1 Introduction

This chapter lays a foundation for all that follows: It contains a road map for the study of engineering statistics. The subject is defined, its importance is described, some basic terminology is introduced, and the important issue of measurement is discussed. Finally, the role of mathematical models in achieving the objectives of engineering statistics is investigated.

1.1 Engineering Statistics: What and Why

In general terms, what a working engineer does is to design, build, operate, and/or improve physical systems and products. This work is guided by basic mathematical and physical theories learned in an undergraduate engineering curriculum. As the engineer's experience grows, these quantitative and scientific principles work alongside sound engineering judgment. But as technology advances and new systems and products are encountered, the working engineer is inevitably faced with questions for which theory and experience provide little help. When this happens, what is to be done?

On occasion, consultants can be called in, but most often an engineer must independently find out "what makes things tick." It is necessary to **collect and interpret data** that will help in understanding how the new system or product works. Without specific training in data collection and analysis, the engineer's attempts can be haphazard and poorly conceived. Valuable time and resources are then wasted, and sometimes erroneous (or at least unnecessarily ambiguous) conclusions are reached. To avoid this, it is vital for a working engineer to have a toolkit that includes the best possible principles and methods for gathering and interpreting data.

The goal of engineering statistics is to provide the concepts and methods needed by an engineer who faces a problem for which his or her background does not serve as a completely adequate guide. It supplies principles for how to efficiently acquire and process empirical information needed to understand and manipulate engineering systems.

Definition 1

Engineering statistics is the study of how best to

1. collect engineering data,
2. summarize or describe engineering data, and
3. draw formal inferences and practical conclusions on the basis of engineering data,

all the while recognizing the reality of variation.

To better understand the definition, it is helpful to consider how the elements of engineering statistics enter into a real problem.

Example 1

Heat Treating Gears

The article "Statistical Analysis: Mack Truck Gear Heat Treating Experiments" by P. Brezler (*Heat Treating*, November, 1986) describes a simple application of engineering statistics. A process engineer was faced with the question, "How should gears be loaded into a continuous carburizing furnace in order to minimize distortion during heat treating?" Various people had various semi-informed opinions about how it should be done—in particular, about whether the gears should be laid flat in stacks or hung on rods passing through the gear bores. But no one really knew the consequences of laying versus hanging.

Data collection

In order to settle the question, the engineer decided to get the facts—to collect some data on "thrust face runout" (a measure of gear distortion) for gears laid and gears hung. Deciding exactly how this data collection should be done required careful thought. There were possible differences in gear raw material lots, machinists and machines that produced the gears, furnace conditions at different times and positions within the furnace, technicians and measurement devices that would produce the final runout measurements, etc. The engineer did not want these differences either to be mistaken for differences between the two loading techniques or to unnecessarily cloud the picture. Avoiding this required care.

Data summarization

In fact, the engineer conducted a well-thought-out and executed study. Table 1.1 shows the runout values obtained for 38 gears laid and 39 gears hung after heat treating. In raw form, the runout values are hardly understandable. They lack organization; it is not possible to simply look at Table 1.1 and tell what is going on. The data needed to be summarized. One thing that was done was to compute some numerical summaries of the data. For example, the process engineer found

$$\text{Mean laid runout} = 12.6$$

$$\text{Mean hung runout} = 17.9$$

Table 1.1
Thrust Face Runouts (.0001 in.)

Gears Laid	Gears Hung
5, 8, 8, 9, 9,	7, 8, 8, 10, 10,
9, 9, 10, 10, 10,	10, 10, 11, 11, 11,
11, 11, 11, 11, 11,	12, 13, 13, 13, 15,
11, 11, 12, 12, 12,	17, 17, 17, 17, 18,
12, 13, 13, 13, 13,	19, 19, 20, 21, 21,
14, 14, 14, 15, 15,	21, 22, 22, 22, 23,
15, 15, 16, 17, 17,	23, 23, 23, 24, 27,
18, 19, 27	27, 28, 31, 36

Further, a simple graphical summarization was made, as shown in Figure 1.1.

Variation

From these summaries of the runouts, several points are obvious. One is that there is variation in the runout values, even within a particular loading method. Variability is an omnipresent fact of life, and all statistical methodology explicitly recognizes this. In the case of the gears, it appears from Figure 1.1 that there is somewhat more variation in the hung values than in the laid values.

But in spite of the variability that complicates comparison between the loading methods, Figure 1.1 and the two group means also carry the message that the laid runouts are on the whole smaller than the hung runouts. By how much? One answer is

$$\text{Mean hung runout} - \text{Mean laid runout} = 5.3$$

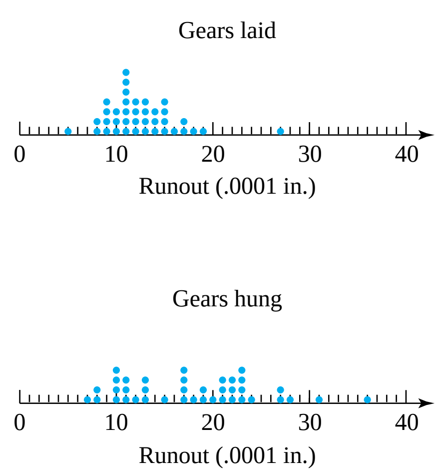

Figure 1.1 Dot diagrams of runouts

Example 1 *(continued)*

But how "precise" is this figure? Runout values are variable. So is there any assurance that the difference seen in the present means would reappear in further testing? Or is it possibly explainable as simply "stray background noise"? Laying gears is more expensive than hanging them. Can one know whether the extra expense is justified?

