 6.5 cm is only 51%; see Problem 2.32. We know the mean of the stick population, with high confidence, to be between 5.5 and 6.5 cm, but the distribution of stick lengths leading to that well-characterized mean is broad.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

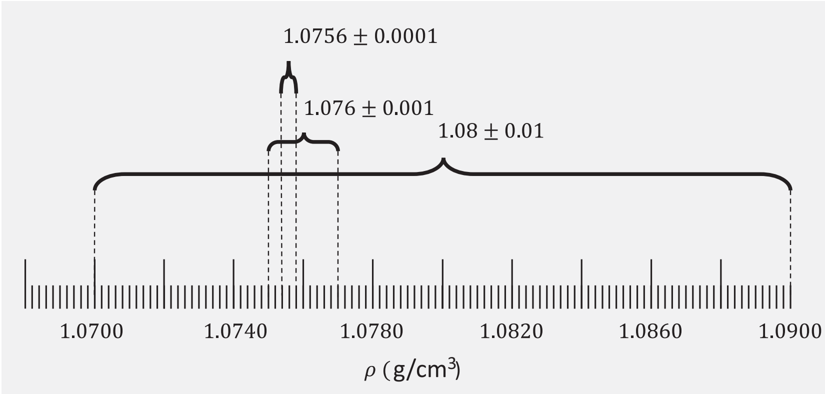

Figure 2.11 Error limits shown correspond to density reported to five sig figs (narrowest curly brackets), four sig figs, and three sig figs. When we specify an answer within the significant figures convention, it implies that the true value will be found in the range created by toggling the least certain digit by 1. When we specify fewer significant figures, we are presenting a broader range and indicating we are less certain of the true value.

Since the question concerns significant figures, can we perhaps apply the sig-figs rules from Appendix B? Unfortunately, these rules are only applicable when values of known precision are combined or manipulated. This is not our current circumstance; rather, the uncertainty in density is due to the sample-to-sample variability of the data (replicate variability). We must determine the uncertainty from the variability of the data and subsequently assign the correct number of significant figures.

We can address the question in this example by returning to the fundamental meaning of the sig-figs convention and thinking in terms of deviation $(\bar{x} - \mu)$. The sig figs idea is that the last digit retained is uncertain by plus or minus one digit. Another way of expressing the question of this example is as follows: what are the probabilities that the true value of the density is found in the following intervals (Figure 2.11):

$$\text{true value} \stackrel{?}{=} 1.0756 \pm 0.0001 \text{ g/cm}^3 \text{ (5 sig figs)}$$
$$\text{true value} \stackrel{?}{=} 1.076 \pm 0.001 \text{ g/cm}^3 \text{ (4 sig figs)}$$
$$\text{true value} \stackrel{?}{=} 1.08 \pm 0.01 \text{ g/cm}^3 \text{ (3 sig figs)}$$

The three $\pm$ values are three possible values of maximum deviation $(\bar{x}-\mu)_{max}$. The sampling distribution of the sample mean [the Student's t distribution, $\nu = (n-1)$] allows us to assess the probability that the true value of the

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

variable (in this case the solution density) is found in a chosen interval of scaled deviation t; we can calculate the probability for each of the potential intervals shown above, and whichever answer gives us an acceptable probability is the number of sig figs we report.

The true value of the density is unknown, but we can write the deviation between the estimate (sample mean $\bar{x}$) and the true μ as $(\bar{x} - \mu)$, which for five significant figures would have to be equal to no more than 0.0001 g/cm^3.

Maximum deviation between the estimate and the true value of ρ for 5 sig figs
$$\left|(\bar{x} - \mu)_{max}\right| = 0.0001 \text{ g/cm}^3$$

The variable t in the Student's t distribution is the deviation $(\bar{x} - \mu)$ expressed in units of replicate standard error $s/\sqrt{n}$, which is based on sample properties (sample standard deviation s, number of samples n).

Scaled deviation
$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$$

We know the replicate standard error $s/\sqrt{n}$ from the dataset, and thus we can determine the value of the scaled deviation t_{limit} such that the dimensionless interval between $-t_{limit}$ and t_{limit} corresponds to a maximum deviation of 0.0001 g/cm^3.

Scaled maximum deviation
$$t_{limit} = \frac{\text{max deviation}}{s/\sqrt{n}} = \frac{(\bar{x} - \mu)_{max}}{s/\sqrt{n}} \tag{2.39}$$

$$t_{limit,10^{-4}} = \frac{0.0001 \text{ g/cm}^3}{0.01129736 \text{ g/cm}^3/\sqrt{10}} = 0.02799 \text{ (unitless)}$$

The probability that, when we take a sample, the observed scaled deviation t will be in the interval $-t_{limit} \leq t \leq t_{limit}$ is given by the area under the pdf of the Student's t distribution with $\nu = (n - 1) = 9$ in the interval between $-t_{limit}$ and t_{limit}.

$$\text{Pr} = \int_{-t_{limit}}^{t_{limit}} f(t', \nu)dt' \tag{2.40}$$

Using the same Excel function introduced in Example 2.4, and for the three sig-figs intervals under consideration [corresponding to maximum deviations $(\bar{x} - \mu)$ of 10^{-4}, 10^{-3}, and 10^{-2} g/cm^3], we obtain the following probabilities.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

For five significant figures:

$$t_{limit,10^{-4}} = \frac{(\bar{x}-\mu)_{max}}{s/\sqrt{n}}$$
$$= \frac{0.0001}{3.5725 \times 10^{-3}} = 0.02799$$

$$\int_{-0.02799}^{0.02799} f(t';9)dt' = 1 - \text{T.DIST.2T}(0.02799, 9) = \boxed{2\% \text{ (5 sig figs)}}$$

For four significant figures:

$$t_{limit,10^{-3}} = \frac{0.001}{3.5725 \times 10^{-3}} = 0.2799$$

$$\int_{-0.2799}^{0.2799} f(t';9)dt' = 1 - \text{T.DIST.2T}(0.2799, 9) = \boxed{21\% \text{ (4 sig figs)}}$$

For three significant figures:

$$t_{limit,10^{-2}} = \frac{0.01}{3.5725 \times 10^{-3}} = 2.799$$

$$\int_{-2.799}^{2.799} f(t';9)dt' = 1 - \text{T.DIST.2T}(2.799, 9) = \boxed{98\% \text{ (3 sig figs)}}$$

The calculation shows us that intervals associated with five significant figures ($\pm 10^{-4}$ g/cm^3, only 2% confidence of capturing the true value; review Figure 2.7) and four significant figures ($\pm 10^{-3}$ g/cm^3, just 21% confidence) are not justified; based on the variability of the data, we should report no more than three significant figures or expect a deviation of at least $\pm 10^{-2}$ g/cm^3 if we want to be reasonably sure that the reported interval includes the true value of the density, given the variability of the observations. Specifically, we can be 98% confident that the true value of the density is within the interval associated with reporting three significant figures. If we choose to report four significant figures, we are taking a substantial risk, as the sample statistics imply that we should only have 21% confidence that we will bracket the true value of the density with this narrower choice.

From Examples 2.4 and 2.5, we see that we can answer some interesting and practical questions with the Student's t distribution and Excel. While sig-figs rules are helpful when manipulating quantities of known uncertainty, sampling gives us direct access to the variability of a quantity. The Excel (or MATLAB) functions make using the statistics fast and easy.

At the beginning of Section 2.2.3 we defined our restated goal as

$$\begin{array}{c}\text{Restated goal: to determine error limits}\\ \text{on an estimated value that, 95\%}\\ \text{of the time, captures the true value of } x\end{array} \qquad \text{Answer} = (\text{value}) \pm (\text{error limits}) \tag{2.41}$$

We have made progress on this goal. Based on a sample $(n, \bar{x}, s)$, we now know how to calculate the probability of finding the true mean μ in a chosen interval. The steps are given here.

Calculate the likelihood of finding the true mean μ in a chosen interval, based on a sample of size n with mean $\bar{x}$ and standard deviation s:

1. Choose the magnitude of the maximum deviation $|(\bar{x} - \mu)_{max}|$. This determines the chosen interval for the error limits.
2. Calculate the maximum scaled deviation t_{limit} from $|(\bar{x} - \mu)_{max}|$, the definition of t, and the sample properties n and s (Equation 2.39, repeated here).

$$\begin{array}{c}\text{Scaled}\\ \text{maximum}\\ \text{deviation}\end{array} \qquad t_{limit} = \frac{\text{max deviation}}{s/\sqrt{n}} = \frac{(\bar{x} - \mu)_{max}}{s/\sqrt{n}} \qquad \text{(Equation 2.39)}$$

3. Use Equation 2.34 (repeated here) to calculate the probability that the observed deviation $|(\bar{x} - \mu)|$ is no larger than the chosen value of maximum deviation.

$$\begin{array}{c}\text{Probability that}\\ t \text{ is between}\\ -t_{limit} \text{ and } t_{limit}\end{array} = \int_{-t_{limit}}^{+t_{limit}} f(t'; n-1)\,dt' \qquad \text{(Equation 2.34)}$$

$$= 1 - \text{T.DIST.2T}(t_{limit}, n-1)$$

4. Report the answer for the predicted mean as $\bar{x} \pm$ (max deviation) at a confidence of (result).

The definition of the scaled sampling deviation t and the identification of the Student's t distribution (with $\nu = n - 1$) as the appropriate probability density distribution for the sampling distribution of the mean, when the standard deviation is unknown, are the advances that make these error-limit and probability calculations possible.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Calculating the probability for a chosen maximum deviation clarifies error-limit and sig-figs choices, but choosing the maximum deviation (step 1) is not always the preferred way of addressing error questions. We would prefer to choose the confidence level with which we are comfortable and turn the problem around and calculate the limits $\pm t_{limit}$ that correspond to the chosen confidence level. We show how to do this in Section 2.3.2. The probability we choose is 95%, and the interval calculated, when expressed in terms of error limits on $\bar{x}$, is called the 95% confidence interval of the mean. In the next section we discuss the issue of how to back-calculate t_{limit} from sample sets to obtain 95% confidence intervals.

We have a final comment on the validity of using the Student's t distribution for estimating sampling properties of the mean. The use of the Student's t distribution is based on the assumption that the underlying distribution of x is the normal distribution with unknown standard deviation, but rigorous calculations show that even if the underlying distribution is not normal, if it is at least a centrally peaked distribution, we may continue to use the Student's t distribution as the sampling distribution of the sample mean [5, 38].

2.3.2 Confidence Intervals of the Mean

The sig-figs convention, based on plus/minus "1" in the last digit, is a coarse expression of error limits that we have seen may not precisely reflect the likely uncertainty in measurements. In Example 2.5 the four significant figures choice for average density (that is, chosen maximum deviation of $\pm 10^{-3}$ g/cm^3) gave a too low amount of confidence at 21% (the true value is captured within this range only about 1 in 5 times that a sample is processed), but the three significant figures choice ($\pm 10^{-2}$ g/cm^3) forced us to the perhaps too conservative 98% confidence level. The jump from 21% confidence to 98% confidence was rather abrupt, and it was forced by thinking of error limits as having to be plus/minus "1" of a decimal place. It may make more sense to choose our confidence level and let the plus/minus increment be whatever corresponds to that confidence level. We explore this approach in the next example.

