STATISTICAL METHODS FOR QUALITY ASSURANCE: Basics, Measurement, Control, Capability, and Improvement

Stephen B. Vardeman and J. Marcus Jobe

September 27, 2007

Contents

Preface

This is the preface. More here later.

CHAPTER 1

Introduction

This opening chapter first introduces the subject of quality assurance and the relationship between it and the subject of statistics in Section 1.1. Then Section 1.2 provides context for the material of this book. Standard emphases in modern quality assurance are introduced and a six-step process-oriented quality assurance cycle is put forward as a framework for approaching projects in this field. Some connections between modern quality assurance and popular business process improvement programs are discussed next. Some of the simplest quality assurance tools are then introduced in Sections 1.3 through 1.5 . There is a brief discussion of process mapping/analysis in Section 1.3,.discussion of some simple principles of quality assurance data collection follows in Section 1.4, and simple statistical graphics are considered in Section 1.5.

1.1 The Nature of Quality and the Role of Statistics

This book's title raises at least two basic questions: "What is 'quality'?" and "What do 'statistical methods' have to do with assuring it?"

Consider first the word "quality." What does it mean to say that a particular good is a quality product? And what does it mean to call a particular service a quality service? In the case of manufactured goods (like automobiles and dishwashers), issues of reliability (the ability to function consistently and effectively across time), appropriateness of configuration, and fit and finish of parts come to mind. In the realm of services (like telecommunications and transportation services) one thinks of consistency of availability and performance, esthetics, and convenience. And in evaluating the "quality" of both goods and services, there is an implicit understanding that these issues will be

balanced against corresponding costs to determine overall "value." Here is a popular definition of quality that reflects some of these notions.

Definition 1 ***Quality*** *in a good or service is fitness for use. That fitness includes aspects of both product design and conformance to the (ideal) design.*

Quality of design has to do with appropriateness; the choice and configuration of features that define what a good or service is supposed to be like and is supposed to do. In many cases it is essentially a matter of matching product "species" to an arena of use. One needs different things in a vehicle driven on the dirt roads of the Baja peninsula than in one used on the German autobahn. Vehicle quality of design has to do with providing the "right" features at an appropriate price. With this understanding, there is no necessary contradiction between thinking of both a Rolls Royce and a Toyota economy car as quality vehicles. Similarly, both a particular fast food outlet and a particular four star restaurant might be thought of as quality eateries.

Quality of conformance has to do with living up to specifications laid down in product design. It is concerned with small variation from what is specified or expected. Variation inevitably makes goods and services undesirable. Mechanical devices whose parts vary substantially from their ideal/design dimensions tend to be noisy, inefficient, prone to breakdown, and difficult to service. They simply don't work well. In the service sector, variation from what is promised/expected is the principal source of customer dissatisfaction. A city bus system that runs on schedule every day that it is supposed to run can be seen as a quality transportation system. One that fails to do so cannot. And an otherwise elegant hotel that fails to ensure the spotless bathrooms its customers expect will soon be without those customers.

This book is concerned primarily with tools for assuring quality of conformance. This is not because quality of design is unimportant. Designing effective goods and services is a highly creative and important activity. But it is just not the primary topic of this text.

Then what does the subject of statistics have to do with the assurance of quality of conformance? To answer this question, it is helpful to have clearly in mind a definition of statistics.

Definition 2 ***Statistics*** *is the study of how best to*

1. *collect data,*
2. *summarize or describe data, and*
3. *draw conclusions or inferences based on data,*

all in a framework that recognizes the reality and omnipresence of variation.

If quality of conformance has to do with small variation and one wishes to assure it, it will be necessary to measure, monitor, find sources of, and seek ways to reduce variation. All of these require data (information on what is happening in a system producing a product) and therefore the tool of statistics. The intellectual framework

of the subject of statistics, emphasizing as it does the concept of variation, makes it a natural for application in the world of quality assurance. We will see that both simple and also somewhat more advanced methods of statistics have their uses in the quest to produce quality goods and services.

Section 1.1 Exercises

1. "Quality" and "statistics" are related. Briefly explain this relationship, using the definitions of both words.
2. Why is variation in manufactured parts undesirable? Why is variation undesirable in a service industry?
3. If a product or service is designed appropriately, does that alone guarantee quality? Why or why not?
4. If a product or service conforms to design specifications, does that alone guarantee quality? Why or why not?

1.2 Modern Quality Philosophy and Business Practice Improvement Strategies

The global business environment is fiercely competitive. No company can afford to "stand still" if it hopes to stay in business. Every healthy company has explicit strategies for constantly improving its business processes and products.

Over the past several decades, there has been a blurring of distinctions between "quality improvement" and "general business practice improvement." (Formerly, the first of these was typically thought of as narrowly focused on characteristics of manufactured goods.) So there is now much overlap in emphases, language, and methodologies between the areas. The best strategies in both arenas must in the end boil down to good methodical/scientific data-based problem solving.

In this section we first provide a discussion of some elements of modern quality philosophy and an intellectual framework around which we have organized the topics of this book (and that can serve as a road map for approaching quality improvement projects). We then provide some additional discussion and critique of the modern general business environment and its better known process improvement strategies.

1.2.1 Modern Quality Philosophy and a Six-Step Process-Oriented Quality Assurance Cycle

Modern quality assurance methods and philosophy are focused not (primarily) on products, but rather on the **processes** used to produce them. The idea is that if one gets

processes to work effectively, resulting products will automatically be good. On the other hand, if one only focuses on screening out or reworking bad product, root causes of quality problems are never discovered or eliminated. The importance of this process orientation can be illustrated by an example.

Example 3 ***Process Improvement in a Clean Room***. *One of the authors of this text once toured a "clean room" at a division of a large electronics manufacturer. Integrated circuit (IC) chips critical to the production of the division's most important product were made in the room and it was the bottleneck of the whole production process for that product. Initial experience with that (very expensive) facility included 14% yields of good IC chips, with over 80 people working there trying to produce the precious components.*

Early efforts at quality assurance for these chips centered on final testing and sorting good ones from bad. But it was soon clear that those efforts alone would not produce yields adequate to supply the numbers of chips needed for the end product. So a project team went to work on improving the production process. The team found that by carefully controlling the quality of some incoming raw materials, adjusting some process variables, and making measurements on wafers of chips early in the process (aimed at identifying and culling ones that would almost certainly in the end consist primarily of bad chips) the process could be made much more efficient. At the time of the tour, process improvement efforts had raised yields to 65% (effectively quadrupling production capacity with no capital expenditure!), drastically reduced material waste, and cut the staff necessary to run the facility from the original 80 to only eight technicians. Process-oriented efforts are what enabled this success story. No amount of attention to the yield of the process as it was originally running would have produced these important results.