Drawing inferences from data

These questions point to the need for methods of formal statistical inference from data and translation of those inferences into practical conclusions. Methods presented in this text can, for example, be used to support the following statements about hanging and laying gears:

1. One can be roughly 90% sure that the difference in long-run mean runouts produced under conditions like those of the engineer's study is in the range

   3.2 to 7.4

2. One can be roughly 95% sure that 95% of runouts for gears laid under conditions like those of the engineer's study would fall in the range

   3.0 to 22.2

3. One can be roughly 95% sure that 95% of runouts for gears hung under conditions like those of the engineer's study would fall in the range

   .8 to 35.0

These are formal quantifications of what was learned from the study of laid and hung gears. To derive practical benefit from statements like these, the process engineer had to combine them with other information, such as the consequences of a given amount of runout and the costs for hanging and laying gears, and had to apply sound engineering judgment. Ultimately, the runout improvement was great enough to justify some extra expense, and the laying method was implemented.

The example shows how the elements of statistics were helpful in solving an engineer's problem. Throughout this text, the intention is to emphasize that the topics discussed are not ends in themselves, but rather tools that engineers can use to help them do their jobs effectively.

Section 1 Exercises

1. Explain why engineering practice is an inherently statistical enterprise.
2. Explain why the concept of variability has a central place in the subject of engineering statistics.
3. Describe the difference between descriptive and (formal) inferential statistics.

1.2 Basic Terminology

Engineering statistics requires learning both new words and new technical meanings for familiar words. This section introduces some common jargon for types of statistical studies, types of data that can arise in those studies, and types of structures those data can have.

1.2.1 Types of Statistical Studies

When an engineer sets about to gather data, he or she must decide how active to be. Will the engineer turn knobs and manipulate process variables or simply let things happen and try to record the salient features?

Definition 2

> An **observational study** is one in which the investigator's role is basically passive. A process or phenomenon is watched and data are recorded, but there is no intervention on the part of the person conducting the study.

Definition 3

> An **experimental study** (or, more simply, an *experiment*) is one in which the investigator's role is active. Process variables are manipulated, and the study environment is regulated.

Most real statistical studies have both observational and experimental features, and these two definitions should be thought of as representing idealized opposite ends of a continuum. On this continuum, the experimental end usually provides the most **efficient and reliable** ways to collect engineering data. It is typically much quicker to manipulate process variables and watch how a system responds to the changes than to passively observe, hoping to notice something interesting or revealing.

Inferring causality

In addition, it is far easier and safer to infer **causality** from an experiment than from an observational study. Real systems are complex. One may observe several instances of good process performance and note that they were all surrounded by circumstances X without being safe in assuming that circumstances X cause good process performance. There may be important variables in the background that are changing and are the true reason for instances of favorable system behavior. These so-called **lurking variables** may govern both process performance and circumstances X. Or it may simply be that many variables change haphazardly without appreciable impact on the system and that by chance, during a limited period of observation, some of these happen to produce X at the same time that good performance occurs. In either case, an engineer's efforts to create X as a means of making things work well will be wasted effort.

On the other hand, in an experiment where the environment is largely regulated except for a few variables the engineer changes in a purposeful way, an inference of causality is much stronger. If circumstances created by the investigator are consistently accompanied by favorable results, one can be reasonably sure that they caused the favorable results.

Example 2

Pelletizing Hexamine Powder

Cyr, Ellson, and Rickard attacked the problem of reducing the fraction of nonconforming fuel pellets produced in the compression of a raw hexamine powder in a pelletizing machine. There were many factors potentially influencing the percentage of nonconforming pellets: among others, Machine Speed, Die Fill Level, Percent Paraffin added to the hexamine, Room Temperature, Humidity at manufacture, Moisture Content, "new" versus "reground" Composition of the mixture being pelletized, and the Roughness of the chute entered by the freshly stamped pellets. Correlating these many factors to process performance through passive observation was hopeless.

The students were, however, able to make significant progress by conducting an experiment. They chose three of the factors that seemed most likely to be important and purposely changed their levels while holding the levels of other factors as close to constant as possible. The important changes they observed in the percentage of acceptable fuel pellets were appropriately attributed to the influence of the system variables they had manipulated.

Besides the distinction between observational and experimental statistical studies, it is helpful to distinguish between studies on the basis of the **intended breadth of application of the results**. Two relevant terms, popularized by the late W. E. Deming, are defined next:

Definition 4

An **enumerative study** is one in which there is a particular, well-defined, finite group of objects under study. Data are collected on some or all of these objects, and conclusions are intended to apply only to these objects.

Definition 5

An **analytical study** is one in which a process or phenomenon is investigated at one point in space and time with the hope that the data collected will be representative of system behavior at other places and times under similar conditions. In this kind of study, there is rarely, if ever, a particular well-defined group of objects to which conclusions are thought to be limited.

Most engineering studies tend to be of the second type, although some important engineering applications do involve enumerative work. One such example is the reliability testing of critical components—e.g., for use in a space shuttle. The interest is in the components actually in hand and how well they can be expected to perform rather than on any broader problem like "the behavior of all components of this type." Acceptance sampling (where incoming lots are checked before taking formal receipt) is another important kind of enumerative study. But as indicated, most engineering studies are analytical in nature.

Example 2 (*continued*)

The students working on the pelletizing machine were not interested in any particular batch of pellets, but rather in the question of how to make the machine work effectively. They hoped (or tacitly assumed) that what they learned about making fuel pellets would remain valid at later times, at least under shop conditions like those they were facing. Their experimental study was analytical in nature.

Particularly when discussing enumerative studies, the next two definitions are helpful.

Definition 6

A **population** is the entire group of objects about which one wishes to gather information in a statistical study.

Definition 7

A **sample** is the group of objects on which one actually gathers data. In the case of an enumerative investigation, the sample is a subset of the population (and can in some cases include the entire population).

Figure 1.2 shows the relationship between a population and a sample. If a crate of 100 machine parts is delivered to a loading dock and 5 are examined in order to verify the acceptability of the lot, the 100 parts constitute the population of interest, and the 5 parts make up a (single) sample of size 5 from the population. (Notice the word usage here: There is *one* sample, not *five* samples.)