Example 2.6: Error limits driven by confidence level: measured fluid density, revisited. *A team of engineering students obtained 10 replicate determinations of the density of a 20 wt% aqueous sugar solution (see Example 2.5 for the data). With 95% confidence and based on their data, what is the value of the density along with appropriate error limits on the determined value? (Answer is a range.)*

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Solution: The calculation of sample mean density $\bar{\rho}$ and sample standard deivation s from the data proceeds as in Example 2.5, and the remaining problem is to construct the appropriate error limits such that the probability of capturing the true value of density is 95%. As with the calculations in Example 2.5, we arrive at a confidence value by integrating the probability density function of the sampling distribution between $-t_{limit}$ and t_{limit}.

$$\Pr[-t_{limit} \leq t \leq t_{limit}] = \int_{-t_{limit}}^{t_{limit}} f(t';n-1)dt' \tag{2.42}$$

where $f(t;n-1)$ is the pdf of the Student's t distribution with $(n-1)$ degrees of freedom. Previously we chose the maximum deviation $|(\bar{x}-\mu)_{max}|$, determined t_{limit} from the maximum deviation and properties of the sample, and calculated the probability of observing that deviation from the Student's t distribution. For the current example, the desired confidence is given as 95%; $\Pr = 0.95$ is thus the value of the integral in Equation 2.42. What is not known in this case is the value of the scaled deviation t_{limit} to be used in the limits of the integration so as to obtain that chosen value of the probability.

$$\Pr[? \leq t \leq ?] = 0.95 = \int_{?}^{?} f(t';n-1)dt' \tag{2.43}$$

To determine the 95% probability limits, we must back-calculate t_{limit} from the value of the integral (that is, 0.95) so that Equation 2.43 holds. Once we know t_{limit}, we can write the range in which we expect to find the true value of the density as follows:

$$\begin{array}{cc} \begin{array}{c}\text{Maximum}\\ \text{scaled}\\ \text{deviation}\end{array} & t_{limit} = \dfrac{|(\bar{x}-\mu)|_{max}}{s/\sqrt{n}} \\ \begin{array}{c}\text{Maximum deviation}\\ \text{(error limits)}\end{array} & \pm(\bar{x}-\mu) = \pm t_{limit}\left(\dfrac{s}{\sqrt{n}}\right) \end{array} \tag{2.44}$$

Solving Equation 2.44 for μ, we expect, with 95% confidence, the true mean μ will be in this interval:

$$\mu = \bar{x} \pm t_{limit}\frac{s}{\sqrt{n}} \tag{2.45}$$

Excel performs the inversion we seek (Equation 2.43) with its function T.INV.2T(α,$n-1$) (see also Appendices C and D). The significance level α is defined as 1 minus the confidence level (probability) sought.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$$\text{Significance level} \quad \boxed{\alpha \equiv 1 - \mathrm{Pr}} \tag{2.46}$$

For 95% confidence, $\alpha = 0.05$. The parameter $\nu = (n - 1)$ is the number of degrees of freedom of the sampling process. The general integral of interest in this type of problem is

$$\begin{array}{c}\text{Probability that} \\ t \text{ is between} \\ -t_{limit} \text{ and } t_{limit}\end{array} \quad \mathrm{Pr}[-t_{limit} \leq t \leq t_{limit}]$$

$$\boxed{\mathrm{Pr} = (1 - \alpha) = \int_{-t_{limit}}^{t_{limit}} f(t', \nu) dt'} \tag{2.47}$$

The back-calculation in Equation 2.47 of t_{limit} from known α and $\nu = (n - 1)$ is performed by Excel or MATLAB:

$$\begin{array}{c}\text{Calculate limits } \pm t_{limit} \\ \text{from Equation 2.47} \\ \text{for set significance } \alpha \\ \text{(Excel)}\end{array} \quad \boxed{t_{limit} = t_{\frac{\alpha}{2},\nu} = \mathrm{T.INV.2T}(\alpha, \nu)} \tag{2.48}$$

$$\text{(MATLAB)} \quad \boxed{t_{limit} = t_{\frac{\alpha}{2},\nu} = -\mathrm{tinv}\left(\frac{\alpha}{2}, \nu\right)} \tag{2.49}$$

We explain the subscript "$\alpha/2$" nomenclature in the discussion that follows.

The value of t_{limit} we seek is the value that will give an area of $(1 - \alpha)$ when $f(t', n - 1)$ is integrated between $-t_{limit}$ and t_{limit} (Figure 2.12). There are two pdf "tails" containing excluded areas: one between $-\infty$ and $-t_{limit}$ and one between t_{limit} and ∞ (compare with Equation 2.36). The total area in the two tails is α. In the Excel function T.INV.2T() (2T="two tailed"), one variable we specify is the total amount of probability in the two tails (α); the other variable is the number of degrees of freedom $\nu = (n - 1)$. The area under $f(t; \nu)$ below $-t_{limit}$ is $\alpha/2$; thus the convention is to write $t_{limit} = t_{\frac{\alpha}{2},\nu}$.

We now apply this calculation to the current problem of fluid density error limits. For 95% confidence ($\alpha = 0.05$), with $\nu = (n - 1) = 9$ degrees of freedom, the t_{limit} we obtain is:

$$\begin{aligned} t_{limit} = t_{0.025,9} &= \mathrm{T.INV.2T}(0.05, 9) \\ &= 2.262157 \end{aligned}$$

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

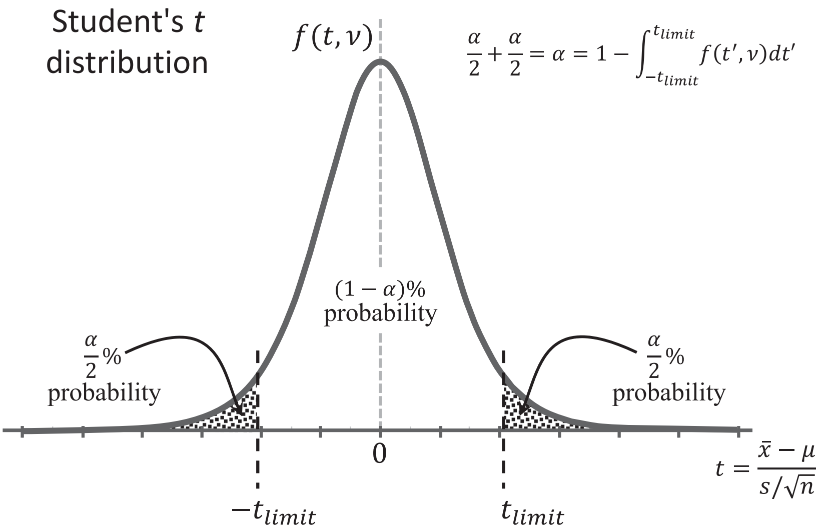

Figure 2.12 For confidence intervals, we are interested in knowing how far out toward the tails we need to go to capture $(1-\alpha)\%$ of the probability between the limits $-t_{limit}$ and t_{limit}. The small probability that resides in the two tails represents improbable observed values of the sample mean.

Once again the calculation is reduced to a simple function call in Excel or MATLAB. From t_{limit} we now calculate the range in which we expect to find the true value of the density.

Maximum scaled deviation
$$t_{limit}=\frac{(\bar{x}-\mu)_{max}}{s/\sqrt{n}}$$

$$\mu=\bar{x}\pm t_{limit}\frac{s}{\sqrt{n}}$$

$$=1.075588\pm 2.262157\left(\frac{0.01129736}{\sqrt{10}}\right)$$

20 wt% solution density (95% confidence)
$$=\boxed{1.076\pm 0.008\ \text{g/cm}^3} \quad (2.50)$$

This interval is shown in Figure 2.13 along with the sig figs–based error limits from Example 2.5. The 95% confidence interval is more precise than the broad, three sig-figs error limits, while still corresponding to a very reasonable (and known) confidence level, 95%.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

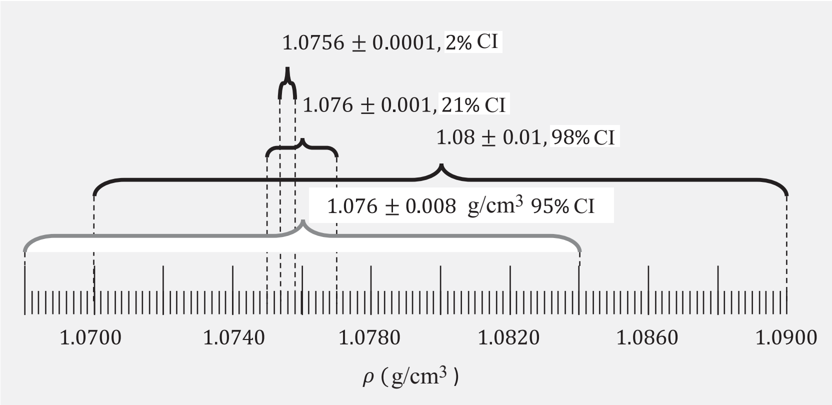

Figure 2.13 Confidence intervals from Examples 2.5 and 2.6. The confidence level varies depending on the error limits chosen. With 95% confidence intervals, we specify a confidence level and calculate the error limits.

Note that in writing our answer in Equation 2.50, we retain only one digit on error. Also, the error is in the third digit after the decimal, and this uncertainty determines how many digits we report for the density (we usually keep only one uncertain digit). For more on significant figures when writing error limits (including exceptions to the one-uncertain-digit rule), see Section 2.4.

The process followed in Example 2.6 is a statistical way of knowing the stochastic "give and take" amounts we mentioned earlier in the chapter. A common choice is to report the amount of "give and take" that will, with 95% confidence, make your estimate right: "right" means we are 95% confident that the calculated range captures the true value of the mean of the distribution, μ. The range that, with 95% confidence, includes the true value of the mean is called a *95% confidence interval of the mean* (Figure 2.14).

$$\begin{array}{c}\text{95\% confidence interval}\\ \text{of the mean}\end{array} = \left(\begin{array}{c}\text{range that, with 95\% confidence,}\\ \text{contains } \mu\text{, the true mean}\\ \text{of the underlying distribution}\end{array}\right) \qquad (2.51)$$

The 95% confidence interval[12] of the mean is the usual way to determine error limits when only random errors are present. For a stochastic variable

[12] If we create a large number of 95% confidence intervals from different samples, 95% of them will contain the true mean μ. The range for any *one sample*, however, either does or does not contain μ. Thus, the probability that any one confidence interval contains the true mean μ is not 95%: It is either one or zero. Rather, we are 95% confident that the mean is in that interval.

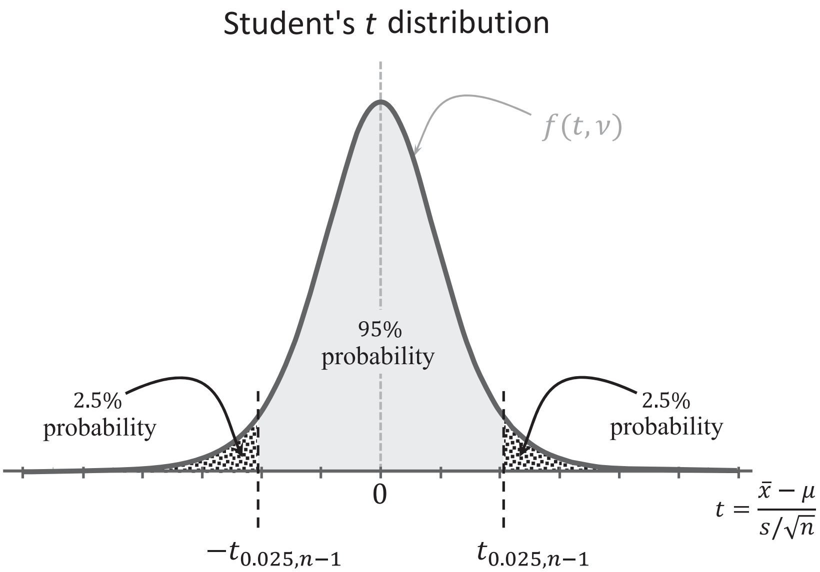

Figure 2.14 The Student's t probability distribution, $f(t;\nu)$, is used to construct 95% confidence intervals on the mean. The central 95% region represents the most likely values we will observe for the sample mean in a sample of size n. The 5% of the probability that resides in the two tails represents improbable observed values of sample mean – improbable, but not impossible. They are observed about 5% of the time (they will be observed 1 in 20 times that a sample of size n is tested).

x sampled n times with sample mean $\bar{x}$ and sample variance s^2, the 95% confidence interval of the mean is calculated as follows:

95% confidence interval of the mean $\bar{x}$ (replicate error only)

$$\boxed{\mu = \bar{x} \pm t_{0.025,n-1} e_{s,random}} \tag{2.52}$$

$$e_{s,random} \equiv \frac{s}{\sqrt{n}}$$

$$\begin{aligned} t_{0.025,n-1} &= \text{T.INV.2T}(0.05, n-1) && \text{Excel} \\ &= -tinv(0.025, n-1) && \text{MATLAB} \end{aligned}$$

where $e_{s,random} = \frac{s}{\sqrt{n}}$ is the standard error on the mean of x and $t_{0.025,n-1}$ is a value associated with the Student's t distribution that ensures that 95% of the probability in the distribution is captured between the limits given in Equation 2.52. Three-digit values of $t_{0.025,n-1}$ are given in Table 2.3 as a function of sample size n; these values are calculated more precisely with

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Table 2.3. *The Student's t distribution approximate values of* $t_{0.025,n-1}$ *for use in constructing 95% confidence intervals for samples of size n. The numbers in* **bold** *are equal to "***2***" to one digit. Accurate values of* $|t_{0.025,n-1}|$ *may be calculated with Excel's function call T.INV.2T*$(0.05, n-1)$ *or with* $-tinv(0.025, n-1)$ *with MATLAB.*

n	2	3	4	5	6	7	8	9	10	20	50	100	∞
$n-1$	1	2	3	4	5	6	7	8	9	19	49	99	∞
$t_{0.025,n-1}$	12.71	4.30	3.18	2.78	2.57	**2.45**	**2.36**	**2.31**	**2.26**	**2.09**	**2.01**	**1.98**	**1.96**

Excel using the function call T.INV.2T(0.05,$n-1$).[13] Note that for large n, $t_{0.025,n-1}$ approaches a value of about 2.

For practice, we now apply the Student's t distribution and Equation 2.52 to determine 95% confidence intervals on the commuting-time measurements in Examples 2.1 and 2.2.