It is important to note that while process-oriented quality improvement efforts have center stage, product-oriented methods still have their place. In the clean room of Example 3, process improvement efforts in no way eliminated the need for end-of-the-line testing of the IC chips. Occasional bad chips still needed to be identified and culled. Product-oriented inspection was still necessary, but it alone was not sufficient to produce important quality improvements.

A second important emphasis of modern quality philosophy is its **customer orientation**. This has two faces. First, the final or end user of a good or service is viewed as being supremely important. Much effort is expended by modern corporations in seeing that the "voice of the customer" (the will of the end user) is heard and carefully considered in all decisions involved in product design and production. There are many communication and decision-making techniques (such as "quality function deployment") that are used to see that this happens.

But the customer orientation in modern quality philosophy extends beyond concentration on an end user. All workers are taught to view their efforts in terms of processes that have both "vendors" from whom they receive input and "customers" to whom they pass work. One's most immediate customer need not be the end user of a company

product. But it is still important to do one's work in a way that those who handle one's personal "products" are able to do so without difficulties.

A third major emphasis in modern quality assurance is that of **continual improvement**. What is state-of-art today will be woefully inadequate tomorrow. Consumers are expecting (and getting!) ever more effective computers, cars, home entertainment equipment, package delivery services, and communications options. Modern quality philosophy says that this kind of improvement must and will continue. This is both a statement of what "ought" to be, and a recognition that in a competitive world, if an organization does not continually improve what it does and makes, it will not be long before aggressive competition drives it from the marketplace.

This text presents a wide array of tools for quality assurance. But students do not always immediately see where they might fit into a quality assurance/improvement effort or how to begin a class project in the area. So, it is useful to present an outline for approaching modern quality assurance that places the methods of this book into their appropriate context. Table 1.1 on page 6 presents a six-step process-oriented quality assurance cycle (that is the intellectual skeleton of this book) and the corresponding technical tools we discuss.

A sensible first step in any quality improvement project is to attempt to thoroughly understand the current and ideal configurations of the processes involved. This matter of *process mapping* can be aided by very simple tools like the flowcharts and Ishikawa diagrams discussed in Section 1.3.

Effective measurement is foundational to efforts to improve processes and products. If one cannot reliably measure important characteristics of what is being done to produce a good or service, there is no way to tell whether design requirements are being met and customer needs genuinely addressed. Chapter 2 introduces some basic concepts of metrology and statistical methodology for quantifying and improving the performance of measurement systems.

When adequate measurement systems are in place, one can begin to collect data on process performance. But there are pitfalls to be avoided in this collection, and if data are to be genuinely helpful in addressing quality assurance issues, they typically need to be summarized and presented effectively. So Sections 1.4 and 1.5 contain discussions of some elementary principles of quality assurance data collection and effective presentation of such data.

Once one recognizes uniformity as essentially synonymous with quality of conformance (and variation as synonymous with "unquality"), one wants processes to be perfectly consistent in their output. But that is too much to hope for in the real world. Variation is a fact of life. The most that one can expect is that a process be consistent in its pattern of variation, that it be describable as physically stable. Control charts are tools for monitoring processes and issuing warnings when there is evidence in process data of physical instability. These essential tools of quality assurance are discussed in Chapter 3.

Even those processes that can be called physically stable need not be adequate for current or future needs. (Indeed modern quality philosophy views *all* processes as inadequate and in need of improvement!) So it is important to be able to characterize

TABLE 1.1. A Six-Step Process-Oriented Quality Assurance Cycle (and Corresponding Tools)

	Step	Tools
1.	Attempt a logical analysis of how a process works (or should work) and where potential trouble spots, sources of variation, and data needs are located.	• Flowcharts (§1.3) • Ishikawa/fishbone/cause-and-effect diagrams (§1.3)
2.	Formulate appropriate (customer-oriented) measures of process performance and develop corresponding measurement systems.	• Basic concepts of measurement/ metrology (Ch. 2) • Statistical quantification of measurement precision (Ch. 2) • Regression and calibration (Ch. 2)
3.	Habitually collect and summarize process data.	• Simple quality assurance data collection principles (§1.4) • Simple statistical graphics (§1.5)
4.	Assess and work toward process stability.	• Control charts (Ch. 3)
5.	Characterize current process and product performance.	• Statistical graphics for process characterization (§4.1) • Measures of process capability and performance and their estimation (§4.2, §4.3) • Probabilistic tolerancing and propagation of error (§4.4) • Estimation of variance components (§4.5)
6.	Work to improve those processes that are unsatisfactory.	• Design and analysis of experiments (Ch. 5, Ch. 6)

TABLE 1.2. Elements of TQM Emphasis

1.	Customer focus
2.	Process/system orientation
3.	Continuous improvement
4.	Self-assessment and benchmarking
5.	Change to flat organizations "without barriers"
6.	"Empowered" people/teams and employee involvement
7.	Management (and others') commitment (to TQM)
8.	Appreciation/understanding of variability

in precise terms what a process is currently doing and to have tools for finding ways of improving it. Chapter 4 of this text discusses a number of methods for quantifying current process and product performance, while Chapters 5 and 6 deal with methods of experimental design and analysis especially helpful in process improvement efforts.

The steps outlined in Table 1.1 are a useful framework for approaching most process-related quality assurance projects. They are presented here not only as a road map for this book, but also as a list of steps to follow for students wishing to get started on a class project in process-oriented quality improvement.

1.2.2 The Modern Business Environment and General Business Process Improvement

Intense global competition has fueled a search for tools to use in improving all aspects of what modern companies do. At the same time, popular understanding of the realm of "quality assurance" has broadened substantially in the past few decades. As a result, distinctions between what is the improvement of general business practice and what is process-oriented quality improvement have blurred. General business emphases and programs like Total Quality Management, ISO 9000 certification, Malcolm Baldrige Prize competitions, and Six Sigma programs have much in common with the kind of quality philosophy just discussed.

TQM

Take for example, "TQM," an early instance of the broad business influence of modern quality philosophy. The name *Total Quality Management* was meant to convey the notion that in a world economy, successful organizations will *manage* the *total*ity of what they do with a view toward producing *quality* work. TQM was promoted as appropriate in areas as diverse as manufacturing, education, and government. The matters listed in Table 1.2 came up most frequently when TQM was discussed.

Items 1,2, and 3 in Table 1.2 are directly related to the emphases of modern quality assurance discussed above. The TQM process orientation in 2 is perhaps a bit broader than the discussion of the previous subsection, as it sees an organization's many processes fitting together in a large **system**. (The billing process needs to mesh with various production processes, which need to mesh with the product-development

process, which needs to mesh with the sales process, and so on.) There is much planning and communication needed to see that these work together in harmony within a single organization. But there is also recognition that *other* organizations, external suppliers and customers, need to be seen as part of "the system." A company's products can be only as good as the raw materials with which it works. TQM thus emphasized involving a broader and broader "superorganization" (our terminology) in process- and system-improvement efforts.