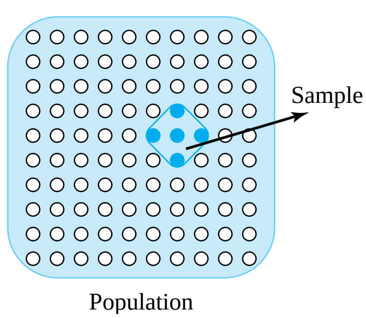

Figure 1.2 Population and sample

There are several ways in which the meanings of the words *population* and *sample* are often extended. For one, it is common to use them to refer to not only objects under study but also data values associated with those objects. For example, if one thinks of Rockwell hardness values associated with 100 crated machine parts, the 100 hardness values might be called a population (of numbers). Five hardness values corresponding to the parts examined in acceptance sampling could be termed a sample from that population.

Example 2
(continued)

Cyr, Ellson, and Rickard identified eight different sets of experimental conditions under which to run the pelletizing machine. Several production runs of fuel pellets were made under each set of conditions, and each of these produced its own percentage of conforming pellets. These eight sets of percentages can be referred to as eight different samples (of numbers).

Also, although strictly speaking there is no concrete population being investigated in an analytical study, it is common to talk in terms of a **conceptual population** in such cases. Phrases like "the population consisting of all widgets that could be produced under these conditions" are sometimes used. We dislike this kind of language, believing that it encourages fuzzy thinking. But it is a common usage, and it is supported by the fact that typically the same mathematics is used when drawing inferences in enumerative and analytical contexts.

1.2.2 Types of Data

Engineers encounter many types of data. One useful distinction concerns the degree to which engineering data are intrinsically numerical.

Definition 8

Qualitative or **categorical** data are the values of basically nonnumerical characteristics associated with items in a sample. There can be an order associated with qualitative data, but aggregation and counting are required to produce any meaningful numerical values from such data.

Consider again 5 machine parts constituting a sample from 100 crated parts. If each part can be classified into one of the (ordered) categories (1) conforming, (2) rework, and (3) scrap, and one knows the classifications of the 5 parts, one has 5 qualitative data points. If one aggregates across the 5 and finds 3 conforming, 1 reworkable, and 1 scrap, then numerical summaries have been derived from the original categorical data by counting.

In contrast to categorical data are numerical data.

Definition 9

Quantitative or **numerical** data are the values of numerical characteristics associated with items in a sample. These are typically either **counts** of the number of occurrences of a phenomenon of interest or **measurements** of some physical property of the items.

Returning to the crated machine parts, Rockwell hardness values for 5 selected parts would constitute a set of quantitative measurement data. Counts of visible blemishes on a machined surface for each of the 5 selected parts would make up a set of quantitative count data.

It is sometimes convenient to act as if infinitely precise measurement were possible. From that perspective, measured variables are **continuous** in the sense that their sets of possible values are whole (continuous) intervals of numbers. For example, a convenient idealization might be that the Rockwell hardness of a machine part can lie anywhere in the interval $(0, \infty)$. But of course this is only an idealization. All real measurements are to the nearest unit (whatever that unit may be). This is becoming especially obvious as measurement instruments are increasingly equipped with digital displays. So in reality, when looked at under a strong enough magnifying glass, all numerical data (both measured and count alike) are **discrete** in the sense that they have isolated possible values rather than a continuum of available outcomes. Although $(0, \infty)$ may be mathematically convenient and completely adequate for practical purposes, the real set of possible values for the measured Rockwell hardness of a machine part may be more like $\{.1, .2, .3, \ldots\}$ than like $(0, \infty)$.

Well-known conventional wisdom is that measurement data are preferable to categorical and count data. Statistical methods for measurements are simpler and more informative than methods for qualitative data and counts. Further, there is typically far more to be learned from appropriate measurements than from qualitative data taken on the same physical objects. However, this must sometimes be balanced against the fact that measurement can be more time-consuming (and thus expensive) than the gathering of qualitative data.

Example 3

Pellet Mass Measurements

As a preliminary to their experimental study on the pelletizing process (discussed in Example 2), Cyr, Ellson, and Rickard collected data on a number of aspects of machine behavior. Included was the mass of pellets produced under standard operating conditions. Because a nonconforming pellet is typically one from which some material has broken off during production, pellet mass is indicative of system performance. Informal requirements for (specifications on) pellet mass were from 6.2 to 7.0 grams.

Example 3 *(continued)*

Information on 200 pellets was collected. The students could have simply observed and recorded whether or not a given pellet had mass within the specifications, thereby producing qualitative data. Instead, they took the time necessary to actually measure pellet mass to the nearest .1 gram—thereby collecting measurement data. A graphical summary of their findings is shown in Figure 1.3.

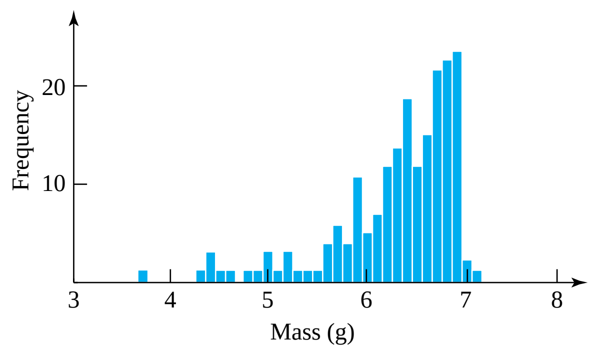

Figure 1.3 Pellet mass measurements

Notice that one can recover from the measurements the conformity/nonconformity information—about 28.5% (57 out of 200) of the pellets had masses that did not meet specifications. But there is much more in Figure 1.3 besides this. The shape of the display can give insights into how the machine is operating and the likely consequences of simple modifications to the pelletizing process. For example, note the **truncated** or chopped-off appearance of the figure. Masses do not trail off on the high side as they do on the low side. The students reasoned that this feature of their data had its origin in the fact that after powder is dispensed into a die, it passes under a paddle that wipes off excess material before a cylinder compresses the powder in the die. The amount initially dispensed to a given die may have a fairly symmetric mound-shaped distribution, but the paddle probably introduces the truncated feature of the display.