Example 2.7: Commuting time, revisited. *Over the course of a year, Eun Young took 20 measurements of her commuting time under all kinds of conditions. Ten of her observations are shown in Table 2.1 (Example 2.1) and 10 other measurements taken using the same stopwatch and the same timing protocols are shown in Table 2.2 (Example 2.2). Calculate an estimate of Eun Young's time-to-commute first using the first dataset and then using all 20 data points. Express your answers for the estimates of commuting time with error limits consisting of 95% confidence intervals of the mean.*

Solution: Using the first set of $n = 10$ data points found in Table 2.1, we employ Equation 2.52 with $\bar{x} = 28$ min, $s = 11$ min, and $t_{0.025,9} = 2.26$ [T.INV.2T(0.05, 9)]. The 95% confidence interval of the mean is:

95% confidence interval
of the mean
Set 1; $n = 10$

$$\mu = \bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

$$= 28 \pm (2.26)\frac{11}{\sqrt{10}}$$

$$= 28 \pm 8 \text{ min}$$

Using all $n = 20$ data points found in Tables 2.1 and 2.2, we calculate $\bar{x} = 30$ min, $s = 10$ min, and from T.INV.2T(0.05, 19) we find $t_{0.025,19} = 2.09$. The 95% confidence interval of the mean for these data is as follows:

[13] In this formula, the $\alpha = 0.05$ corresponds to the 95% confidence level selected; for a 99% confidence level, replace 0.05 with 0.01.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

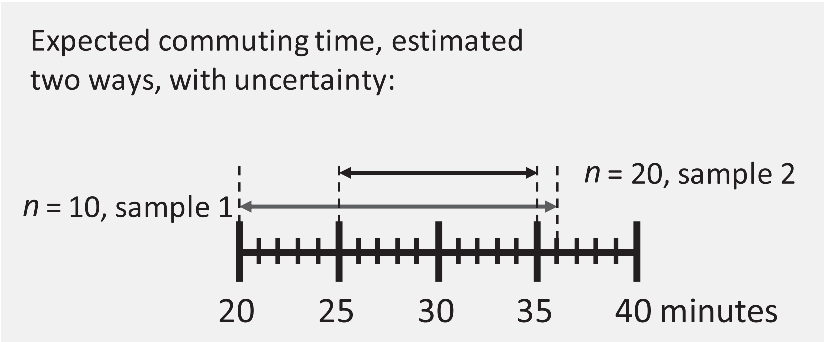

Figure 2.15 The first dataset with $n = 10$ produced a wider 95% confidence interval than the second dataset with $n = 20$. The two predictions overlap and thus are consistent in their predictions for the true value of the mean of the variable, μ.

95% confidence interval of the mean Combined set; $n = 20$

$$\begin{aligned} \mu &= \bar{x} \pm t_{0.025,n-1} \frac{s}{\sqrt{n}} \\ &= 30 \pm (2.09) \frac{10}{\sqrt{20}} \\ &= 30 \pm 5 \text{ min} \end{aligned}$$

Taking a sample size of 10 yielded an estimate of between 20 and 36 min for the mean commuting time (Figure 2.15, sample 1). Taking a sample of size 20 yielded an estimate that agrees that the true mean is between those two numbers, but that narrows the prediction to be between 20 and 30 min at the same level of confidence (Figure 2.15, sample 2). Because of the $\sqrt{n}$ in the denominator of Equation 2.52, increasing n narrows the confidence interval and more precisely predicts the mean commuting time. Note that in neither case could we rely on sig-figs rules to express the appropriate precision. The true uncertainty is reflected in the variability of the data and had to be determined through sampling.

As Example 2.7 shows, when we have data replicates, it is straightforward to use software to calculate an expected value of the mean and a 95% confidence interval for that value. For predictions of the sample mean, obtaining more replicates narrows the range of the confidence interval (makes us more precise in our estimate of the true value of the mean at the same level of confidence).

$$\text{Predicted value of the mean:} \quad \mu = \bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$$
$$= \lim_{n\to\infty}\left(\bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}\right)$$
$$= \bar{x} \tag{2.53}$$

The 95% confidence interval quantifies the "give or take" that should be associated with an estimate of the mean of a stochastic quantity such as commuting time.

Note that the current discussion has focused on error limits on the mean. As n increases to infinity, we become certain of the mean. One example we have discussed is the typical commuting time for Eun Young – the answer for typical commuting time is the mean commuting time and the error limits are the error limits on the mean. A different but related question is, what duration do we expect for Eun Young's commute *tomorrow*? This is a question about a "next" value for a variable. The answer to this question is also the mean of the dataset, but the variability is different – the variability expected in tomorrow's commuting time is much larger than the variability expected for the mean. More replicates refine our estimate of the mean, but larger samples do not make Eun Young's commute less variable from day to day. To address a "next" value of a variable, we use a prediction interval rather than a confidence interval; see Section 2.3.3 and Example 2.15.

In Example 2.8 we carry out another 95% confidence interval calculation for laboratory data.

Example 2.8: Density from laboratory replicates. *Ten student groups measure the density of Blue Fluid 175 following the same technique: using an analytical balance they weigh full and empty* 10.00 ± 0.04 *ml pycnometers.*[14] *Their raw results for density are given in Table 2.4. Note that they used ten different pycnometers. With 95% confidence, what is the density of the Blue Fluid 175 implied by these data? Assume only random errors are present.*

Solution: We use Excel to calculate the sample mean and standard deviation of the data in Table 2.4: the mean of the dataset is $\bar{\rho} = 1.73439$ g/cm^3 and the standard deviation is $s = 0.00485$ g/cm^3 (excess digits have been retained to avoid round-off error in downstream calculations). According to our understanding of the distribution of random events, with 95% confidence, the true value of a variable x measured n times is within the range $\bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$

[14] For more on pycnometers, see Figure 3.13 and Examples 3.8 and 4.4.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Table 2.4. *Raw data replicates of the room-temperature density of Blue Fluid 175 obtained by ten student groups.*

Index	Density, g/cm^3
1	1.7375
2	1.7272
3	1.7374
4	1.7351
5	1.73012
6	1.7377
7	1.7398
8	1.72599
9	1.7354
10	1.7377

(Equation 2.52); the quantity $t_{0.025,n-1}$ is given in Table 2.3 (or calculated from the Student's t distribution with Excel or MATLAB) as a function of n. For the density data in Table 2.4, we calculate:

$$\begin{aligned}
n &= 10 \\
t_{0.025,9} &= 2.262157 \quad \text{[using Excel: T.INV.2T(0.05,9)]} \\
\text{density} &= \bar{x} \pm t_{0.025,9}\frac{s}{\sqrt{n}} \\
&= 1.734391 \pm (2.262157)\left(\frac{0.004853}{\sqrt{10}}\right) \\
\rho_{BF175} &= \boxed{1.734 \pm 0.003 \text{ g/cm}^3 \text{ (95\% CI)}}
\end{aligned}$$

From the data collected, the expected value of the density of the Blue Fluid 175 is 1.734 ± 0.003 g/cm^3 with 95% confidence. We are 95% confident that the true value of the density of the fluid is within this interval. Note that the calculated 95% confidence interval is ± 3 in the last digit (± 0.003 g/cm^3). We cannot express this precise level of uncertainty with significant figures alone. Without the error limits, reporting $\rho = 1.734$ g/cm^3 implies by sig figs that the uncertainty is ± 0.001 ("1" in the last digit), which is overly optimistic.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

If we choose to report the uncertainty as ± 0.001 g/cm^3, we can calculate the confidence we should have in such an answer, as we see in the next example.

Example 2.9: Density from laboratory replicates: tempted by four sig figs. *In Example 2.8 we calculated the 95% confidence interval associated with some measurement replicates of density. If we instead take a shortcut and guess that we may report the final answer to 4 sig figs as $\rho = 1.734\ g/cm^3$, what is the confidence level associated with this answer?*

Solution: Error limits of ± 0.001 g/cm^3 imply a maximum deviation from the true value of $|(\bar{x} - \mu)_{max}| = 0.001$ g/cm^3. To calculate the probability of such a maximum deviation, we integrate the pdf of the Student's t distribution between limits of $\pm t_{limit}$, where t_{limit} is calculated from the maximum deviation (see Example 2.5).

$$\begin{aligned}
\text{Maximum scaled deviation} \qquad t_{limit} &= \frac{|(\bar{x} - \mu)_{max}|}{s/\sqrt{n}} \\
&= \frac{(0.001\ \text{g/cm}^3)}{0.004853\ \text{g/cm}^3/\sqrt{10}} \\
&= 0.651612953 \\
\Pr &= \int_{-t_{limit}}^{t_{limit}} f(t'; n-1)dt' \\
&= 1 - \text{T.DIST.2T}(t_{limit}, n-1) \\
&= 1 - \text{T.DIST.2T}(0.6516129, 9) \\
&= 47\%
\end{aligned}$$

Our confidence is less than 50% that the true value of the density is found in the narrow interval $\rho = 1.734 \pm 0.001$ g/cm^3.

Determining uncertainty of replicates is essential in scientific and engineering practice, so it is worthwhile to become familiar with the basic features of the Student's t distribution. Looking at the values in Table 2.3, we see that for two replicates ($n = 2, \nu = 1$) and considering only one digit of precision, we must bracket almost 13 standard errors of the mean to be 95% sure that we have captured the true mean ($t_{0.025,1} \approx 13$, Figure 2.16). Increasing from $n = 2$ replicates to $n = 3$ replicates, however, reduces the number of standard errors needed for 95% confidence to about $t_{0.025,2} \approx 4$, which is a big improvement. Going to $n = 4$ narrows the confidence interval still further to $t_{0.025,3} \approx 3$ standard errors, which is also approximately the value of $t_{0.025,n-1}$ for $n = 5$ and 6 (to one digit). For $n \geq 7$, the value of $t_{0.025,n-1}$ is approximately 2. From

	$t_{0.025,n-1}$
$n = 2$	≈ 13
$n = 3$	≈ 4
$n = 4$	≈ 3
$n = 5$	≈ 3
$n = 6$	≈ 3
$n = 7$	≈ 2

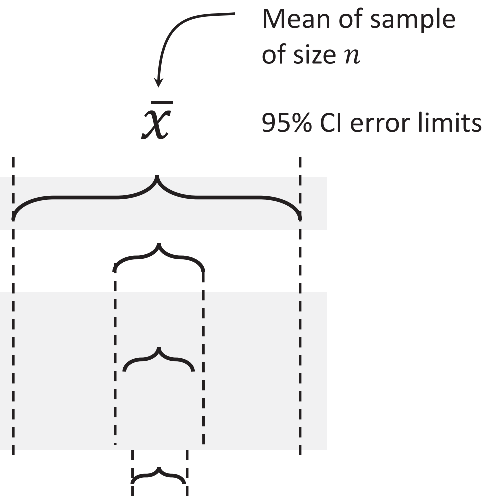

Figure 2.16 The number of standard errors to use when constructing replicate error limits varies strongly as n changes from 2 to 3; the number continues to decrease until approximately $n \approx 7$, when it plateaus at about 2.

the values of $t_{0.025,n-1}$, we can understand why scientists and engineers often take at *least* data triplicates to determine an estimate of a stochastic quantity: we get a large gain in precision when we increase the number of measurements from $n = 2$ to $n = 3$, and less substantial increases when we increase n further, particularly after $n = 7$.

The 95% confidence interval for replicates, given in Equation 2.52, is a widely adopted standard for expressing experimental results. In Appendix A we provide a replicate error worksheet as a reminder of the process used to obtain error limits on measured quantities subject to random error. For $n > 7$, about plus or minus two standard errors ($e_{s,random} = s\sqrt{n}$) corresponds to the 95% confidence interval.

Random error
95% CI of mean ($e_{s,random} = s\sqrt{n}$) $\quad n$ replicates, $\quad \mu = \bar{x} \pm t_{0.025,n-1} e_{s,random} \quad$ (2.54)

$$n > 7 \text{ replicates, } \mu \approx \bar{x} \pm 2e_{s,random}$$

In subsequent chapters we discuss ways to incorporate nonrandom errors into our error limits, and additional error worksheets are discussed there. When expressing 95% confidence intervals with nonrandom errors, we retain the

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

structure of the 95% confidence interval as approximately $\pm 2e_s$ (two standard errors; see Appendix E, Equation E.5), but we use a different standard error that corresponds to the independent combination of replicate, reading, and calibration errors.

Standard errors combine in quadrature

$$\boxed{e^2_{s,cmbd} = e^2_{s,random} + e^2_{s,reading} + e^2_{s,cal}} \tag{2.55}$$

The task of incorporating nonrandom errors into error limits requires us to determine an appropriate standard error e_s for the nonrandom contributions (see Section 1.4).

Nonrandom error
95% CI of estimate
($e_{s,cmbd}$ = combination of random, reading, and calibration error)

$$\boxed{\text{estimate} \approx \bar{x} \pm 2e_{s,cmbd}} \tag{2.56}$$

We pursue these calculations in Chapters 3 and 4.

The examples that follow provide some practice with 95% confidence intervals of the mean and with using the Student's t values from Table 2.3 or from Excel or MATLAB.