In support of continual improvement, TQM proponents emphasized knowing what the "best-in-class" practices are for a given business sector or activity. They promoted **benchmarking activities** to find out how an organization's techniques compare to the best in the world. Where an organization was found to be behind, every effort was to be made to quickly emulate the leader's performance. (Where an organization's methodology is state of the art, opportunities for yet another quantum improvement were to be considered.)

It was standard TQM doctrine that the approach could only be effective in organizations that are appropriately structured and properly unified in their acceptance of the viewpoint. Hence, there was a strong emphasis in the movement on **changing corporate cultures and structures** to enable this effectiveness. Proponents of TQM simultaneously emphasized the importance of involving all corporate citizens in TQM activities, beginning with the highest levels of management, and at the same time reducing the number of layers between the top and bottom of an organization, making it more egalitarian. Cross-functional project teams composed of employees from various levels of an organization (operating in consensus-building modes, with real authority not only to suggest changes but to see that they were implemented, and drawing on the various kinds of wisdom resident in the organization) were standard TQM fare. One of the corporate evils most loudly condemned was the human tendency to create "little empires" inside an organization that in fact compete with each other, rather than cooperate in ways that are good for the organization as a whole.

In a dimension most closely related to the subject of statistics, the TQM movement placed emphasis on understanding and appreciating the consequences of **variability.** In fact, providing training in elementary statistics (including the basics of describing variation through numerical and graphical means, and often some basic Shewhart control charting) was a typical early step in most TQM programs.

TQM had its big names like W.E. Deming, J.M. Juran, A.V. Feigenbaum, and P. Crosby. There were also thousands of less famous individuals, who in some cases provided guidance in implementing the ideas of more famous quality leaders, and in others provided instruction in their own modifications of the systems of others. The sets of terminology and action items promoted by this diverse set of individuals varied consultant to consultant, in keeping with the need for them to have unique products to sell.

Six Sigma

Fashions change and business interest in some of the more managerial emphases of TQM have waned. But interest in business process improvement has not. One particu-

larly popular and long-lived form of corporate improvement emphasis goes under the name "**Six Sigma**." The name originated at Motorola Corporation in the late 1980's. Six Sigma programs at General Electric, AlliedSignal and Dow Chemical (among other leading examples) have been widely touted as at least partially responsible for important growth in profits and company stock values. So huge interest in Six Sigma programs persists.

The name "Six Sigma" is popularly used in at least three different ways. It refers to:

1. a goal for business process performance,
2. a strategy for achieving that performance for all of a company's processes, and
3. an organizational, training and recognition program designed to support and implement the strategy referred to in 2.

As a goal for process performance, the "Six Sigma" name has a connection to the normal distribution. If a (normal) process mean is set 6σ inside specifications/requirements (even should it inadvertently drift a bit, say by as much as 1.5σ) the process produces essentially no unacceptable results. As a formula for organizing and training to implement universal process improvement, Six Sigma borrows from the culture of the martial arts. Properly trained and effective individuals are designated as "black belts," "master black belts," and so on. These individuals with advanced training and demonstrated skills lead company process improvement teams.

Here, our primary interest is in item 2 in the foregoing list. Most Six Sigma programs use the acronym DMAIC and the corresponding steps

1. **D**efine
2. **M**easure
3. **A**nalyze
4. **I**mprove
5. **C**ontrol

as a framework for approaching process improvement. The *Define* step involves setting the boundaries of a particular project, laying out the scope of what is to be addressed, and bringing focus to a general "we need to work on X" beginning. The *Measure* step requires finding appropriate responses to observe, identifying corresponding measurement systems, and collecting initial process data. The *Analyze* step involves producing data summaries and formal inferences adequate to make clear initial process performance. After seeing how a process is operating, there comes an *Improve*ment effort. Often this is guided by experimentation and additional data collected to see the effects of implemented process changes. Further, there is typically an emphasis on variation reduction (improvement in process consistency). Finally, the Six Sigma 5-step cycle culminates in process *Control*. This means process watching/monitoring through the

TABLE 1.3. DMAIC and Statistics

Element	Statistical Topics
Measure	• Measurement concepts • Data collection principles • Regression and linear calibration • Modeling measurement error • Inference in measurement precision studies
Analyze	• Descriptive statistics • Normal plotting and capability indices • Statistical intervals and testing • Confidence intervals and testing
Improve	• Regression analysis and response surface methods • Probabilistic tolerancing • Confidence intervals and testing • Factorial and fractional factorial analysis
Control	• Shewhart control charts

routine collection of and attention to process data. The point is to be sure that improvements made persist over time. Like this book's six step process oriented quality assurance cycle in Table 1.1, the Six Sigma 5-step DMAIC cycle is full of places where statistics is important. Table 1.3 shows where some standard statistical concepts and methods fit into the DMAIC paradigm.

1.2.3 Some Caveats

This book is primarily about technical tools, not philosophy. Nevertheless, some comments about proper context are in order before launching into the technical discussion. It may at first seem hard to imagine anything unhappy issuing from an enthusiastic universal application of quality philosophy and process improvement methods. Professor G. Box, for example, referred to TQM in such positive terms as "the democratization of science." Your authors are generally supportive of the emphases of quality philosophy and process improvement *in the realm of commerce*. But it is possible to lose perspective, and by applying them where they are not really appropriate, to create unintended and harmful consequences.

Consider first the matter of "customer focus." To become completely absorbed with what some customers want amounts to embracing them as the final arbiters of what is to be done. And that is a basically amoral (or ultimately immoral) position. This point holds in the realm of commerce, but is even more obvious when a customer-focus paradigm is applied in areas other than business.

For example, it is laudable to try to make government or educational systems more efficient. But these institutions deal in fundamentally moral arenas. We should want governments to operate morally, whether or not that is currently in vogue with the majority of (customer) voters. People should want their children to go to schools where serious content is taught, real academic achievement is required, and depth of char-

acter and intellect are developed, whether or not it is a "feel-good" experience and popular with the (customer) students, or satisfies the job-training desires of (customer) business concerns. Ultimately, we should fear for a country whose people expect other individuals and all public institutions to immediately gratify their most trivial whims (as deserving customers). The whole of human existence is not economics and commerce. Big words and concepts like "self-sacrifice," "duty," "principle," "integrity," and so on have little relevance in a "customer-driven" world. What "the customer" wants is not always even consistent, let alone moral or wise.

Preoccupation with the analysis and improvement of processes and systems has already received criticism in business circles, as often taking on a life of its own and becoming an end in itself, independent of the fundamental purposes of a company. Rationality is an important part of the human standard equipment and it is only good stewardship to be moderately organized about how things are done. But enough is enough. The effort and volume of reporting connected with planning (and documentation of that planning) and auditing (what has been done in every conceivable matter) has increased exponentially in the past few years in American business, government, and academia. What is happening in many cases amounts to a monumental triumph of form over substance. In a sane environment, smart and dedicated people will naturally do reasonable things. Process improvement tools are sometimes helpful in thinking through a problem. But slavish preoccupation with the details of how things are done and endless generation of vision and mission statements, strategic plans, process analyses, outcome assessments, and so forth can turn a relatively small task for one person into a big one for a group, with an accompanying huge loss of productivity.