Also, from the numerical data displayed in Figure 1.3, one can find a percentage of pellet masses in any interval of interest, not just the interval [6.2, 7.0]. And by mentally sliding the figure to the right, it is even possible to project the likely effects of increasing die size by various amounts.

It is typical in engineering studies to have several response variables of interest. The next definitions present some jargon that is useful in specifying how many variables are involved and how they are related.

Definition 10

Univariate data arise when only a single characteristic of each sampled item is observed.

Definition 11

Multivariate data arise when observations are made on more than one characteristic of each sampled item. A special case of this involves two characteristics—**bivariate data**.

Definition 12

When multivariate data consist of several determinations of basically the same characteristic (e.g., made with different instruments or at different times), the data are called **repeated measures data**. In the special case of bivariate responses, the term **paired data** is used.

It is important to recognize the multivariate character of data when it is present. Having Rockwell hardness values for 5 of 100 crated machine parts and determinations of the percentage of carbon for 5 other parts is not at all equivalent to having both hardness and carbon content values for a single sample of 5 parts. There are two samples of 5 univariate data points in the first case and a single sample of 5 bivariate data points in the second. The second situation is preferable to the first, because it allows analysis and exploitation of any relationships that might exist between the variables Hardness and Percent Carbon.

Example 4

Paired Distortion Measurements

In the furnace-loading scenario discussed in Example 1, radial runout measurements were actually made on all $38 + 39 = 77$ gears both before and after heat treating. (Only after-treatment values were given in Table 1.1.) Therefore, the process engineer had two samples (of respective sizes 38 and 39) of paired data. Because of the pairing, the engineer was in the position of being able (if desired) to analyze how post-treatment distortion was correlated with pretreatment distortion.

1.2.3 Types of Data Structures

Statistical engineering studies are sometimes conducted to compare process performance at one set of conditions to a stated standard. Such investigations involve only one sample. But it is far more common for several sets of conditions to be compared with each other, in which case several samples are involved. There are a variety of standard notions of structure or organization for multisample studies. Two of these are briefly discussed in the remainder of this section.

Definition 13

A **(complete) factorial study** is one in which several process variables (and settings of each) are identified as being of interest, and data are collected under each possible combination of settings of the process variables. The process variables are usually called **factors**, and the settings of each variable that are studied are termed **levels** of the factor.

For example, suppose there are four factors of interest—call them A, B, C, and D for convenience. If A has 3 levels, B has 2, C has 2, and D has 4, a study that includes samples collected under each of the $3 \times 2 \times 2 \times 4 = 48$ different possible sets of conditions would be called a $3 \times 2 \times 2 \times 4$ factorial study.

Example 2 *(continued)*

Experimentation with the pelletizing machine produced data with a $2 \times 2 \times 2$ (or 2^3) factorial structure. The factors and respective levels studied were

Die Volume low volume vs. high volume

Material Flow current method vs. manual filling

Mixture Type no binding agent vs. with binder

Combining these then produced eight sets of conditions under which data were collected (see Table 1.2).

Table 1.2
Combinations in a 2^3 Factorial Study

Condition Number	Volume	Flow	Mixture
1	low	current	no binder
2	high	current	no binder
3	low	manual	no binder
4	high	manual	no binder
5	low	current	binder
6	high	current	binder
7	low	manual	binder
8	high	manual	binder

When many factors and/or levels are involved, the number of samples in a full factorial study quickly reaches an impractical size. Engineers often find that they want to collect data for only some of the combinations that would make up a complete factorial study.

Definition 14

A **fractional factorial study** is one in which data are collected for only some of the combinations that would make up a complete factorial study.

One cannot hope to learn as much about how a response is related to a given set of factors from a fractional factorial study as from the corresponding full factorial study. Some information must be lost when only part of all possible sets of conditions are studied. However, some fractional factorial studies will be potentially more informative than others. If only a fixed number of samples can be taken, which samples to take is an issue that needs careful consideration. Sections 8.3 and 8.4 discuss fractional factorials in detail, including how to choose good ones, taking into account what part of the potential information from a full factorial study they can provide.

Example 2 *(continued)*

The experiment actually carried out on the pelletizing process was, as indicated in Table 1.2, a full factorial study. Table 1.3 lists four experimental combinations, forming a well-chosen half of the eight possible combinations. (These are the combinations numbered 2, 3, 5, and 8 in Table 1.2.)

Table 1.3
Half of the 2^3 Factorial

Volume	Flow	Mixture
high	current	no binder
low	manual	no binder
low	current	binder
high	manual	binder

Section 2 Exercises

1. Describe a situation in your field where an observational study might be used to answer a question of real importance. Describe another situation where an experiment might be used.
2. Describe two different contexts in your field where, respectively, qualitative and quantitative data might arise.
3. What kind of information can be derived from a single sample of n bivariate data points (x, y) that can't be derived from two separate samples of, respectively, n data points x and n data points y?
4. Describe a situation in your field where paired data might arise.

5. Consider a study of making paper airplanes, where two different Designs (say, delta versus t wing), two different Papers (say, construction versus typing), and two different Loading Conditions (with a paper clip versus without a paper clip) are of interest in terms of their effects on flight distance. Describe a full factorial and then a fractional factorial data structure that might arise from such a study.
6. Explain why it is safer to infer causality from an experiment than from an observational study.

1.3 Measurement: Its Importance and Difficulty

Success in statistical engineering studies requires the ability to measure. For some physical properties like length, mass, temperature, and so on, methods of measurement are commonplace and obvious. Often, the behavior of an engineering system can be adequately characterized in terms of such properties. But when it cannot, engineers must carefully define what it is about the system that needs observing and then apply ingenuity to create a suitable method of measurement.