Example 2.10: Power of replicates: density from a pycnometer. *Chris and Pat are asked to measure the density of a solution using pycnometers*[15] *and an analytical balance. They are new to using pycnometers, and they each take a measurement; Chris obtains* $\rho_1 = 1.723\ g/cm^3$ *and Pat obtains* $\rho_2 = 1.701\ g/cm^3$*. Chris averages the two results and reports the average along with the replicate error limits. Calculate Chris's answer and error limits* ($n = 2$)*. Pat takes a third measurement and obtains* $\rho_3 = 1.687\ g/cm^3$*. Pat decides to average all three data points (Pat's two and Chris's measurement). Calculate Pat's answer and error limits, assuming that all three measurements are equally valid and affected only by random error. Comment on the effect of taking the third measurement versus reporting the average of a duplicate only.*

Solution: The 95% confidence interval of the mean of data replicates is

$$\text{95\% CI of the mean:} \quad \bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

For Chris's two data points (Chris's measurement and Pat's first data point), the mean is $\bar{\rho} = 1.7120$ g/cm^3 and the standard deviation is $s = 0.0156$ g/cm^3, and thus Chris obtains:

[15] For more on pycnometers, see Figure 3.13 and Examples 3.8 and 4.4.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$$\begin{aligned}
\text{Chris's answer:} \qquad n &= 2 \\
t_{0.025,1} &= 12.706 \\
\text{density} &= 1.7120 \pm (12.706)\left(\frac{0.0156}{\sqrt{2}}\right) \\
&= 1.71 \pm 0.14 \text{ g/cm}^3
\end{aligned}$$

Note that we chose to provide two uncertain digits in our answer since the leading error digit is "1" (see Section 2.4).

For Pat's answer, we use all three data points, calculating: $\bar{\rho} = 1.70367$ g/cm^3 and $s = 0.0181$ g/cm^3. Pat obtains:

$$\begin{aligned}
\text{Pat's answer:} \qquad n &= 3 \\
t_{0.025,2} &= 4.303 \\
\text{density} &= 1.70367 \pm (4.303)\left(\frac{0.0181}{\sqrt{3}}\right) \\
&= 1.70 \pm 0.05 \text{ g/cm}^3
\end{aligned}$$

Because the leading error digit is 5, we provide only one digit on error.

The two answers are compared in Figure 2.17. Chris and Pat obtain similar estimates of the density, but the error limits on Chris's number are quite a bit wider than those on Pat's, since Pat used a data triplicate versus Chris's duplicate. There is a significant increase in precision when we are able to

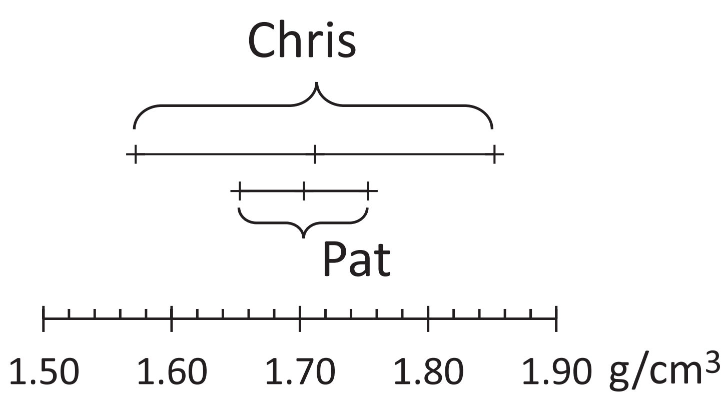

Figure 2.17 Chris and Pat obtain similar estimates of a sample's density (in g/cm^3), but Chris's answer is significantly less precise (wider 95% confidence interval).

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

obtain a data triplicate ($t_{0.025,2} \approx 4.3$ for $n = 3$; see Table 2.3) compared to only having two replicates ($t_{0.025,1} \approx 12.7$ for $n = 2$).

Replication provides the opportunity to obtain precise values of measured quantities. Sometimes, however, there is no opportunity to replicate, as in the next example.

Example 2.11: Estimate density uncertainty without replicates. *We seek to determine the density of a sample of Blue Fluid 175, but there is only a small amount of solution available, and we can make the measurement only once. The technique used is mass by difference with a* 10.00 ± 0.04 *ml pycnometer and an analytic balance (the balance is accurate to* $\pm 10^{-4} g$*). What are the appropriate error limits on this single measurement?*

Solution: We have discussed thus far how to obtain error limits when replicate measurements are available. In the current problem, we do not have any replicates to evaluate, so the techniques of this chapter do not allow us to assess the error limits.

The scenario in this problem is quite common, and we address this question further in Chapter 5. We can estimate error limits of a single data point calculation with a technique called *error propagation*. Error propagation is based on appropriately combining the estimated reading and calibration error of the devices that produce the numbers employed in the calculation. In the current problem, these would be the reading and calibration error of the analytical balance and the calibration error of the pycnometer. For now, we cannot proceed with this problem; we revisit this estimate in Examples 5.1 and 5.4.

Thus far, we have explored the appropriate methods for determining error limits on measurements when random error is present: when random errors are present, we take data replicates and construct 95% confidence intervals around the mean of the sample set. We have also indicated that when more than random error is present, we must estimate reading and calibration error and combine these with replicate error to obtain the correct combined error. In Chapter 3 we turn to determining reading error.

The remaining sections of this chapter address several topics related to replicate measurements as well as some topics related to measurement error in general. First, we discuss *prediction intervals*, which may be used to address problems such as estimating uncertainty in future ("next") data points. Next, we discuss a technique for turning potential systematic errors into more easily handled random errors. Finally, we formally present our convention for reporting significant figures on error. Together these topics complete our quick start on replicate error and set the stage for the discussion of systematic errors, which begins in Chapter 3.

2.3.3 Prediction Intervals for the Next Value of x

In the previous section on confidence intervals for the mean, we saw how to estimate a value for a replicated quantity and how to determine its uncertainty. The uncertainty of the sample mean is proportional to the standard error of replicates, $e_s = s/\sqrt{n}$, which gets smaller as the sample size n gets larger. For large sample sizes, $\lim_{n\longrightarrow\infty} s/\sqrt{n} = 0$ and the 95% confidence interval error limits will be very tight, and the true value of the mean of the underlying distribution of x will be known with a high degree of precision. Note that *it is the true, mean value of the stochastic variable* x, calculated from the mean of samples, that is determined within tighter and tighter error limits.[16] Confidence intervals for the mean do not tell us about error limits for individual measurements. We explore this topic in the next example.

Example 2.12: Using statistics to identify density outliers. *In Example 2.8, we calculated the value of the density of Blue Fluid 175 to be* $\rho_{BF175} = 1.734 \pm 0.003\ g/cm^3$. *The density was determined by calculating the mean of 10 measurements, and the uncertainty in the result was calculated from the 95% confidence interval of the mean. Looking back at the data used to calculate the mean (Table 2.4, plotted in Figure 2.18), we see that several of the original data points are outside of the 95% confidence interval of the mean. Does this indicate that some of the data points are outliers and should be discarded?*

Solution: No, that is not correct. We have calculated the 95% confidence interval *of the mean*. It indicates that we are 95% confident that the true mean value of the Blue Fluid 175 density measurement (the mean of the underlying distribution of density measurements) is in the interval 1.734 ± 0.003 g/cm^3. The confidence interval of the mean gets more and more narrow when we use larger and larger sample sizes; in other words, the more data we obtain, the more precisely we can state the value of the mean (the confidence interval becomes small).

$$\mu = \bar{x} \pm t_{0.025,n-1} \frac{s}{\sqrt{n}} \tag{2.57}$$

$$\text{For } n \text{ large:} \qquad \mu = \lim_{n\longrightarrow\infty} \left(\bar{x} \pm t_{0.025,n-1} \frac{s}{\sqrt{n}} \right) = \bar{x}$$

As n increases, eventually nearly *all* measured data points will lie outside the 95% confidence interval of the mean.

16 This assumes only random error influences the measurement; see Chapters 3 and 4.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

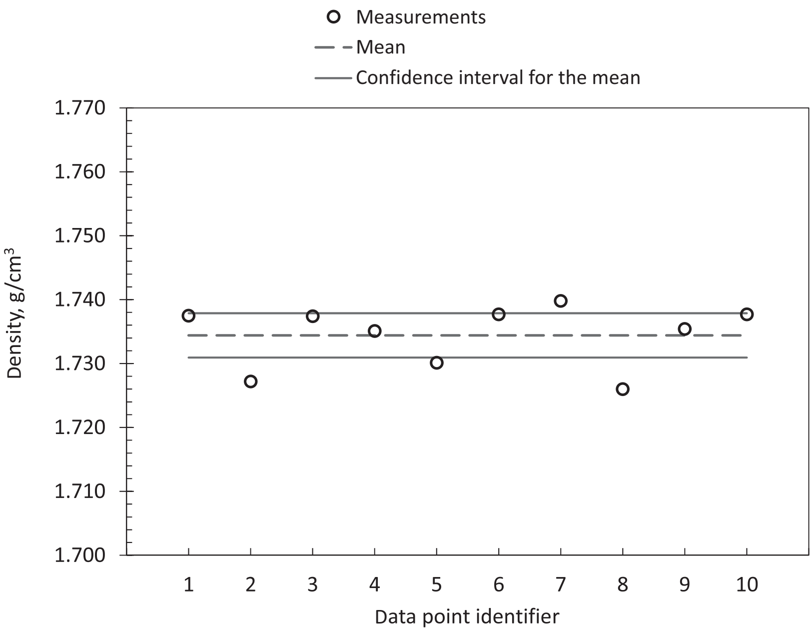

Figure 2.18 Several of the original data points from Example 2.8 fall outside the 95% confidence interval of the mean. The confidence interval of the mean only relates to how well we know the mean; it does not help us evaluate or understand the scatter of the individual observations of the variable.

When we look at the raw data (Figure 2.18), they are scattered, and their scatter reflects the random effects that influenced the values obtained in the individual, noisy measurements. The next (eleventh) measurement of density will also be subject to these random effects, and we expect subsequent measurements to be similarly scattered. The 95% confidence interval *of the mean* does not tell us about the magnitude of these random effects on individual measurements.

> The 95% confidence interval *of the mean* does not tell us about the magnitude of random effects on individual measurements.

The idea that individual data points may be suspect – that is, might be outliers (due to a blunder or other disqualifying circumstances) – can be evaluated with a different tool, the prediction interval. We discuss this approach next.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

If we are interested in expressing an interval that encompasses a prediction of the next likely value of a measured quantity, we construct a *prediction interval*. If the underlying pdf of the variable can be assumed to be approximately normally distributed (meaning the errors follow the normal distribution), then it can be shown with error propagation (see reference [38] and Problem 2.41) that the next data point for x will fall within the following 95% prediction interval of the next value of x [5, 38]:

95% prediction interval for x_{n+1}, the next value of x

$$\boxed{x_{n+1} = \bar{x} \pm t_{0.025,n-1}\, e_{s,next}} \qquad (2.58)$$

$$e_{s,next} \equiv s\sqrt{1+\frac{1}{n}}$$

$$t_{0.025,n-1} = \text{T.INV.2T}(0.05, n-1)$$

The quantity $e_{s,next} = s\sqrt{1+\frac{1}{n}}$ is the standard error for the next value of x. This expression results from a combination (see Chapter 5) of the uncertainty associated with a new data point and the uncertainty associated with the mean $\bar{x}$. Note that this interval shrinks a bit with increasing n due to changes in $t_{0.025,n-1}$ and n (Table 2.3), but for $n \geq 7$, the prediction interval plateaus at approximately $\pm 2s$.

We can practice creating prediction intervals by calculating a prediction interval for the Blue Fluid 175 density data from Example 2.12.

Example 2.13: Prediction interval on students' measurements of fluid density. *Based on the student data on density of Blue Fluid 175 given in Table 2.4, what value do we expect a new student group to obtain for ρ_{BF175}, given that they follow the same experimental procedure? The answer is a range.*

Solution: We assume that the new student group is of comparable ability and attentiveness to procedure as the groups whose data supplied the results in Table 2.4. Thus, the next value obtained will be subject to the same amount of random scatter as shown in the previous data. We can find the answer to the question by calculating the 95% prediction interval of the next value of ρ_{BF175}.

We construct the 95% prediction interval using Equation 2.58 and the data in Table 2.4. To avoid round-off error, extra digits are shown here for use in the intermediate calculations (see Section 2.4). The final answer is given with the appropriate number of significant figures.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$$\begin{bmatrix} \text{95\% PI} \\ \text{of next} \\ \rho_{BF175} \end{bmatrix} = \bar{x} \pm t_{0.025,n-1} s \sqrt{1 + \frac{1}{n}}$$

$$\bar{\rho} = 1.734391 \text{ g/cm}^3$$
$$n = 10$$
$$s = 0.004853 \text{ g/cm}^3$$
$$t_{0.025,9} = 2.262157 \qquad \text{[T.INV.2T(0.05,9)]}$$

$$\begin{bmatrix} \text{95\% PI} \\ \text{of next} \\ \rho_{BF175} \end{bmatrix} = 1.734391 \pm (2.262157)(0.004853)(1.0488)$$

$$= \boxed{1.734 \pm 0.012 \text{ g/cm}^3} \qquad \text{(two uncertain digits)}$$

Our result[17] indicates that, with 95% confidence, a student group will obtain a value between 1.723 and 1.746 g/cm^3.

This prediction interval is wider than the confidence interval of the mean (Figure 2.19). Several individual data points fall outside the 95% confidence interval of the mean, but all of the data points are within the 95% prediction interval for the next value of ρ_{BF175}. None of the data points is particularly unusual, as judged at the 95% confidence level.