There are other aspects of emphases on the analysis of processes, continuous improvement, and the benchmarking notion that deserve mention. A preoccupation with formal benchmarking has the natural tendency to produce homogenization and the stifling of genuine creativity and innovation. When an organization invests a large effort in determining what others are doing, it is very hard to then turn around and say "So be it. That's not what we're about. That doesn't suit our strengths and interests. We'll go a different way." Instead, the natural tendency is to conform, to "make use" of the carefully gathered data and strive to be like others. And frankly, the tools of process-analysis applied in endless staff meetings are not the stuff of which first-order innovations are born. Rather, those almost always come from really bright and motivated people working hard on a problem *individually* and perhaps occasionally coming together for free-form discussions of what they've been doing and what might be possible.

In the end, one has in the quality philosophy and process improvement emphases introduced above a sensible set of concerns, *provided* they are used in limited ways, in appropriate arenas, by ethical and thinking people.

Section 1.2 Exercises

1. A "process orientation" is one of the primary emphases of modern quality assur-

ance. What is the rationale behind this?

2. How does a "customer focus" relate to "quality"?
3. What are motivations for a corporate "continuous improvement" emphasis?
4. Why is effective measurement a prerequisite to success in process improvement?
5. What tools are used for monitoring processes and issuing warnings of apparent process instability?
6. If a process is stable or consistent, is it necessarily producing high quality goods or services? Why or why not?

1.3 Logical Process Identification and Analysis

Often, simply comparing "what is" in terms of process structure to "what is supposed to be" or to "what would make sense" is enough to identify opportunities for real improvement. Particularly in service industry contexts, the mapping of a process and identification of redundant and unnecessary steps can often lead very quickly to huge reductions in cycle times and corresponding improvements in customer satisfaction. But even in cases where how to make such easy improvements is not immediately obvious, a process identification exercise is often invaluable in locating potential process trouble spots, possibly important sources of process variation, and data collection needs.

The simple **flowchart** is one effective tool in process identification. Figure 1.1 is a flowchart for a printing process similar to one prepared by students (Drake, Lach, and Shadle) in a quality assurance course. The figure gives a high-level view of the work flow in a particular shop. Nearly any one of the boxes on the chart could be expanded to provide more detailed information about the printing process.

People have suggested many ways of increasing the amount of information provided by a flowchart. One possibility is the use of different shapes for the boxes on the chart, according to some kind of classification scheme for the activities being portrayed. Figure 1.1 uses only three different shapes, one each for input/output, decisions, and all else. In contrast, Kolarik's *Creating Quality: Concepts, Systems, Strategies and Tools* suggests the use of seven different symbols for flowcharting industrial processes (corresponding to operations, transportation, delays, storage, source inspection, SPC charting, and sorting inspection). Of course, many schemes are possible and potentially useful in particular circumstances.

A second way to enhance the analytical value of the flowchart is to make good use of both spatial dimensions on the chart. Typically, top-to-bottom corresponds at least roughly to time order of activities. That leaves the possibility of using left-to-right positioning to indicate some other important variable. For example, a flowchart might be segmented into several "columns" left to right, each one indicating a different physical location. Or the columns might indicate different departmental spheres of

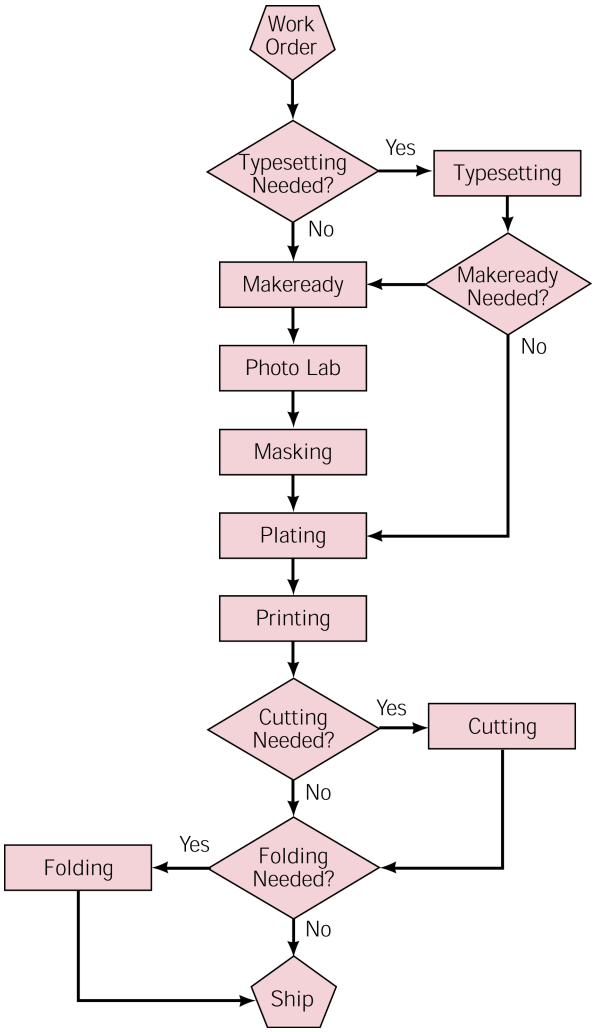


FIGURE 1.1. Flowchart of a printing process

responsibility. Such positioning is an effective way of further organizing one's thinking about a process.

Another simple device for use in process identification/mapping activities is the **Ishikawa diagram** (otherwise known as the **fishbone diagram** or **cause-and-effect diagram**). Suppose one has a desired outcome or (conversely) a quality problem in mind, and wishes to lay out the various possible contributors to the outcome or problem. It is often helpful to place these factors on a tree-like structure, where the further one moves into the tree, the more specific or basic the contributor becomes. For example, if one were interested in quality of an airline flight, general contributors might include on-time performance, baggage handling, in-flight comfort, and so on. In-flight comfort might be further amplified as involving seating, air quality, cabin service, etc. Cabin service could be broken down into components like flight attendant availability and behavior, food quality, entertainment, and so on.

Figure 1.2 is part of an Ishikawa diagram made by an industrial team analyzing an injection molding process. Without this or some similar kind of organized method of putting down the various contributors to the quality of the molded parts, nothing like an exhaustive listing of potentially important factors would be possible. The cause-and-effect diagram format provides an easily made and effective organization tool. It is an especially helpful device in group brainstorming sessions, where people are offering suggestions from many different perspectives in an unstructured way, and some kind of organization needs to be provided "on the fly."