Example 5

Measuring Brittleness

A senior design class in metallurgical engineering took on the project of helping a manufacturer improve the performance of a spike-shaped metal part. In its intended application, this part needed to be strong but very brittle. When meeting an obstruction in its path, it had to break off rather than bend, because bending would in turn cause other damage to the machine in which the part functions.

As the class planned a statistical study aimed at finding what variables of manufacture affect part performance, the students came to realize that the company didn't have a good way of assessing part performance. As a necessary step in their study, they developed a measuring device. It looked roughly as in Figure 1.4. A swinging arm with a large mass at its end was brought to a

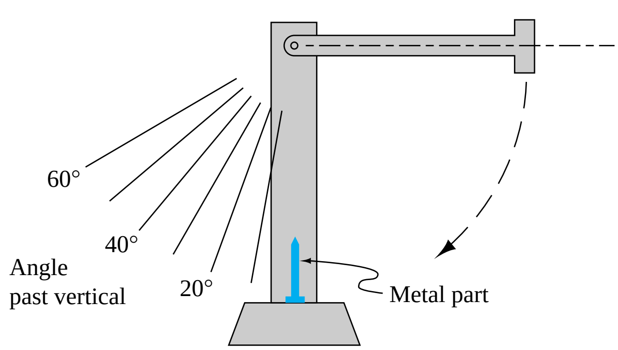

Figure 1.4 A device for measuring brittleness

horizontal position, released, and allowed to swing through a test part firmly fixed in a vertical position at the bottom of its arc of motion. The number of degrees past vertical that the arm traversed after impact with the part provided an effective measure of brittleness.

Example 6

Measuring Wood Joint Strength

Dimond and Dix wanted to conduct a factorial study comparing joint strengths for combinations of three different woods and three glues. They didn't have access to strength-testing equipment and so invented their own. To test a joint, they suspended a large container from one of the pieces of wood involved and poured water into it until the weight was sufficient to break the joint. Knowing the volume of water poured into the container and the density of water, they could determine the force required to break the joint.

Regardless of whether an engineer uses off-the-shelf technology or must fabricate a new device, a number of issues concerning measurement must be considered. These include **validity**, **measurement variation/error**, **accuracy**, and **precision**.

Definition 15
Validity

A measurement or measuring method is called **valid** if it usefully or appropriately represents the feature of an object or system that is of engineering importance.

It is impossible to overstate the importance of facing the question of measurement validity before plunging ahead in a statistical engineering study. Collecting engineering data costs money. Expending substantial resources collecting data, only to later decide they don't really help address the problem at hand, is unfortunately all too common.

Measurement error

The point was made in Section 1.1 that when using data, one is quickly faced with the fact that variation is omnipresent. Some of that variation comes about because the objects studied are never exactly alike. But some of it is due to the fact that measurement processes also have their own inherent variability. Given a fine enough scale of measurement, no amount of care will produce exactly the same value over and over in repeated measurement of even a single object. And it is naive to attribute all variation in repeat measurements to bad technique or sloppiness. (Of course, bad technique and sloppiness *can* increase measurement variation beyond that which is unavoidable.)

An exercise suggested by W. J. Youden in his book *Experimentation and Measurement* is helpful in making clear the reality of measurement error. Consider measuring the thickness of the paper in this book. The technique to be used is as follows. The book is to be opened to a page somewhere near the beginning and one somewhere near the end. The stack between the two pages is to be grasped firmly between the thumb and index finger and stack thickness read to the nearest .1 mm using an ordinary ruler. Dividing the stack thickness by the number of sheets in the stack and recording the result to the nearest .0001 mm will then produce a thickness measurement.

Example 7

Book Paper Thickness Measurements

Presented below are ten measurements of the thickness of the paper in Box, Hunter, and Hunter's *Statistics for Experimenters* made one semester by engineering students Wendel and Gulliver.

Wendel:	.0807, .0826, .0854, .0817, .0824, .0799, .0812, .0807, .0816, .0804
Gulliver:	.0972, .0964, .0978, .0971, .0960, .0947, .1200, .0991, .0980, .1033

Figure 1.5 shows a graph of these data and clearly reveals that even repeated measurements by one person on one book will vary and also that the patterns of variation for two different individuals can be quite different. (Wendel's values are both smaller and more consistent than Gulliver's.)

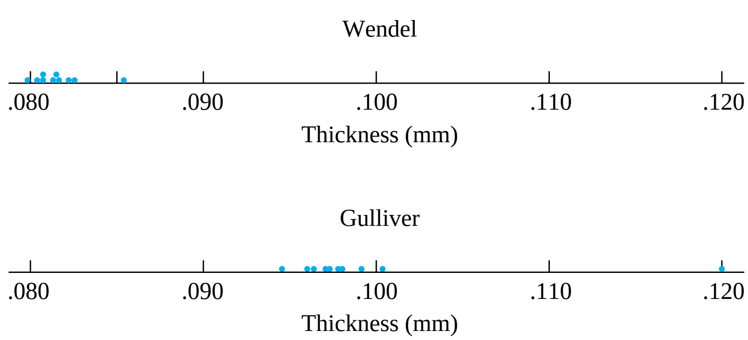

Figure 1.5 Dot diagrams of paper thickness measurements

The variability that is inevitable in measurement can be thought of as having both internal and external components.

Definition 16
Precision

A measurement system is called **precise** if it produces small variation in repeated measurement of the same object.

Precision is the internal consistency of a measurement system; typically, it can be improved only with basic changes in the configuration of the system.

Example 7 (*continued*)

Ignoring the possibility that some property of Gulliver's book was responsible for his values showing more spread than those of Wendel, it appears that Wendel's measuring technique was more precise than Gulliver's.

The precision of both students' measurements could probably have been improved by giving each a binder clip and a micrometer. The binder clip would provide a relatively constant pressure on the stacks of pages being measured, thereby eliminating the subjectivity and variation involved in grasping the stack firmly between thumb and index finger. For obtaining stack thickness, a micrometer is clearly a more precise instrument than a ruler.