An important use for the prediction interval for the next value of x is to determine if new data should be obtained or existing data discarded due to a probable mistake of some sort. The idea is that some experimental outcomes are very unlikely, and if an unlikely value is observed, it is plausible that the data point is a mistake or blunder and should not be used. Example 2.14 considers such an application.

Example 2.14: Evaluating the quality of a new density data point. *One day in lab, ten student groups measured the density of a sample of Blue Fluid 175, obtaining the results shown in Table 2.4. An eleventh student group also measured the density of the same fluid the next day, following the same procedure. The result obtained by Group 11 was $\rho_{BF175} = 1.755$ g/cm^3, which was higher than any value obtained by the first ten groups. Should we suspect that there is some problem with Group 11's result, or is the result within the range we would expect for this measurement protocol?*

[17] Two digits are used in expressing the error limits in this case because the leading error digit is "1"; see Section 2.4.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

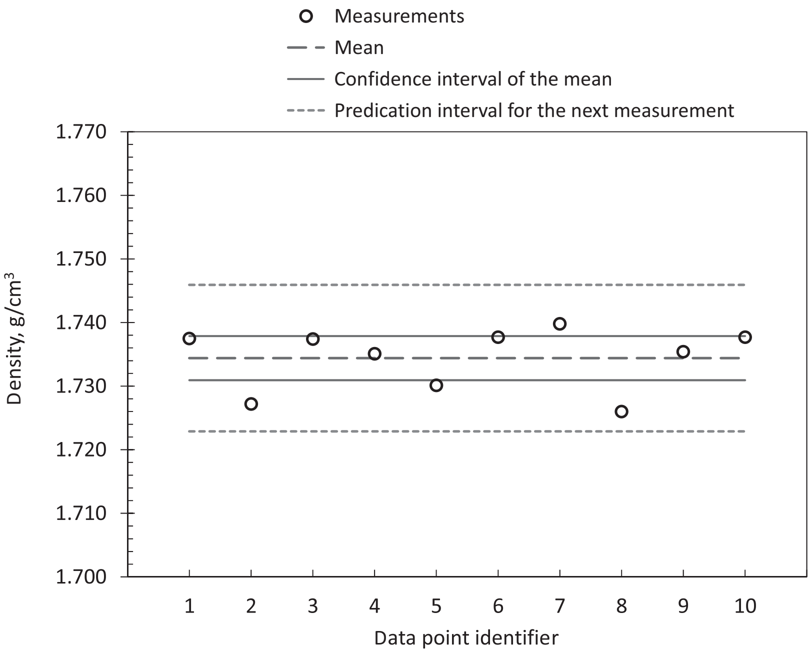

Figure 2.19 For n larger than 3, it is unsurprising that several of the original data points fall outside the 95% confidence interval of the mean, since the confidence interval is constructed with $\pm t_{0.025,n-1}s/\sqrt{n}$ and n is in the denominator. To address the question of how representative any given data point is of the entire set, we construct 95% prediction intervals of the next value of x. All the data points in the example fall within the prediction interval; by definition, if the effects are all random, 95% of the data will fall within the prediction interval.

Solution: We address this question by seeing whether the new result lies within the 95% prediction interval implied by the original data. In Example 2.13 we calculated that prediction interval:

$$\begin{array}{c}\text{Based on}\\ \text{Table 2.4 data}\\ (n=10)\end{array} \quad \begin{bmatrix}\text{95\% PI}\\ \text{of next}\\ \rho_{BF175}\end{bmatrix} = 1.734 \pm 0.012 \text{ g/cm}^3$$

This result tells us that with 95% confidence, the next measured value of ρ_{BF175} will lie in the following interval:

$$1.722 \text{ g/cm}^3 \leq \rho_{BF175} \leq 1.746 \text{ g/cm}^3 \tag{2.59}$$

The value measured by Group 11 was $\bar{\rho} = 1.755$ g/cm^3, which is not within this interval. Thus, we conclude that their result is suspect, since with

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

95% confidence, individual results are expected to be within the prediction interval calculated. Perhaps there was a mistake in the execution of the procedure.

We could also ask a related question: what is the probability that a measurement at least as extreme as this measurement occurs? If the resulting probability is very small, that could be reason for rejecting the point. The posed question may be answered by finding the total probability within the region of the Student's t distribution pdf tail that just includes this rare point; see Problem 2.29.

In Example 2.14 we did not conclude that the examined value is definitely in error; rather, we concluded that the value is suspect. We are using a 95% prediction interval, which means that, on average, one time in 20 we expect a next data point to lie outside of the interval. When working with stochastic variables, we can never be 100% certain that an unlikely outcome is not occurring. What we can do, however, is arrive at a prediction and deliver that prediction along with the level of our confidence, based on the samples examined.

We can use the prediction interval to determine the range of commuting times that Eun Young should expect on future trips, based on her past data (as discussed in Examples 2.1, 2.2, and 2.7). This is a classic question about a next value of a stochastic variable.

Example 2.15: Commuting time, once again revisited. *Over the course of a year, Eun Young took 20 measurements of her commuting time under all kinds of conditions, using the same stopwatch and the same timing protocols; the data are in Tables 2.1 and 2.2. What is Eun Young's most likely commuting time tomorrow? Provide limits that indicate the range of values of commuting time that Eun Young may experience.*

Solution: The most likely amount of time that Eun Young's commute will take is the mean of the 20 observations, with 95% confidence, which we determined in Example 2.7 to be 30 ± 5 min. The error limits on the mean indicate our confidence (at the 95% level) in the value of the mean.

To determine the range of values of commuting time that Eun Young may experience tomorrow ("next"), we need the 95% prediction interval of the next value of the variable. We calculate this with Equation 2.58 with $\alpha = 0.05$.

$$\begin{array}{c}\text{95\% PI}\\ \text{of next}\\ \text{value of } x\end{array} \qquad x_i = \bar{x} \pm t_{0.025,n-1}\, s\sqrt{1+\frac{1}{n}} \tag{2.60}$$

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

For the 20 data points on Eun Young's commuting time, the 95% prediction interval of the mean is:

$$x_i = \bar{x} \pm t_{0.025,n-1}\, s\sqrt{1+\frac{1}{n}}$$

$$\begin{aligned} x &= 29.9 \pm (2.093024)(9.877673)\sqrt{1+\frac{1}{20}} \\ &= 29.9 \pm 21.18476 \\ &= 30 \pm 22 \text{ min} \end{aligned}$$

Thus, Eun Young should expect a commuting time of up to 52 min (95% confidence level).

Replicate sampling, as we see from the discussion here, is a powerful way to estimate the true value of a quantity that we are able to sample. The only limitation on the power of replication is that *replicate error only accounts for random effects*. If nonrandom effects cause measurements to be different from the true value, then no amount of replication will allow us to find the true value of the variable or the proper range of the confidence or prediction intervals.

> If *nonrandom* effects cause measurements to be different from the true value, then no amount of replication will allow us to find the true value of the variable or the proper range of the confidence or prediction intervals.

To see what this dilemma looks like, consider the next example.

Example 2.16: Repeated temperature observations with a digital indicator. *Ian uses a digital temperature indicator equipped with a thermocouple to take the temperature at the surface of a reactor. He records the temperature reading every 5 min for 20 min and the values recorded are 185.1, 185.1, 185.1, 185.1, and 185.1 (all in °C). Assuming there are only random errors in these measurements, what is the temperature at the surface of the reactor? Include the appropriate error limits.*

Solution: Assuming that there are only random errors, we calculate the value of the surface temperature as the average of the five values, and we use the 95% confidence interval of the mean in Equation 2.52 for the uncertainty.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$$
\begin{aligned}
n &= 5 \\
\bar{T} &= 185.1^\circ\mathrm{C} \\
s &= 0.0^\circ\mathrm{C} \\
t_{0.025,4} &= 2.78 \\
\text{surface temperature} &= \bar{T} \pm t_{0.025,4}\left(\frac{s}{\sqrt{n}}\right) \\
&= \boxed{185.1000 \pm 0.0000^\circ\mathrm{C}}
\end{aligned}
$$

According to this calculation, since we obtained five identical measurements, we know the surface temperature of the reactor *exactly*.

Of course, you should be skeptical of this conclusion and already wondering what went wrong with this logic. If that is so, you are correct: something has gone wrong. The sensor used in this analysis gave the same value every time a measurement was taken, and we (correctly) calculated a standard deviation of zero. The conclusion we can draw from the zero standard deviation is that the sensor is very consistent and there is no detectable random error.

It would be wrong to assume that we now know the temperature exactly, however. For one thing, since the indicator only gives temperature to the tenths digit, we cannot tell the difference between 181.12°C and 181.13°C with our indicator. We discuss this sort of error, called reading error, in Chapter 3 (see Problem 3.3). In addition, we have not explored the actual *accuracy* of the temperature indicator. Does the indicator give the actual temperature or does it read a bit high or low? Many home ovens can run high or low due to limitations in sensor accuracy, and good cooks know to adjust for this tendency. Matching a sensor's reading with the true value of the quantity of interest is called *calibration*. Uncertainty due to limitations of calibration is discussed in Chapter 4 (see Problem 4.23 for more on thermocouple accuracy). Both reading error and calibration error are systematic errors, and repeating the measurement will not make these systematic errors visible. We must track down systematic errors by other methods.

In this chapter we have discussed how to account for random errors in measurement. There are also nonrandom effects, and we discuss these beginning in the next chapter. Two common sources of nonrandom error are considered in this book: reading error (Chapter 3) and calibration error (Chapter 4). These nonrandom effects often dominate the uncertainty in experimental results and should not be neglected without investigation. Both reading error and

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

calibration error are systematic errors, and repeating the measurement will not make them visible. Instead, we must track down systematic errors individually.

In the next section we introduce randomization, an experimental technique that helps us ferret out some types of nonrandom error. In Section 2.4 we close with a discussion of the significant figures appropriate for error limits.

2.3.4 Essential Practice: Randomizing Data Acquisition

The methods described in this book allow us to quantify uncertainty. When random errors are present, the values obtained for a measured quantity bounce around, even when all the known factors affecting the quantity are held constant. This makes random errors easy to detect when a measurement is repeated or replicated. As discussed earlier in this chapter, we can quantify random errors by analyzing replicate measurements: once we obtain replicate values, we apply the statistics of the sampling of stochastic variables to report a good value for the property (the mean of replicates) and the error limits (95% confidence interval of the mean).

Soon we will consider systematic contributions to uncertainty due to reading error and calibration errors. These are errors that we know are present. Because we recognize the presence of reading and calibration errors, we can explore the sources of the errors, reason about how these errors affect samples, and construct standard errors for each of these sources (we do this in Chapters 3 and 4). Since random, reading, and calibration errors are independent effects, they simultaneously act on measurements, and thus they add up in ways we can determine (they add like variances – that is, in quadrature. This topic was introduced in Chapter 1, and details are discussed in Section 3.3).

Somewhat more difficult to deal with are systematic errors that we do *not* know are present. Unrecognized systematic errors do not show up as scatter in replicates because these errors correlate, or act systematically; thus they can skew our data without us ever knowing they were present. To see this dilemma in action, we discuss a specific example.

Example 2.17: Calibrating a flow meter. *A team is assigned to calibrate a rotameter flow meter. The rotameter is installed in a water flow loop (see Figure 4.4; water temperature* $= 19.1^{\circ}C$*) constructed of copper tubing, with the flow driven by a pump. To calibrate the rotameter, the flow rate is varied and independently measured by the "pail-and-scale" method: the water flow is collected in a "pail" (or tank, or other reservoir), and the mass of the water collected over a measured time interval is recorded. The mass measurement,*

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Table 2.5. *Data from calibrating a rotameter by the pail-and-scale method with two team members systematically dividing up data acquisition duties. Extra digits provided; see also Table 6.9.*

Index	Mass flow rate (kg/s)	Rotameter reading ($R\%$)
1	0.04315	14.67
2	0.05130	17.33
3	0.05515	18.50
4	0.06512	21.50
5	0.07429	24.33
6	0.08545	27.80
7	0.09585	31.00
8	0.10262	33.00
9	0.11449	36.67
10	0.13201	42.00
11	0.14325	45.50
12	0.15100	51.00
13	0.16882	56.33
14	0.18326	60.67
15	0.19643	65.00
16	0.20447	67.00
17	0.21361	69.67
18	0.22880	74.67
19	0.23625	77.00
20	0.24323	79.00
21	0.26620	85.50
22	0.27831	90.67

along with the collection time, allows the team to calculate the observed mass flow rate for each reading on the rotameter (reading $R\% = \%$ full scale).

The team divides up the calibration flow range, with one team member taking the low flow rates and the other taking the high flow rates. The data are shown in Table 2.5. What is the calibration curve for the rotameter? What is the uncertainty associated with the calibration curve?