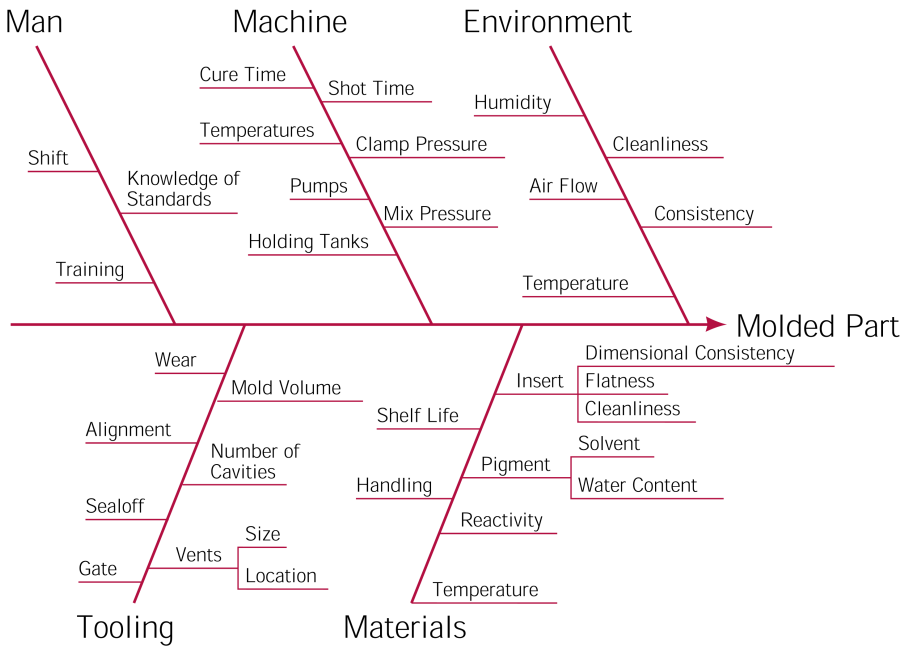


FIGURE 1.2. Cause-and-effect diagram for an injection molding process

Section 1.3 Exercises

1. The top-to-bottom direction on a flowchart usually corresponds to what important aspect of process operation?
2. How might a left-to-right dimension on a flowchart be employed to enhance process understanding?
3. What are other names for an Ishikawa diagram?
4. Name two purposes of the Ishikawa diagram.

1.4 Elementary Principles of Quality Assurance Data Collection

Good (practically useful) data do not collect themselves. Neither do they magically appear on one's desk, ready for analysis and lending insight into how to improve processes. But it sometimes seems that little is said about data collection. And in practice, people sometimes lose track of the fact that no amount of clever analysis will make up for lack of intrinsic information content in poorly collected data. Often, wisely and purposefully collected data will carry such a clear message that they essentially "analyze themselves." So we make some early comments here about general considerations in quality assurance data collection.

A first observation about the collection of quality assurance data is that if they are to be at all helpful, there must be a consistent understanding of exactly how they are to be collected. This involves having **operational definitions** for quantities to be observed and personnel who have been **well-trained** in using the definitions and any relevant measurement equipment. Consider, for example, the apparently fairly "simple" problem of measuring "the" diameters of (supposedly circular) steel rods. Simply handed a gauge and told to measure diameters, one would not really know where to begin. Should the diameter be measured at one identifiable end of the rods, in the center, or where? Should the first diameter seen for each rod be recorded, or should perhaps the rods be rolled in the gauge to get maximum diameters (for those cases where rods are not perfectly circular in cross section)?

Or consider a case where one is to collect qualitative data on defects in molded glass automobile windshields. Exactly what constitutes a "defect"? Surely a bubble one inch in diameter directly in front of the driver's head position is a defect. But would a 10^{-4}-inch diameter flaw in the same position be a problem? Or what about a one-inch diameter flaw at the very edge of the windshield that would be completely covered by trim molding? Should such a flaw be called a defect? Clearly, if useful data

are to be collected in a situation like this, very careful operational definitions need to be developed and personnel need to be taught to use them.

The importance of consistency of observation/measurement in quality assurance data collection cannot be overemphasized. When, for example, different technicians use measurement equipment in substantially different ways, what looks (in process monitoring data) like a big process change can in fact be nothing more than a change in the person doing the measurement. This is a matter we will consider from a more technical perspective Chapter 2. But here we can make the qualitative point that if operator-to-operator variation in measuring is of the same magnitude as important physical effects, and multiple technicians are going to make measurements, operator differences must be reduced through proper training and practice before there is reason to put much faith in data that are collected.

A second important point in the collection of quality assurance data has to do with **when and where** they are gathered. The closer in time and space that data are taken to an operation whose performance they are supposed to portray, the better. The ideal here is typically for well-trained workers actually doing the work or running the equipment in question to do their own data collection. There are several reasons for this. For one thing, it is such people who are in a position (after being trained in the interpretation of process monitoring data and given the authority to act on them) to react quickly and address any process ills suggested by the data that they collect. (Quick reaction to process information can prevent process difficulties from affecting additional product and producing unnecessary waste.) For another, it is simply a fact of life that data collected far away in time and space from a process rarely lead to important insights into "what is going on." Your authors have seen many student groups (against good advice) take on company projects of the variety "Here are some data we've been collecting for the past three years. Tell us what they mean." These essentially synthetic postmortem examinations never produce anything helpful for the companies involved. Even if an interesting pattern is found in such data, it is very rare that root causes can be identified completely after the fact.

If one accepts that much of the most important quality assurance data collection will be done by people whose primary job is not data collection but rather working in or on a production process, a third general point comes into focus. That is that routine data collection should be made as **convenient** as possible and where at all feasible, the methods used should make the data **immediately useful**. These days, quality assurance data are often entered as they are collected (sometimes quite automatically) into computer systems that produce real-time displays intended to show those who gathered them their most important features.

Whether automatic or pencil-and-paper data recording methods are used, thought needs to go into the making of the forms employed and displays produced. There should be no need for transfer to another form or medium before using the data. Figure 1.3 is a so-called two-variable "check sheet." Rather than making a list of (x, y) pairs and later transferring them to a piece of graph paper or a computer program for making a scatterplot, use of a pencil-and-paper form like this allows immediate display of any relationship between x and y. (Note that the use of different symbols or even colors

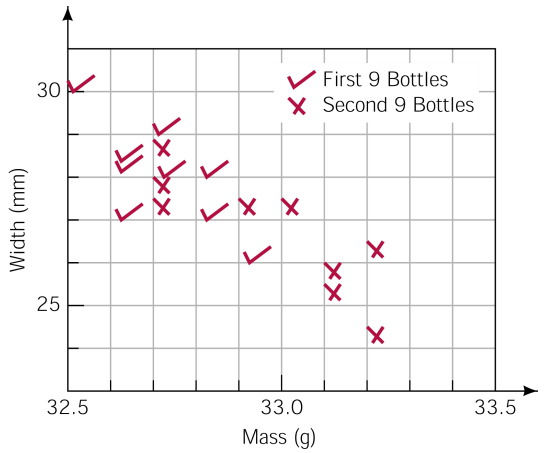


FIGURE 1.3. Check sheet for bottle mass and width of bottom piece for 18 PVC bottles

can carry information on variables besides x and y, like time order of observation.) The point here is that if one's goal is process improvement, data are for using, and their collection and immediate display needs to be designed to be practically effective.