Precision of measurement is important, but for many purposes it alone is not adequate.

Definition 17 *Accuracy*

A measurement system is called **accurate** (or sometimes, **unbiased**) if on average it produces the true or correct value of a quantity being measured.

Accuracy is the agreement of a measuring system with some external standard. It is a property that can typically be changed without extensive physical change in a measurement method. **Calibration** of a system against a standard (bringing it in line with the standard) can be as simple as comparing system measurements to a standard, developing an appropriate conversion scheme, and thereafter using converted values in place of raw readings from the system.

Example 7 (*continued*)

It is unknown what the industry-standard measuring methodology would have produced for paper thickness in Wendel's copy of the text. But for the sake of example, suppose that a value of .0850 mm/sheet was appropriate. The fact that Wendel's measurements averaged about .0817 mm/sheet suggests that her future accuracy might be improved by proceeding as before but then multiplying any figure obtained by the ratio of .0850 to .0817—i.e., multiplying by 1.04.

Maintaining the U.S. reference sets for physical measurement is the business of the National Institute of Standards and Technology. It is important business. Poorly calibrated measuring devices may be sufficient for local purposes of comparing local conditions. But to establish the values of quantities in any absolute sense, or to expect local values to have meaning at other places and other times, it is essential to calibrate measurement systems against a constant standard. A millimeter must be the same today in Iowa as it was last week in Alaska.

The possibility of bias or inaccuracy in measuring systems has at least two important implications for planning statistical engineering studies. First, the fact that

Accuracy and statistical studies

measurement systems can lose accuracy over time demands that their performance be monitored over time and that they be recalibrated as needed. The well-known phenomenon of **instrument drift** can ruin an otherwise flawless statistical study. Second, whenever possible, a single system should be used to do all measuring. If several measurement devices or technicians are used, it is hard to know whether the differences observed originate with the variables under study or from differences in devices or technician biases. If the use of several measurement systems is unavoidable, they must be calibrated against a standard (or at least against each other). The following example illustrates the role that human differences can play.

Example 8

Differences Between Technicians in Their Use of a Gauge

Cowan, Renk, Vander Leest, and Yakes worked with a company on the monitoring of a critical dimension of a high-precision metal part produced on a computer-controlled lathe. They encountered large, initially unexplainable variation in this dimension between different shifts at the plant. This variation was eventually traced not to any real shift-to-shift difference in the parts but to an instability in the company's measuring system. A single gauge was in use on all shifts, but different technicians used it quite differently when measuring the critical dimension. The company needed to train the technicians in a single, standardized method of using the gauge.

An analogy that is helpful in understanding the difference between precision and accuracy involves comparing measurement to target shooting. In target shooting, one can be on or off target (accurate or inaccurate) with a small or large cluster of shots (showing precision or imprecision). Figure 1.6 illustrates this analogy.

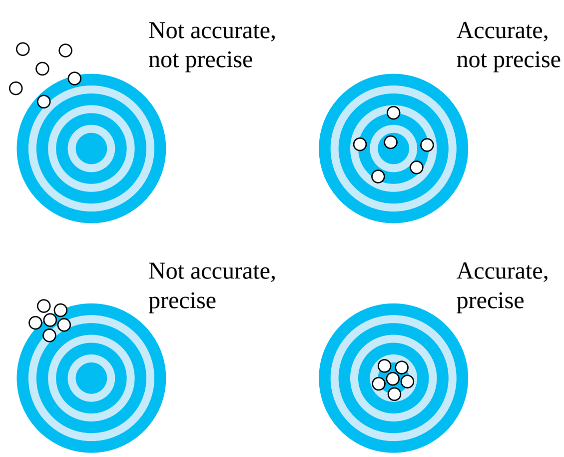

Figure 1.6 Measurement/Target shooting analogy

Good measurement is hard work, but without it data collection is futile. To make progress, engineers must obtain valid measurements, taken by methods whose precision and accuracy are sufficient to let them see important changes in system behavior. Usually, this means that measurement inaccuracy and imprecision must be an order of magnitude smaller than the variation in measured response caused by those changes.

Section 3 Exercises

1. Why might it be argued that in terms of producing useful measurements, one must deal first with the issue of validity, then the issue of precision, and only then the issue of accuracy?
2. Often, in order to evaluate a physical quantity (for example, the mean yield of a batch chemical process run according to some standard plant operating procedures), a large number of measurements of the quantity are made and then averaged. Explain which of the three aspects of measurement quality—validity, precision, and accuracy—this averaging of many measurements can be expected to improve and which it cannot.
3. Explain the importance of the stability of the measurement system to the real-world success of a statistical engineering study.

1.4 Mathematical Models, Reality, and Data Analysis

This is not a book on mathematics. Nevertheless, it contains a fair amount of mathematics (that most readers will find to be reasonably elementary—if unfamiliar and initially puzzling). Therefore, it seems wise to try to put the mathematical content of the book in perspective early. In this section, the relationships of mathematics to the physical world and to engineering statistics are discussed.

Mathematical models and reality

Mathematics is a construct of the human mind.While it is of interest to some people in its own right, engineers generally approach mathematics from the point of view that it can be useful in describing and predicting how physical systems behave. Indeed, although they exist only in our minds, mathematical theories are guides in every branch of modern engineering.

Throughout this text, we will frequently use the phrase *mathematical model*.

Definition 18

> A **mathematical model** is a description or summarization of salient features of a real-world system or phenomenon in terms of symbols, equations, numbers, and the like.

Mathematical models are themselves not reality, but they can be extremely effective descriptions of reality. This effectiveness hinges on two somewhat opposing properties of a mathematical model: (1) its degree of **simplicity** and (2) its **predictive**

ability. The most powerful mathematical models are those that simultaneously are simple *and* generate good predictions. A model's simplicity allows one to maneuver within its framework, deriving mathematical consequences of basic assumptions that translate into predictions of process behavior. When these are empirically correct, one has an effective engineering tool.