Solution: This text presents the error-analysis tools needed to create a calibration curve for data such as that in Table 2.5: in Chapter 4 we introduce calibration and calibration error, and in Chapter 6, which builds on Chapter 5, we show how to produce an ordinary least-squares fit of a model to data. We discuss the current problem out of sequence to illustrate some

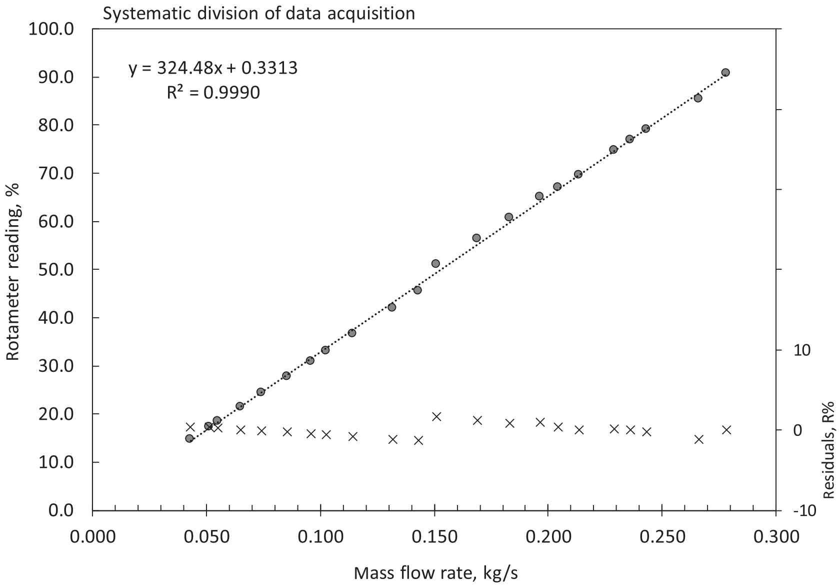

Figure 2.20 The rotameter calibration data are plotted for the case in which the work was systematically divided between the two experimenters, with one providing the low-flow-rate data and the other providing the high-flow-rate data. Also shown are the residuals at each value of x; this is the difference between the data and the fit. We observe a pattern in the residuals, a hint that a systemic error may be affecting the data.

common experimental-design and systematic errors that occur. We introduce the software tools we need as we go along. In the chapters that follow we offer a more thorough discussion of the model-fitting tools used here.

Systematic errors can creep into our data in ways that are difficult to anticipate. In the problem statement for this example, we learned a little bit about the experimental design for the rotameter calibration – in particular, the calibration team shared data acquisition duties, with one team member acquiring the low-flow-rate data points and the other team member acquiring the high-flow-rate ones. This arrangement seems harmless; both team members are presumably equally qualified to take the data. Splitting up the data acquisition in this way spreads out the work in an equitable way.

The data in Table 2.5 are plotted in Figure 2.20, and they look fine. Included in the figure is an ordinary least-squares curve fit, and the equation for the fit and its R^2 value (coefficient of determination; see Equation 6.17) are shown. This curve fit is obtained in Excel by right-clicking on the data in the plot and

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

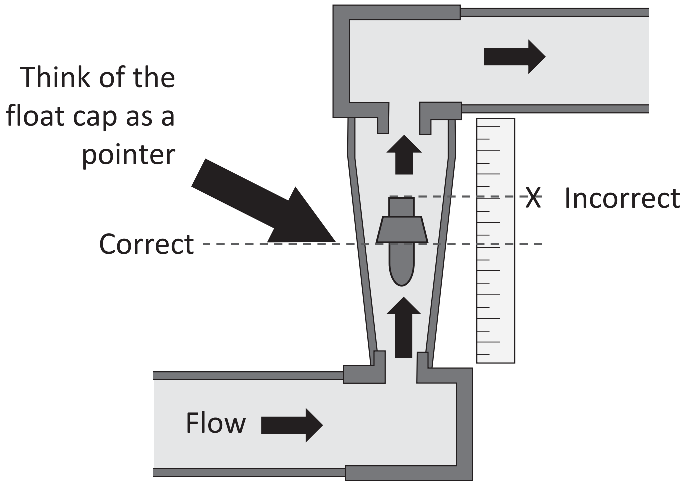

Figure 2.21 The rotameter is designed to be read from the position "pointed to" by the angled portion of the float. Sometimes new users think the reading comes from the location of the top of the float.

choosing to add a "Trendline"; in the Trendline dialog box we check the boxes that instruct Excel to display both the equation and the R^2 on the plot. The data follow a straight line, and R^2 is almost 1, indicating that the linear model is a good choice, as it is capturing the variation in the data. (We will have more to say on R^2 in Example 2.18 and in Chapter 6.)

The data look good, and the linear fit appears to be excellent. Unfortunately, there is a hidden systematic error in the data. The slope obtained by the trendline fit is 324 $R\%/(\text{kg/s})$; when the systematic error is found and eliminated, the correct slope obtained is 306 $R\%/(\text{kg/s})$. (A complete dataset without the systematic error is given in Table 6.9.) The data have an unrecognized systematic error.

We return to this problem in Example 2.18.

What hidden issue has affected the rotameter calibration? And how could we have prevented the problem? It turns out that the hidden systematic error in the data is the result of a misunderstanding in how to read the rotameter (Figure 2.21). The float in the rotameter has an angled indicator that "points" to a position on a scale; the reading that is expected by the manufacturer is the number that is pointed to by the indicator. Investigation revealed that one team member thought that the reading was determined by the position of the *top* of the float rather than by the position of the angled indicator.

The use of two different reading methods does not seem to have affected the data much: the data in Figure 2.20 are smooth and linear, and when a straight line is fit to the data, they produce an excellent value of $R^2 = 0.9990$. There is nothing to make us distrust the calibration. Having been told that there is a problem with the data, however, we can look closer and see that the lower-flow-rate data and the higher-flow-rate data seem to be slightly offset; they meet at rotameter reading 45. The effect is small and nearly invisible, however.

How would such an error ever be discovered? The hidden mistake in these data could have been found if the team had *randomized* their trials. Randomization is a technique that seeks to expose *each* experimental trial to the sources of stochastic variation that are present. For example, we may believe that it would make no difference who reads the rotameter during the calibration work. This is a reasonable assumption. To test this supposition, we can assign data trials randomly to different people. This random assignment may have no effect, as hypothesized. If there is an effect, however, the effect will appear randomly.

Why, we might ask, would we want to introduce variability? Why not just use one person so that the data are consistent? The reason to randomize, for example, the identity of the data-taker, is to guard against it *mattering* who took the data. If only one person takes the data, and if that person makes an error consistently, we will not be able to detect the mistake. Randomizing the data taker tests the hypothesis that it does not matter who takes the data. When we design a process to depend critically on one single aspect, such as entrusting all the data-taking to one individual's expertise or taking data in a systematic way from low to high values, we are exposing our experiments to the risk that all the data may contain hidden systematic errors. On a grand scale, when we scientists and engineers publish our results and invite others to attempt to reproduce the results, we are asserting that any competent investigator will obtain the same results as we obtained. If we have done our experiments correctly, and if our colleagues do the same, we will all get the same results. For our results to be of the highest quality, we should use randomization wherever possible to double-check that no systematic effects have been unintentionally introduced.

Our next question might be, why will randomizing help us find systematic errors? And, will we get any useful data out of the more scattered results we obtain when we randomize? If we randomize our experimental design and we were right that the change does not make a difference, we have shown that the change did not make a difference. This allows us to simplify future experiments with confidence, as we know definitively that the variable we randomized does not affect the outcome.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

If we randomize an aspect of our experimental design and there *is* an effect, the effect will become visible, as it will introduce additional scatter in the data. The presence of scatter is a message to the experimenter that there are random effects in the experiment as currently designed and executed. If the scatter is not too large and can be tolerated, nothing additional needs to be done. If the scatter is large and cannot be tolerated, its presence becomes a reason to reevaluate the experimental design to find the source of the scatter. The magnitude of the scatter becomes a detector of systematic error that needs to be addressed. Unrecognized systematic errors are a serious threat to the quality of the conclusions we may make with our data; knowing that there is an unrecognized effect – one that shows up as extra scatter when the process is randomized – is a welcome first step to tracking down a potentially serious problem with our studies.

We can see the randomization effect on the rotameter calibration in Example 2.17 if we start over with a different dataset in which each of the researchers again obtained half the data points, but in this case, the flow rates were assigned randomly.

Example 2.18: Calibrating a flow meter, revisited. *A team is assigned to calibrate a rotameter flow meter. The rotameter is installed in a water flow loop (see Figure 4.4; water temperature* $= 19.1°C$*) constructed of copper tubing, with the flow driven by a pump. To calibrate the rotameter, the flow rate is varied and independently measured by the "pail-and-scale" method.*

The team divides up the calibration flow range, with flow rates assigned randomly to the two team members. The data are shown in Table 2.6. What is the calibration curve for the rotameter? What is the uncertainty associated with the calibration curve?

Solution: The data in Table 2.6 are plotted in Figure 2.22.[18] These data look more scattered than the original plot in Figure 2.20. An ordinary least-squares fit to a line has been obtained with a slope of 309 $R\%/(\text{kg/s})$ and yields an R^2 of 0.9954.

The randomization of the data acquisition led to data with more scatter. The scatter in the data is a message: something is affecting the data. The message is received when the team reflects on what they see, holds a team meeting to review the procedure, and discovers and corrects the problem.

[18] These data were obtained by the author with each rotameter reading recorded in both the correct way and the incorrect way for each mass flow rate measurement. Thus, we could produce this second version of a plot for the same set of experiments. All the datasets referenced in Examples 2.17 and 2.18 (and Problems 6.7, 6.10, and 6.17) are also given in Table 6.9 (limited number of digits provided).

Table 2.6. *Data from calibrating a rotameter by the pail-and-scale method; data acquisition duties were divided randomly. Extra digits provided; see also Table 6.9.*

Index	Mass flow rate (kg/s)	Rotameter reading (R%)
1	0.04315	17.67
2	0.05130	17.33
3	0.05515	18.50
4	0.06512	24.50
5	0.07429	27.00
6	0.08545	27.80
7	0.09585	31.00
8	0.10262	33.00
9	0.11449	39.00
10	0.13201	45.00
11	0.14325	45.50
12	0.15100	48.33
13	0.16882	53.80
14	0.18326	60.67
15	0.19643	62.33
16	0.20447	62.50
17	0.21361	67.00
18	0.22880	74.67
19	0.23625	77.00
20	0.24323	79.00
21	0.26620	83.00
22	0.27831	90.67

For the rotameter data, plotting the points with different symbols according to who took the data made it clear that there was a systematic effect tied to operator identity. The take-away from this example is that we need to be on guard when we plan our experiments – systematic errors can sneak in through seemingly innocuous choices. Another take-away is that scatter in data can be a good thing if it shows that something unrecognized is affecting the data systematically.

This example also shows that values of R^2 that approach 1 *do not guarantee that data are of high quality and free from systematic errors*. As we discuss in Chapter 6, R^2 reflects the degree to which a chosen model (in our case, a straight line with nonzero slope) represents the data, compared to the

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

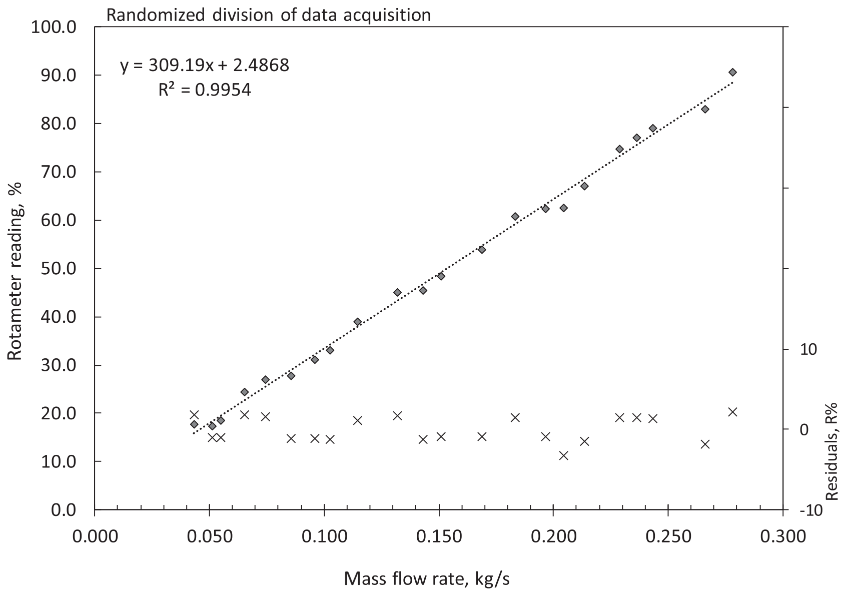

Figure 2.22 The rotameter calibration data are plotted for the task randomly divided between two experimenters. The scatter from the model line is greater than in Figure 2.20; this is reflected in the value of the standard deviation of y at a value of x, $s_{y,x}$ as discussed in the text. The scatter barely affects R^2. The residuals in this dataset (data − model) have a more random pattern compared to those in Figure 2.20.

assumption that the data are constant (flat line). In an ordinary least-squares fit, the statistic that reflects the scatter of data with respect to the model line is $s_{y,x}$, the standard deviation of y at a given x (see Chapter 6, Equation 6.25). For the data discussed in this example, the values of $s_{y,x}$ are

Shared data, assigned systematically (Example 2.17)
$$s_{y,x} = 0.783\ R\% \tag{2.61}$$

Shared data, assigned randomly: (Example 2.18)
$$s_{y,x} = 1.614\ R\% \tag{2.62}$$

The statistic $s_{y,x}$ correctly reflects the qualities of the fits. When the data are all taken correctly, $s_{y,x}$ is small and the data points are all close to the model line; $s_{y,x}$ for correctly taken data is 0.456 $R\%$ (see Table 6.9 and Problem 6.7). In contrast, when the results randomly include data that are read incorrectly, $s_{y,x}$ is large. The systematic case for Example 2.17 has an intermediate value for

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

$s_{y,x}$, indicating that the error in reading the rotameter is hidden in the relatively smooth data (with the wrong slope).