A fourth general principle of quality assurance data collection regards adequate **documentation**. One typically collects process data hoping to locate (and subsequently eliminate) possible sources of variation. If this is to be done, care needs to be taken to keep track of conditions associated with each data point. One needs to know not only that a measured widget diameter was $1.503\,\mathrm{mm}$, but also the machine on which it was made, who was running the machine, what raw material lot was used, when it was made, what gauge was used to do the measuring, who did the measuring, and so on. Without such information there is, for example, no way to ever discover consistent differences between two machines that contribute significantly to overall variation in widget diameters. A sheet full of numbers without their histories is of little help in quality assurance.

Several additional important general points about the collection of quality assurance data have to do with the **volume** of information one is to handle. In the first place, a small or moderate amount of carefully collected (and immediately used) data will typically be worth much more than even a huge amount that is haphazardly collected (or never used). One is almost always better off trying to learn about a process based on a small data set collected with specific purposes and questions in mind than when rummaging through a large "general purpose" database assembled without the benefit of such focus.

Further, when trying to answer the question "How much data do I need to…?" one needs at least a qualitative understanding (hopefully gained in a first course in statistics) of what things govern the information content of a sample. For one thing (even in cases where one is gathering data from a particular finite lot of objects rather than from a process) it is the absolute (and not relative) size of a sample that governs

its information content. So blanket rules like "Take a 10% sample" are not rational. Rather than seeking to choose sample sizes in terms of some fraction of a universe of interest, one should think instead in terms of 1) the size of the unavoidable background variation and of 2) the size of an effect that is of practical importance. If there is no variation at all in a quantity of interest, a sample of size $n = 1$ will characterize it completely! On the other hand, if substantial variation is inevitable and small overall changes are of practical importance, huge sample sizes will be needed to illuminate important process behavior.

A final general observation is that one must take careful account of **human nature, psychology, and politics** when assigning data collection tasks. If one wants useful information, he or she had better see that those who are going to collect data are convinced that doing so will genuinely aid (and *not* threaten) them, and that accuracy is more desirable than "good numbers" or "favorable results." People who have seen data collected by themselves or colleagues used in ways that they perceive as harmful (for instance, identifying one of their colleagues as a candidate for termination) will simply not cooperate. Nor will people who see nothing coming of their honest efforts at data collection. People who are to collect data need to believe that these can help them do a better job and help their organization be successful.

Section 1.4 Exercises

1. Why is it more desirable to have data that provide a true picture of process behavior than to obtain "good numbers" or "favorable results"?

2. What personnel issues can almost surely guarantee that a data collection effort will ultimately produce nothing useful.?

3. Why is it important to have agreed upon operational definitions for characteristics of interest before beginning data collection?

4. Making real use of data collected in the past by unnamed others can be next to impossible Why?

5. How can the problem alluded to in question 4 be avoided?

6. A checksheet is a simple but informative tool. How many variables of potential interest can a form like this portray?

7. What is another virtue of a well-designed checksheet (besides that alluded to in question 6)?

8. Is a large volume of data necessarily more informative than a moderate amount? Explain.

1.5 Simple Statistical Graphics and Quality Assurance

The old saying "a picture is worth a thousand words" is especially true in the realm of statistical quality assurance. Simple graphical devices that have the potential to be applied effectively by essentially all workers have a huge potential impact. In this section, the usefulness of simple histograms, Pareto charts, scatterplots, and run charts in quality assurance efforts is discussed. This is done with the hope that readers will see the value of routinely using these simple devices as the important data organizing and communication tools that they are.

Essentially every elementary statistics book ever written has a discussion of the making of a **histogram** from a sample of measurements. Most even provide some terminology for describing various histogram shapes. That background will not be repeated here. Instead we will concentrate on the interpretation of patterns sometimes seen on histograms in quality assurance contexts, and on how they can be of use in quality improvement efforts.

Figure 1.4 is a bimodal histogram of widget diameters.

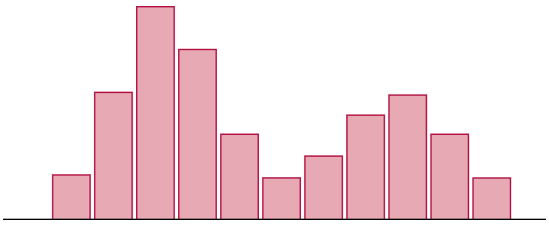


FIGURE 1.4. A bimodal histogram

Observing that the histogram has two distinct "humps" is not in and of itself particularly helpful. But asking the question "Why is the data set bimodal?" begins to be more to the point. Bimodality (or multimodality) in a quality assurance data set is a strong hint that there are two (or more) effectively different versions of something at work in a process. Bimodality might be produced by two different workers doing the same job in measurably different ways, two parallel machines that are adjusted somewhat differently, and so on. The systematic differences between such versions of the same process element produce variation that often can and should be eliminated, thereby improving quality. Viewing a plot like Figure 1.4, one can hope to identify and eliminate the physical source of the bimodality and effectively be able to "slide the two humps together" so that they coincide, thereby greatly reducing the overall variation.

The modern trend toward reducing the size of supplier bases and even "single sourcing" has its origin in the kind of phenomenon pictured in Figure 1.4. Different suppliers of a good or service will inevitably do some things slightly differently. As a result, what they supply will inevitably differ in systematic ways. Reducing a company's number of

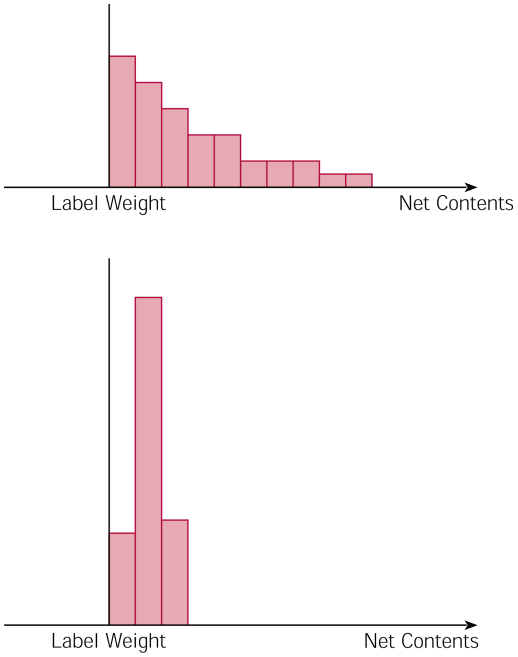


FIGURE 1.5. Two distributions of bottle contents

vendors then has two effects. Variation in the products that it makes from components or raw materials supplied by others is reduced and the costs (in terms of lost time and waste) often associated with switchovers between different material sources are also reduced.