The elementary "laws" of mechanics are an outstanding example of effective mathematical modeling. For example, the simple mathematical statement that the acceleration due to gravity is constant,

$$a = g$$

yields, after one easy mathematical maneuver (an integration), the prediction that beginning with 0 velocity, after a time t in free fall an object will have velocity

$$v = gt$$

And a second integration gives the prediction that beginning with 0 velocity, a time t in free fall produces displacement

$$d = \frac{1}{2}gt^2$$

The beauty of this is that for most practical purposes, these easy predictions are quite adequate. They agree well with what is observed empirically and can be counted on as an engineer designs, builds, operates, and/or improves physical processes or products.

Mathematics and statistics

But then, how does the notion of mathematical modeling interact with the subject of engineering statistics? There are several ways. For one, data collection and analysis are essential in **fitting or estimating parameters** of mathematical models. To understand this point, consider again the example of a body in free fall. If one postulates that the acceleration due to gravity is constant, there remains the question of what numerical value that constant should have. The parameter g must be evaluated before the model can be used for practical purposes. One does this by gathering data and using them to estimate the parameter.

A standard first college physics lab has traditionally been to empirically evaluate g. The method often used is to release a steel bob down a vertical wire running through a hole in its center and allowing 60-cycle current to arc from the bob through a paper tape to another vertical wire, burning the tape slightly with every arc. A schematic diagram of the apparatus used is shown in Figure 1.7. The vertical positions of the burn marks are bob positions at intervals of $\frac{1}{60}$ of a second. Table 1.4 gives measurements of such positions. (We are grateful to Dr. Frank Peterson of the ISU Physics and Astronomy Department for supplying the tape.) Plotting the bob positions in the table at equally spaced intervals produces the approximately quadratic plot shown in Figure 1.8. Picking a parabola to fit the plotted points involves identifying an appropriate value for g. A method of curve fitting (discussed in Chapter 4) called *least squares* produces a value for g of 9.79m/sec^2, not far from the commonly quoted value of 9.8m/sec^2.

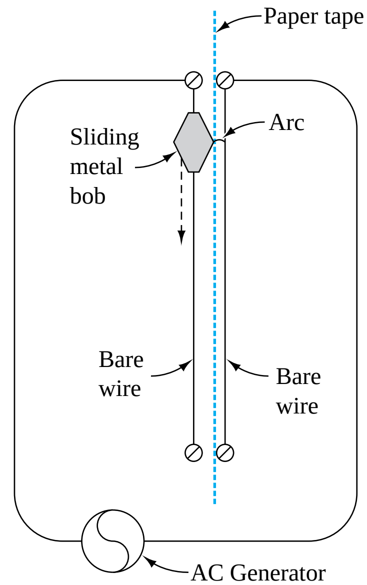

Figure 1.7 A device for measuring g

Table 1.4
Measured Displacements of a Bob in Free Fall

Point Number	Displacement (mm)	Point Number	Displacement (mm)
1	.8	13	223.8
2	4.8	14	260.0
3	10.8	15	299.2
4	20.1	16	340.5
5	31.9	17	385.0
6	45.9	18	432.2
7	63.3	19	481.8
8	83.1	20	534.2
9	105.8	21	589.8
10	131.3	22	647.7
11	159.5	23	708.8
12	190.5		

Notice that (at least before Newton) the data in Table 1.4 might also have been used in another way. The parabolic shape of the plot in Figure 1.8 could have suggested the form of an appropriate model for the motion of a body in free fall. That is, a careful observer viewing the plot of position versus time should conclude that there is an approximately quadratic relationship between position and time (and

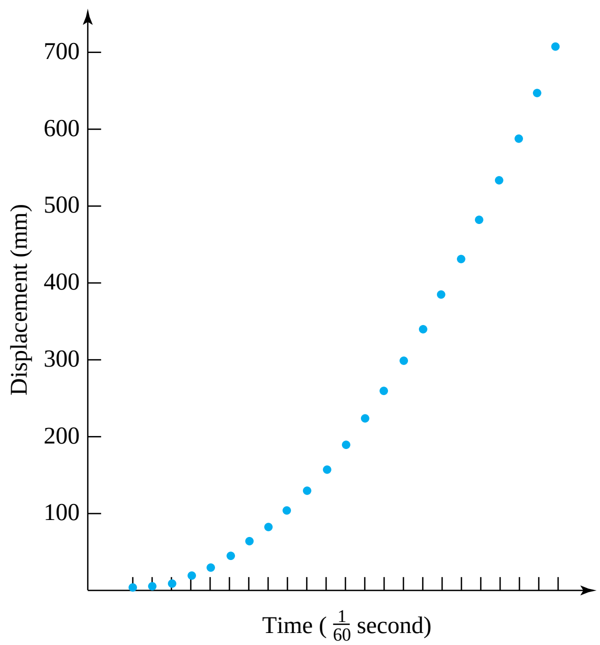

Figure 1.8 Bob positions in free fall

from that proceed via two differentiations to the conclusion that the acceleration due to gravity is roughly constant). This text is full of examples of how helpful it can be to use data both to identify potential forms for empirical models and to then estimate parameters of such models (preparing them for use in prediction).

This discussion has concentrated on the fact that statistics provides raw material for developing realistic mathematical models of real systems. But there is another important way in which statistics and mathematics interact. The mathematical theory of probability provides a framework for quantifying the uncertainty associated with inferences drawn from data.

Definition 19

Probability is the mathematical theory intended to describe situations and phenomena that one would colloquially describe as involving chance.