Obtaining an accurate calibration curve begins with taking good calibration data. Good data are obtained by following the best practices for minimizing random and systematic errors. Randomization is a key tool for identifying and eliminating systematic errors.

The discussion in this section focuses on the difference between random and systematic errors. We prefer random errors because we have a powerful tool for dealing with random error: replication. When the only differences among replicate trials are the amounts of random error present, then averaging the results gives us the true value of the measurement to a high degree of precision. A requirement for this to be valid, however, is that *only random error be present*. Systematic error will not disappear when replicates are averaged, and thus no matter the number of replicates, it remains our obligation to identify and eliminate systematic errors from our measurements.

Randomization is a tool for making systematic errors visible, as we have discussed. If we randomize our data acquisition, switching things that we do not think make a difference, then unrecognized effects can show up as additional "random" error. It may seem undesirable to design our protocols to amplify random error, but that is an inappropriate conclusion. Randomization is an important part of experimental troubleshooting since it is better to discover and correct systematic effects than to leave them in place, hidden, and to believe, incorrectly, that our results are high quality.

Randomization works against the *selective* introduction of errors into a process. If some aspect of a process introduces errors, these errors are easier to identify and fix if they affect all the data points rather than just some points. If a subset of trials are isolated from a source of error, that error source becomes systematic and possibly invisible. Replicates that result from data acquisition that is free from arbitrary systematic effects are called true replicates. In the next example we explore the kind of questions we can ask to ensure that our experiments produce true replicates.

Example 2.19: True replicates of viscosity with Cannon–Fenske viscometers. *Students are asked to measure the kinematic viscosity of a 30wt% aqueous sugar solution using Cannon–Fenske ordinary viscometers (Figure 2.23). The viscometers' procedure calls for loading an amount of solution (following a standard process), thermal-soaking the loaded viscometer in a temperature-controlled bath, and measuring the fluid's efflux time by drawing the fluid into the top reservoir and timing the travel of the fluid meniscus between timing marks. The viscometers are pre-calibrated by the manufacturer,*

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

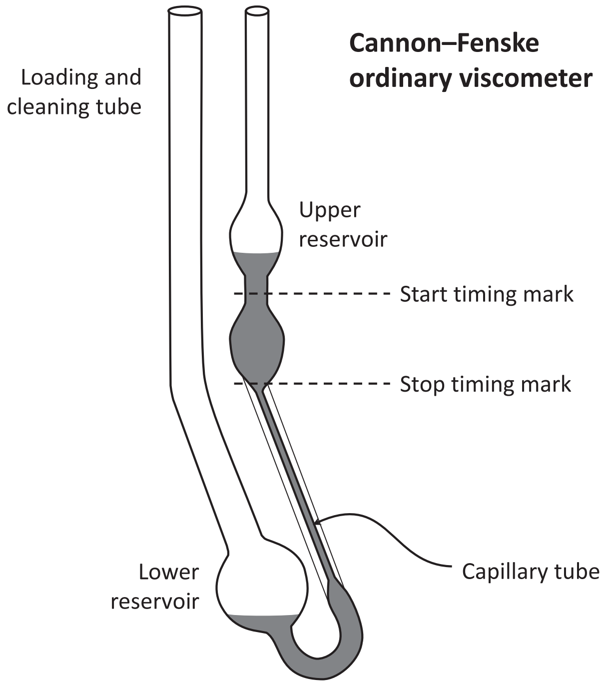

Figure 2.23 The Cannon–Fenske ordinary viscometer uses gravity to drive a fluid through a capillary tube. The viscosity comes from measuring the efflux time Δt_{eff} for the fluid meniscus to pass from the upper timing mark to the lower timing mark. The calibration coefficient $\tilde{\alpha}$ is supplied by the manufacturer. The viscosity $\tilde{\mu}$ is equal to $\tilde{\alpha}\rho\Delta t_{eff}$, where ρ is the fluid density. For more on Cannon–Fenske viscometers, see Example 5.8.

and kinematic viscosity $\tilde{\mu}/\rho$ is obtained by multiplying the efflux time by a manufacturer-supplied calibration constant. How can the student teams produce valid replicates of kinematic viscosity with these viscometers?

Solution: This is a question about experimental design. To address this question, we remind ourselves that a true replicate is exposed to all the elements that introduce error into the process. To plan for true replicates, then, we must reflect on all the elements that might potentially introduce error into the process.

We identify the following issues:

1. Loading the standardized volume may be subject to uncertainty in volume.
2. The temperature bath must accurately maintain the instrument at the desired temperature.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

3. The timing of the travel of the meniscus will be impacted by the operator's reaction time with the stopwatch.
4. The viscometer's calibration constant must be known accurately at the temperature of the experiment.
5. The particular viscometer used must not be defective; that is, it must perform the way the manufacturer warranties it to perform.
6. The viscometer must be clean and dry before use.
7. We did not mention it in the problem statement, but it is also important that when the liquid is flowing the viscometer is maintained in a vertical position and held motionless, since gravity drives the flow in these instruments.

We have many issues to consider when formulating our experimental plan.

Some of the viscometer operation issues presented can be addressed by replication. For example, the flow can be timed repeatedly. Since the operator's response time may lead to random positive or negative deviations in the measured efflux time, the average of repeated timings will give a good value of efflux time for a single loading of a viscometer. Likewise, the uncertainty in sample volume may be addressed by repeating the loading of samples. Each time the sample is loaded there will likely be a small positive or negative error in the volume added, and if this step is repeated, this random effect can be averaged out. The possibility of a defective or dirty viscometer can be explored by filling three or more different viscometers and averaging results across these different, but presumably equivalent, devices. This kind of repetition would also address a calibration problem associated with a single viscometer.

The issue of the water temperature cannot be addressed by replication (unless multiple baths are feasible), but must instead be addressed by calibration. The temperature of the bath must be measured and controlled with devices of established accuracy to eliminate the impact of a temperature offset.

The issue of the vertical placement of the viscometer may be addressed by using specially designed holders that ensure reproducible vertical placement. If these are not available, replication with different, careful operators will allow this source of uncertainty to be distributed across the replicates and, if the effect is random, it will average out.

Following a discussion of these issues, the class agreed that their true replicates would be obtained as follows:

1. A well-calibrated temperature indicator would be used to determine the bath temperature. All groups would use the same, carefully calibrated bath, which would be designed to hold the viscometers vertically.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

2. To reduce the impact of timing issues on efflux time, each group would draw up and time the flowing solution three times and use the average of the three timings to produce a single value of efflux time, which would be converted to a single value of kinematic viscosity.
3. Three groups would measure mean efflux time using three different viscometers and following the standard procedure, resulting in three replicates of viscosity.
4. The three independent measurements of kinematic viscosity (by the three groups) would be considered true replicates and would be averaged to yield the final kinematic viscosity value and its error limits (replicate error only).

To see this protocol at work, see Example 5.8, which considers the data analysis of Cannon–Fenske measurements on an aqueous sugar solution.

The thinking process used in Example 2.19 is a general solution and is recommended when a high degree of accuracy is called for: possible sources of both random and systematic error are identified; procedures are chosen to reduce or at least randomize the errors; and finally, true replicates are taken and averaged. Reflecting alone does not guarantee that we will think of all the sources of error in our measurements, but certainly it is an essential step toward ensuring better data acquisition and error reduction.

The last section of this chapter addresses the convention for significant figures on error limits.

2.4 Significant Figures on Error

When we determine a number from a measurement of some sort, we do not know that number with absolute certainty. In the previous sections we saw that for data subjected to random error only, we can take multiple measurements, average the results, and express the expected value of the quantity we are measuring as the calculated average along with an appropriate 95% confidence interval of the mean (Equation 2.52).

When presenting the result of such an exercise, we are faced with the choice of how many digits to show for both the expected value, $\bar{x}$, and for the error limits, $\pm t_{0.025,n-1}e_s$. The accepted practice for writing these results is to follow the significant-figures convention – that is, retain all certain digits and one uncertain digit. Since the error limits indicate the amount that the measurement may vary, we can adopt the following rule for error:

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Rule 1: Sig Figs on Error
Report only one digit on error.
(one uncertain digit)

Thus, a density 95% confidence interval limit of, for example, $2e_s = \pm 0.0323112$ g/cm^3 for a mean of $\bar{x} = 1.2549921$ g/cm^3 would be expressed as

$$\rho = 1.25 \pm 0.03 \text{ g/cm}^3 \quad \text{(95\% CI, one uncertain digit)}$$

Note that the error limits make the digit 5 on the density uncertain, and therefore to follow the significant-figures convention, we round our results and report only up to that one, uncertain, decimal place.

Although Rule 1 says we keep only one digit on error, we do make an exception to the one-digit rule in some cases, as we see in Rule 2.

Rule 2: Sig Figs on Error
We may report two digits on error
if the error digit is 1 or 2.
(two uncertain digits)

If the uncertainty for density had been $2e_s = 0.0123112$ g/cm^3 (error digit is 1 – that is ± 0.01 g/cm^3) on an expected value of 1.2549921 g/cm^3, the result would be expressed with an additional uncertain error digit:

$$\rho = 1.255 \pm 0.012 \text{ g/cm}^3 \quad \text{(95\% CI, two uncertain digits)} \tag{2.63}$$

The reasoning behind this second rule is that when the error digit is 1 or 2, the next digit to the right has a large effect when rounding, and keeping the extra digit will allow the reader to calculate the 95% confidence interval with less round-off effect.[19] The user of the number in Equation 2.63 must remember, however, to correctly interpret the density to have only three significant figures, even though four digits are shown (two digits are uncertain). The presence of the error limit with two digits shown makes the two-digit uncertainty clear.

It bears repeating that it is good practice to tell the reader what system you are using to express your uncertainty. We use 95% confidence limits or about two standard errors. Others may use 68% (approximately one standard error) or 99% (approximately three standard errors; see Problem 2.37). If the author fails to indicate which meaning is intended, there is no sure way of knowing which standard is being employed.

[19] When the digit you are rounding is 5, there is no good justification to choose to round up or to choose to round down. The best we can do is to randomize this choice. This can be achieved by targeting obtaining an even number after rounding. This practice, over many roundings, is unbiased, whereas always rounding the digit 5 up (or down) is systematic and can introduce bias.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

Discussing the sig-figs rules brings up another related issue, that of round-off error. When we round off or truncate numbers in a calculation, we introduce calculation errors. Calculations done by computers may be the result of thousands or millions of individual mathematical operations. Round-off errors are undesirable, and for this reason, within the internal functioning of calculators and computers, those devices retain many digits (at least 32) so that the round-off errors affect only digits that are well away from the digits that we are going to retain.

In our own calculations, we should follow the same practices. If we are doing follow-up calculations with a value that we calculated or measured, in the intermediate calculations we should use all the digits we have. This requires us to record extra digits from intermediate calculations and to use the most precise values of constants obtained elsewhere. In performing 95% confidence and prediction interval calculations, when a value of $t_{0.025,n-1}$ is needed from the Student's t distribution, we should use the most accurate available value, by employing, for example, Excel's T.INV.2T() function call rather than using the truncated numbers from Table 2.3. Shortcuts (using truncated or rounded values) may be employed for estimates, but high-precision numbers are best to use for important calculations. Keeping extra digits is comparable to what a computer or calculator does internally: it keeps all the digits it has. It is only at the last step, when we report to others our final answer of a quantity, that we must report the appropriate number of significant figures. Rounding off intermediate calculations can severely degrade the precision of a final calculation and should be avoided.

2.5 Summary

In this chapter, we present the basics of using statistics to quantify the effects of random errors on measurements. The methods discussed are summarized here; the discussion of the impact of reading errors begins in Chapter 3, followed by calibration error in Chapter 4 (both are systematic errors).

Summary of Properties of a Sample of n Observations of Stochastic Variable x, Subject Only to Random Error

- A quantity subject to random variation is called a stochastic variable. Values obtained from experimental measurements are continuous stochastic variables.
- Stochastic variables are represented by an expected value and error limits on the expected value.