Other shapes on histograms can also give strong clues about what is going on in a process (and help guide quality improvement efforts). For example, sorting operations often produce distinctive truncated shapes. Figure 1.5 shows two different histograms for the net contents of some containers of a liquid. The first portrays a distribution that is almost certainly generated by culling those containers (filled by an imprecise filling process) that are below label contents. The second looks as if it might be generated by a very precise filling process aimed only slightly above the labeled contents. The histograms give both hints at how the guaranteed minimum contents are achieved in the two cases, and also a pictorial representation of the waste produced by imprecision in filling. A manufacturer supplying a distribution of net contents like that in the first histogram must both deal with the rework necessitated by the part of the first distribution that has been "cut off" and also suffer the "give away cost" associated with the fact that much of the truncated distribution is quite a bit above the label value.

Figure 1.6 is a histogram for a very interesting set of data from *Engineering Statistics and Quality Control* by I.W. Burr. The very strange shape of the data set almost certainly also arose from a sorting operation. But in this case, it appears that the *center* part of the distribution is missing. In all probability, one large production run was

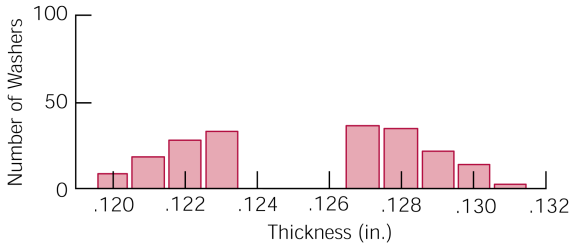


FIGURE 1.6. Thicknesses of 200 mica washers (specifications $1.25 \pm .005$ in)

made to satisfy several orders for parts of the same type. Then a sorting operation graded those parts into classes depending upon how close actual measurements were to nominal. Customers placing orders with tight specifications probably got (perhaps at a premium price) parts from the center of the original distribution, while others with looser specifications likely received shipments with distributions like the one in Figure 1.6.

Marking engineering specifications on a histogram is a very effective way of communicating to even very nonquantitative people what is needed in the way of process improvements. Figure 1.7 on page 22 shows a series of three histograms with specifications for a part dimension marked on them. In the first of those three histograms, the production process seems quite "capable" of meeting specifications for the dimension in question (in the sense of having adequate intrinsic precision), but clearly needs to be "reaimed" so that the mean measurement is lower. The second histogram portrays the output of a process that is properly aimed, but incapable of meeting specifications. The intrinsic precision is not good enough to fit the distribution between the engineering specifications. The third histogram represents data from a process that is both properly aimed *and* completely capable of meeting specifications.

Capability of a Process to Meet Specifications

Another kind of bar chart that is quite popular in quality assurance contexts is the so-called **Pareto diagram**. This tool is especially useful as a political device for getting people to prioritize their efforts and focus first on the biggest quality problems an organization faces. One makes a bar chart where problems are listed in decreasing order of frequency, dollar impact, or some other measure of importance. Often, a broken line graph indicating the cumulative importance of the various problem categories is also added to the display. Figure 1.8 on page 22 shows a Pareto diagram of assembly problems identified on a production run of 100 pneumatic hand tools. By the measure of frequency of occurrence, the most important quality problem to address is that of leaks.

The name "Pareto" is that of a mathematician who studied wealth distributions and concluded that most of the money in Italy belonged to a relatively few people. His name has become associated with the so-called "Pareto principle" or "80–20 principle." This states that "most" of anything (like quality problems or hot dog consumption) is traceable to a relatively few sources (like root causes of quality problems or hot dog eaters). Conventional wisdom in modern quality assurance is that attention to the relatively few major causes of problems will result in huge gains in efficiency and

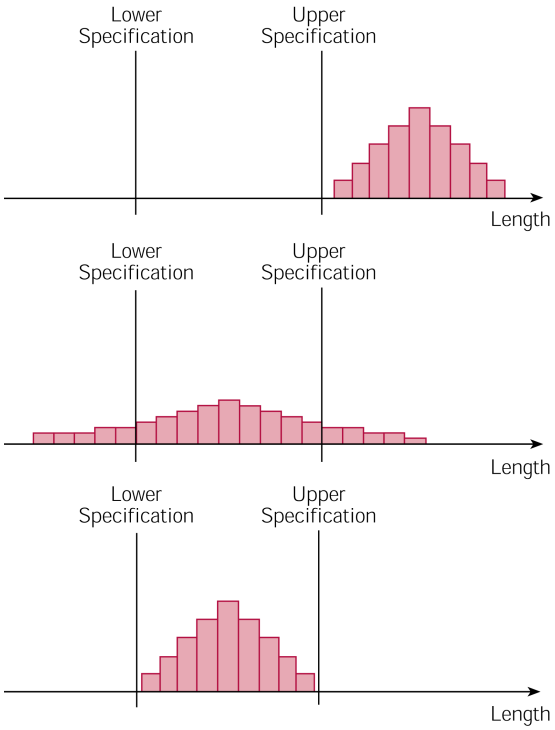


FIGURE 1.7. Three distributions of a critical machined dimension

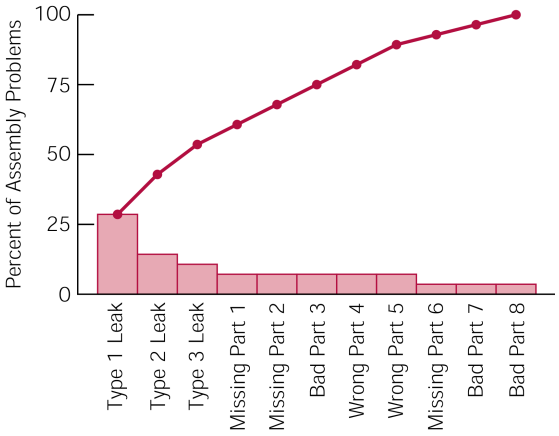


FIGURE 1.8. Pareto chart of assembly problems

quality.

Discovering *relationships* between variables is often important in discovering means of process improvement. An elementary but most important start in looking for such relationships is often the making of simple **scatterplots** (plots of (x, y) pairs). Consider Figure 1.9. This consists of two scatterplots of the numbers of occurrences of two different quality problems in lots of widgets. The stories told by the two scatterplots are quite different. In the first, there seems to be a positive correlation between the numbers of problems of the two types, while in the second no such relationship is evident. The first scatterplot suggests that a single root cause may be responsible for both types of problems and that in looking for it, one can limit attention to causes that could possibly produce both effects. The second scatterplot suggests that two different causes are at work and one will need to look for them separately.