If, for example, five students arrive at the five different laboratory values of g,

$$9.78,\ 9.82,\ 9.81,\ 9.78,\ 9.79$$

questions naturally arise as to how to use them to state both a best value for g and some measure of precision for the value. The theory of probability provides guidance in addressing these issues. Material in Chapter 6 shows that probability considerations support using the class average of 9.796 to estimate g and attaching to it a precision on the order of plus or minus .02m/sec^2.

We do not assume that the reader has studied the mathematics of probability, so this text will supply a minimal introduction to the subject. But do not lose sight of the fact that probability is not statistics—nor vice versa. Rather, probability is a branch of mathematics and a useful subject in its own right. It is met in a statistics course as a tool because the variation that one sees in real data is closely related conceptually to the notion of chance modeled by the theory of probability.

Section 4 Exercises

1. Explain in your own words the importance of mathematical models to engineering practice.

Chapter 1 Exercises

1. Calibration of measurement equipment is most clearly associated with which of the following concepts: validity, precision, or accuracy? Explain.
2. If factor A has levels 1, 2, and 3, factor B has levels 1 and 2, and factor C has levels 1 and 2, list the combinations of A, B, and C that make up a full factorial arrangement.
3. Explain how paired data might arise in a heat treating study aimed at determining the best way to heat treat parts made from a certain alloy.
4. Losen, Cahoy, and Lewis purchased eight spanner bushings of a particular type from a local machine shop and measured a number of characteristics of these bushings, including their outside diameters. Each of the eight outside diameters was measured once by two student technicians, with the following results. (The units are inches.) Considering both students' measurements, what type of data are given here? Explain.

Bushing	**1**	**2**	**3**	**4**
Student A	.3690	.3690	.3690	.3700
Student B	.3690	.3695	.3695	.3695
Bushing	**5**	**6**	**7**	**8**
Student A	.3695	.3700	.3695	.3690
Student B	.3695	.3700	.3700	.3690

5. Describe a situation from your field where a full factorial study might be conducted (name at least three factors, and the levels of each, that would appear in the study).
6. Example 7 concerns the measurement of the thickness of book paper. Variation in measurements is a fact of life. To observe this reality firsthand, measure the thickness of the paper used in this book ten times. Use the method described immediately before Example 7. For each determination, record the measured stack thickness, the number of sheets, and the quotient to four decimal places. If you are using this book in a formal course, be prepared to hand in your results and compare them with the values obtained by others in your class.
7. Exercise 6 illustrates the reality of variation in physical measurement. Another exercise that is similar in spirit, but leads to qualitative data, involves the spinning of U.S. pennies. Spin a penny on a hard surface 20 different times; for each trial, record whether the penny comes to rest with heads or tails showing. Did all the trials have the same outcome? Is the pattern you observed the one you expected to see? If not, do you have any possible explanations?

8. Consider a situation like that of Example 1 (involving the heat treating of gears). Suppose that the original gears can be purchased from a variety of vendors, they can be made out of a variety of materials, they can be heated according to a variety of regimens (involving different times and temperatures), they can be cooled in a number of different ways, and the furnace atmosphere can be adjusted to a variety of different conditions. A number of features of the final gears are of interest, including their flatness, their concentricity, their hardness (both before and after heat treating), and their surface finish.

(a) What kind of data arise if, for a single set of conditions, the Rockwell hardness of several gears is measured both before and after heat treating? (Use the terminology of Section 1.2.) In the same context, suppose that engineering specifications on flatness require that measured flatness not exceed .40 mm. If flatness is measured for several gears and each gear is simply marked Acceptable or Not Acceptable, what kind of data are generated?

(b) Describe a three-factor full factorial study that might be carried out in this situation. Name the factors that will be used and describe the levels of each. Write out a list of all the different combinations of levels of the factors that will be studied.

9. Suppose that you wish to determine “the” axial strength of a type of wooden dowel. Why might it be a good idea to test several such dowels in order to arrive at a value for this “physical constant”?

10. Give an example of a 2×3 full factorial data structure that might arise in a student study of the breaking strengths of wooden dowels. (Name the two factors involved, their levels, and write out all six different combinations.) Then make up a data collection form for the study. Plan to record both the breaking strength and whether the break was clean or splintered for each dowel, supposing that three dowels of each type are to be tested.

11. You are a mechanical engineer charged with improving the life-length characteristics of a hydrostatic transmission. You suspect that important variables include such things as the hardnesses, diameters and surface roughnesses of the pistons and the hardnesses, and inside diameters and surface roughnesses of the bores into which the pistons fit. Describe, in general terms, an observational study to try to determine how to improve life. Then describe an experimental study and say why it might be preferable.

12. In the context of Exercise 9, it might make sense to average the strengths you record. Would you expect such an average to be more or less precise than a single measurement as an estimate of the average strength of this kind of dowel? Explain. Argue that such averages can be no more (or less) accurate than the individual measurements that make them up.

13. A toy catapult launches golf balls. There are a number of things that can be altered on the configuration of the catapult: The length of the arm can be changed, the angle the arm makes when it hits the stop can be changed, the pull-back angle can be changed, the weight of the ball launched can be changed, and the place the rubber cord (used to snap the arm forward) is attached to the arm can be changed. An experiment is to be done to determine how these factors affect the distance a ball is launched.

(a) Describe one three-factor full factorial study that might be carried out. Make out a data collection form that could be used. For each launch, specify the level to be used of each of the three factors and leave a blank for recording the observed value of the response variable. (Suppose two launches will be made for each setup.)

(b) If each of the five factors mentioned above is included in a full factorial experiment, a minimum of how many different combinations of levels of the five factors will be required? If there is time to make only 16 launches with the device during the available lab period, but you want to vary all five factors, what kind of a data collection plan must you use?

14. As a variation on Exercise 6, you could try using only pages in the first four chapters of the book. If there were to be a noticeable change in the ultimate precision of thickness measurement, what kind of a change would you expect? Try this out by applying the method in Exercise 6 ten times to stacks of pages from only the first four chapters. Is there a noticeable difference in precision of measurement from what is obtained using the whole book?