- When a stochastic variable is sampled, the mean of a sample set (mean $\bar{x}$; size n; sample standard deviation s) is an unbiased estimate of the expected value of x. This is a formal statement of the correctness of using averages of true replicates to estimate the value of a stochastic variable, x.
- With 95% confidence, the true value of x (the mean of the underlying distribution of the stochastic variable) is within the range $\approx \bar{x} \pm 2s\left(\frac{1}{\sqrt{n}}\right)$ (**for** $n \geq 7$) or the range $\bar{x} \pm t_{0.025,n-1}\left(\frac{s}{\sqrt{n}}\right)$ (**for all** n**, but especially for** $n < 7$).
- With 95% confidence, the next value of x, if the measurement were repeated, is within the range $\approx \bar{x} \pm 2s\sqrt{\left(1+\frac{1}{n}\right)}$ (**for** $n \geq 7$) or the range $\bar{x} \pm t_{0.025,n-1}s\sqrt{\left(1+\frac{1}{n}\right)}$ (**for all** n**, but especially for** $n < 7$) as established by a previous sample set of size n, mean $\bar{x}$, and sample standard deviation s.
- The accepted convention is to use one digit (one uncertain digit) on error (except if the error digit is 1 or 2, in which case use two uncertain digits on error). Report the value x to no more than the number of decimal places in the error.
- Do not round off digits in intermediate calculations; carry several extra digits and round only the final answer to the appropriate number of significant figures.
- It is recommended to round in an unbiased way. When the digit you are rounding is 5, there is no good justification to choose to round up or to choose to round down. The best we can do is to randomize this choice. This can be achieved by seeking to obtain an even number after rounding. This practice, over many roundings, is unbiased, whereas always rounding the digit 5 up (or down) is systematic and can introduce bias.
- **Bonus advice:** In engineering practice, we usually have no more than two or three significant figures, and we can even expect to have only one significant figure in some cases. Only with extreme care can we get four significant figures. If you have not rigorously designed and executed your measurement with the aim of eliminating error, you have no more than three sig figs, and quite likely you have two or one sig figs. Our advice: **avoid reporting four or more significant figures in an engineering report**. Two significant figures is the most likely precision in engineering work; only if you can justify it should you use three significant figures or more.

2.6 Problems

1. Which of the following are stochastic variables and which are not? What could be the source of the stochastic variations?

   (a) Weight of a cup of sugar
   (b) The number of days in a week
   (c) The temperature at noon in Manila, Philippines
   (d) The number of counties in the state of Michigan
   (e) The number typical of counties in a U.S. state

2. In Example 2.1 we calculate Eun Young's mean commuting time to be 28 min, but in Example 2.2 we calculate her commuting time to be 32 min. Which mean is correct? Explain your answer.
3. In Example 2.1 we calculate Eun Young's commuting time to be 28 min. Looking at the data used to calculate this mean time, how much would you expect her actual commuting time to vary? We are not asking for a mathematical calculation in this question; rather, using your intuition, what would you expect the commuting time to be, most of the time? (Once you have made your estimate, see Example 2.15 for the mathematical answer.)
4. How do we determine probabilities from probability density functions (pdf) for stochastic variables? In other words, for the stochastic variable commuting time (see Examples 2.1 and 2.2), if we knew its pdf, how would we calculate, for example, the probability of the commute taking between 50 and 55 min?
5. In Example 2.3 we presented the pdf for the duration of George's daily work commute (Equation 2.13). What is the probability that George's commuting time is 35 min or longer?
6. For the pdf provided in Example 2.3 (Equation 2.13), what is the probability that it takes George between 20 and 25 min to commute?
7. From the pdf provided in Example 2.3 (Equation 2.13), what is George's mean commuting time?
8. Reproduce the plot in Figure 2.5, which shows the pdf of George's commuting time, using mathematical software. Describe in a few sentences the implications of the shape of the pdf.
9. Reproduce the plot in Figure 2.6, which shows the pdf of the normal distribution, using mathematical software. What is the normal distribution? Describe it in a few sentences.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

10. Reproduce the plot in Figure 2.9, which shows the pdf of the Student's t distribution, using mathematical software. What is the Student's t distribution? Describe it in a few sentences.
11. A sample set of packaged snack food items is sent to the lab. Each item is weighed. For a sample for which $n = 25$, $\bar{x} = 456.323$ g, and $s = 6.4352$ g (extra digits supplied), calculate the error limits on the mean mass and assign the correct number of significant figures. Discuss your observations.
12. A sample of 16 maple tree leaves is collected; we measure the mass of each leaf. For the leaf mass data shown in Table 2.7, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the mean maple leaf mass in the sample? Include the appropriate error limits.
13. A sample of 16 maple tree leaves is collected; we measure the length from the leaf stem to the tip of the leaf. For the leaf length data shown in Table 2.7, calculate the sample mean $\bar{x}$, standard deviation s, and

Table 2.7. *Sample of maple tree leaves masses and lengths.*

i	Mass (g)	Length (cm)
1	0.93	9.5
2	1.38	11.7
3	1.43	10.9
4	1.41	10.4
5	0.78	8.4
6	1.07	10.0
7	2.17	11.5
8	1.43	12.0
9	1.34	11.5
10	0.92	8.8
11	0.73	7.8
12	0.85	8.5
13	1.49	11.5
14	1.11	9.5
15	1.29	9.8
16	0.59	8.4

Table 2.8. *Sample of lilac bush leaves masses and lengths.*

i	Mass (g)	Length (cm)
1	0.65	5.9
2	0.37	4.7
3	0.66	5.9
4	0.41	4.7
5	0.99	6.4
6	0.74	5.7
7	1.01	6.0
8	0.64	5.3
9	0.42	4.2
10	0.5	5.0
11	0.53	4.4
12	0.6	5.2
13	0.47	4.5
14	0.44	4.2
15	0.56	4.9
16	1.06	6.6
17	0.73	5.5

standard error from replicates $s/\sqrt{n}$. What is the mean maple leaf length in the sample? Include the appropriate error limits.

14. A sample of 17 leaves from a lilac bush is collected; we measure the mass of each leaf. For the leaf mass data supplied in Table 2.8, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the mean lilac leaf mass in the sample? Include the appropriate error limits.
15. A sample of 17 leaves from a lilac bush is collected; we measure the length across the broadest part of the leaf. For the leaf length data supplied in Table 2.8, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the mean lilac leaf length in the sample? Include the appropriate error limits.
16. A sample of 15 leaves from a flowering crab tree is collected; we measure the mass of each leaf. For the leaf mass data supplied in Table 2.9, calculate the sample mean $\bar{x}$, standard deviation s, and standard error

Table 2.9. *Sample of flowering crab tree leaves masses and lengths.*

i	Mass (g)	Length (cm)
1	0.56	5.5
2	0.37	4.5
3	0.32	4.3
4	0.36	5.0
5	0.47	4.3
6	0.61	5.6
7	0.43	4.8
8	0.36	4.1
9	0.49	4.7
10	0.45	4.7
11	0.27	3.5
12	0.61	4.9
13	0.27	4.7
14	0.59	5.2
15	0.50	5.2

from replicates $s/\sqrt{n}$. What is the mean flowering crab leaf mass in the sample? Include the appropriate error limits.

17. A sample of 15 leaves from a flowering crab tree is collected; we measure the length across the longest part of the leaf, from stem to tip. For the leaf length data supplied in Table 2.9, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the mean flowering crab leaf length in the sample? Include the appropriate error limits.
18. The process for manufacturing plastic 16-oz drinking cups produces seemingly identical cups. We weigh 19 cups to see how much their masses vary. For the data shown in Table 2.10, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the mean cup mass in the sample? Include the appropriate error limits.
19. For the Cannon–Fenske viscometer efflux time replicates in Table 5.3, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the 95% confidence interval of efflux time for each of the three viscometers? See Table 5.4 for some of the answers to this problem.

Table 2.10. *Sample of masses of plastic cups.*

i	Mass (g)
1	8.47
2	8.48
3	8.53
4	8.45
5	8.44
6	8.46
7	8.49
8	8.52
9	8.48
10	8.51
11	8.42
12	8.45
13	8.47
14	8.49
15	8.44
16	8.47
17	8.45
18	8.48
19	8.47

20. For the three viscosity replicates in Table 5.4, calculate the sample mean $\bar{x}$, standard deviation s, and standard error from replicates $s/\sqrt{n}$. What is the 95% confidence interval of solution viscosity?
21. For the maple-leaf mass data given in Table 2.7, what are the sample mean and sample standard deviation? Calculate the level of confidence associated with reporting two, three, or four significant figures on the mean mass. Express your answer for the mean with the appropriate number of significant figures.
22. For the plastic cup mass data given in Table 2.10, what are the sample mean and sample standard deviation? Calculate the level of confidence associated with reporting two, three, or four significant figures on the mean mass. Express your answer for the mean with the appropriate number of significant figures.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

23. For the Blue Fluid 175 density data given in Example 2.8, what are the sample mean and sample standard deviation? Calculate the level of confidence associated with reporting two, three, or four significant figures on the mean density. Express your answer for the mean with the appropriate number of significant figures.
24. In Example 2.4 we reported on measurements of length for a sample of 25 sticks from a shop. If the mean stick length is reported with three significant figures, what is the level of confidence we are asserting? Repeat for two and one sig figs. How many sig figs will you recommend reporting?
25. Sometimes lab workers take a shortcut and assume an error limit such as 1% error as a rule of thumb. For the plastic cup mass data in Table 2.10, what are the $\pm 1\%$ error limits on the mean? What level of confidence is associated with these error limits? Comment on your answer.
26. For the sample of packaged food weights described in Problem 2.11, what is the probability that the true mean is within $\pm 1\%$ of the measured mean? Comment on your answer.
27. Based on the sample of plastic cup data given in Table 2.10, what is the probability that a cup (a "next" cup) will weigh less than 8.42 g?
28. The density of an aqueous sugar solution was measured ten times as reported in Example 2.5. What are the mean and standard deviation of the density data? What is the standard deviation of the mean? An eleventh data point is taken and the result is 1.1003 g/cm^3 (extra digits supplied). How many standard deviations (of the mean) is this result from the mean? In your own words, what is the significance of this positioning?
29. In Example 2.14 we identified a new data point on density ($\rho_{BF,n+1} = 1.755$ g/cm^3) as being outside the expected 95% prediction interval of the dataset, implying that there may have been a problem with the measurement. If we broadened our prediction interval to, for example, 96% confidence, perhaps the data point would be included. What is the smallest prediction interval (the smallest percent confidence) that we could choose to use to make the new obtained value in that example consistent with the other members of the dataset? Comment on your findings.
30. Using the "worst-case method" discussed in the text, estimate the duration of Eun Young's commute, along with error limits. Perform your calculation for both datasets provided (Examples 2.1 and 2.2). Calculate also the 95% confidence interval and the 95% prediction interval for the

combined dataset. Comment on your answers. What do you tell Eun Young about the probable duration of her commute tomorrow?

31. For Eun Young's commute, as discussed in Examples 2.1 and 2.2, what is the probability that tomorrow the commute will take more than 40 min? Use the combined dataset ($n = 20$) for your calculation.
32. For the stick vendor's data from Example 2.4, what is the 95% prediction interval for the next value of stick length? What is the probability that a stick chosen at random from the vendor's collection is between 5.5 and 6.5 cm? *Hint: we know the next value in terms of deviation. What is the probability?*
33. For the maple leaf data in Table 2.7, calculate the 95% confidence interval of the mean and the 95% prediction interval of the next value of leaf mass. Create a plot like Figure 2.19 showing the data and the intervals. In your own words, what do these two limits represent?
34. For the lilac leaf data in Table 2.8, calculate the 95% confidence interval of the mean and the 95% prediction interval of the next value of leaf length. Create a plot like Figure 2.19 showing the data and the intervals. In your own words, what do these two limits represent?
35. For the flowering crab leaf data in Table 2.9, calculate the 95% confidence interval of the mean and the 95% prediction interval of the next value of both leaf mass and length. Create a plot like Figure 2.19 showing the data and the intervals. In your own words, what do these two limits represent?
36. For the plastic cup data in Table 2.10, calculate the 95% confidence interval of the mean and the 95% prediction interval of the next value of cup mass. Create a plot like Figure 2.19 showing the data and the intervals. In your own words, what do these two limits represent?
37. For data that follow the normal distribution (pdf given in Equation 2.17), calculate the confidence level associated with error limits of $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$.
38. For all probability density distributions, the integral of the pdf from $-\infty$ to ∞ is 1. Verify that this is the case for the normal distribution (Equation 2.17).
39. Do these two numbers agree: $62 \pm 14°$C; $58 \pm 2°$C? Justify your answer with a graphic.
40. An (x, y) dataset is given here: (20, 16.234); (30, 16.252); (40, 16.271); (50, 16.211); (60, 16.201); (70, 16.292); (80, 16.235). Plot y versus x. Does y depend on x or is y independent of x within a reasonable amount of uncertainty in y? Justify your answer.

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press

41. Equation 2.58 gives the 95% prediction interval for the next value of x:

95% prediction interval for x_i, the next value of x:
$$x_i = \bar{x} \pm t_{0.025,n-1}\, s\sqrt{1+\frac{1}{n}} \quad \text{(Equation 2.58)}$$

Using the error-propagation techniques of Chapter 5, show that Equation 2.58 holds.

42. Based on the sample of plastic cup data given in Table 2.10, what is the probability that a cup will weigh more than 8.55 g?

https://doi.org/10.1017/9781108777513.003 Published online by Cambridge University Press