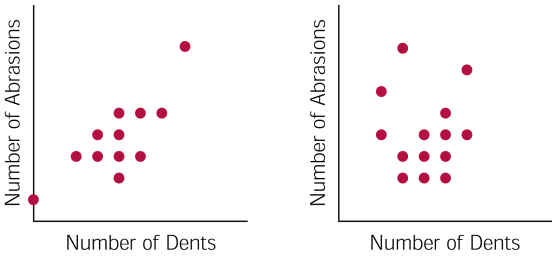


FIGURE 1.9. Two scatterplots of numbers of occurrences of manufacturing defects

It is true, of course, that one can use numerical measures (like the sample correlation) to investigate the extent to which two variables are related. But a simple scatterplot can be understood and used even by people with little quantitative background. Besides, there are things that can be seen in plots (like, for example, nonlinear relationships) that will be missed by looking only at numerical summary measures.

The habit of plotting data is one of the best habits a quality engineer can develop. And one of the most important ways of plotting is in a scatterplot against time order of observation. Where there is only a single measurement associated with each time period and one connects consecutive plotted points with line segments, it is common to call the resulting plot a **run chart**. Figure 1.10 on page 24 is a run chart of some data studied by a student project group (Williams and Markowski). Pictured are 30 consecutive outer diameters of metal parts turned on a lathe.

Investigation of the somewhat strange pattern on the plot led to a better understanding of how the turning process worked (and could have led to appropriate compensations to eliminate much of the variation in diameters seen on the plot). The first 15 diameters generally decrease with time, then there is a big jump in diameter, after which diameters again decrease. Checking production records, the students found that the lathe in question had been shut down and allowed to cool off between parts 15 and 16. The pattern seen on the plot is likely related to the dynamics of the lathe hydraulics. When cold, the hydraulics did not push the cutting tool into the workpiece

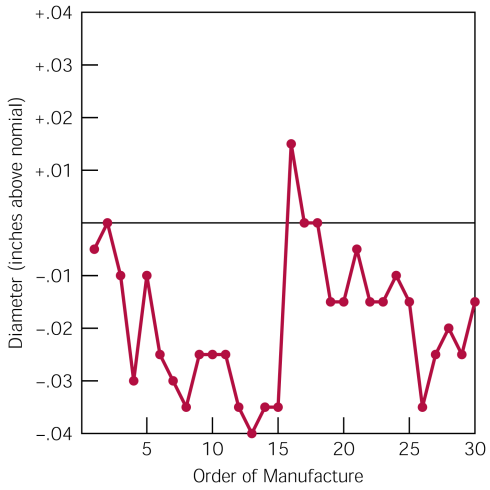


FIGURE 1.10. A run chart for 30 consecutive outer diameters turned on a lathe

as effectively as when they were warm. Hence the diameters tended to decrease as the lathe warmed up. (The data collection in question did not cover a long enough period to see the effects of tool wear, which would have tended to increase part diameters as the length of the cutting tool decreased.) If one knows that this kind of phenomenon exists, it is possible to compensate for it (and increase part uniformity) by setting artificial target diameters for parts made during a warm-up period below those for parts made after the lathe is warmed up.

Section 1.5 Exercises

1. In what ways can a simple histogram help in evaluating process performance?

2. What aspect(s) of process performance can not be pictured by a histogram?

3. The run chart is a graphical representation of process data that is not "static"; it gives more than a snapshot of process performance. What about the run chart makes it an improvement over the histogram for monitoring a process?

4. Consider Figure 1.7. The bottom histogram appears "best" with respect to being completely within specification limits and reasonably mound-shaped. Describe run charts for two different scenarios that could have produced this "best" histogram and yet reflect undesirable situations, i.e., an unstable process.

5. What is the main use of a Pareto diagram?

6. What is the rationale behind the use of a Pareto diagram?

1.6 Chapter Summary

Modern quality assurance is concerned with quality of design and quality of conformance. Statistical methods, dealing as they do with data and variation, are essential tools for producing quality of conformance. Most of the tools presented in this text are useful in the process-oriented approach to assuring quality of conformance that is outlined in Table 1.1. After providing general background on modern quality and business process improvement emphases, this chapter has introduced some of simple tools. Section 1.3 considered elementary tools for use in process mapping. Important qualitative principles of engineering and quality assurance data collection were presented in Section 1.4. And Section 1.5 demonstrated how effective simple methods of statistical graphics can be when wisely used in quality improvement efforts.

1.7 Chapter 1 Exercises

1. An engineer observes several values for a variable of interest. The average of these measurements is exactly what the engineer desires for any single response. Why should the engineer be concerned about variability in this context? How does the engineer's concern relate to product quality?

2. What is the difference between quality of conformance and quality of design?

3. Suppose 100% of all brake systems produced by an auto manufacturer have been inspected and meet safety standards. What type of quality is this? Why?

4. Describe how a production process might be characterized as exhibiting quality of conformance but potential customers are wisely purchasing a competitor's version of the product.

5. In Example 3, initial experience at an electronics manufacturing facility involved 14% yields of good IC chips.

   (a) Explain how this number (14%) was probably obtained.

   (b) Describe how the three parts of Definition 2 are involved in your answer for part (a).

6. The improved yield discussed in Example 3 came as a result of improving the chip production process. Material waste and the staff necessary to run the facility were reduced. What motivation do engineers have to improve processes if improvement might lead to their own layoff? Discuss the issues this matter raises.

7. Suppose an engineer must choose among vendors 1, 2, and 3 to supply tubing for a new product. Vendor 1 charges \$20 per tube, vendor 2 charges \$19 per tube, and vendor 3 charges \$18 per tube. Vendor 1 has implemented the Six-Step Process-Oriented Quality Assurance Cycle (and corresponding tools) in Table 1.1. As a result, only 1 tube in a million from vendor 1 is nonconforming. Vendor 2 has just begun implementation of the six steps and is producing 10% nonconforming tubes. Vendor 3 does not apply quality assurance methodology and has no idea what percent of its tubing is nonconforming. What is the price per *conforming* item for vendors 1, 2, and 3?

8. The following matrix (suggested by Dr. Brian Joiner) can be used to classify production outcomes. Good result of production means there is a large proportion of product meeting engineering specifications. (Bad result of production means there is a low proportion of product meeting requirements.) Good method of production means that quality variables are consistently near their targets. (Bad method of production means there is considerable variability about target values.)

		Result of Production	
		Good	Bad
Method of Production	Good	1	2
	Bad	3	4

   Describe product characteristics for items produced under circumstances corresponding to each of cells 1, 2, 3, and 4.

9. **Plastic Packaging.** Hsiao, Linse, and McKay investigated the production of some plastic bags, specifically hole positions on the bags. Production of these bags is done on a model 308 poly bag machine using preprinted, prefolded plastic film delivered on a roll. The plastic film is drawn through a series of rollers to punches that make holes in the bag lips. An electronic eye scans the film after it is punched and triggers heated sills which form the seals on the bags. A conveyor transports the bags to a machine operator who counts and puts them onto wickets (by placing the holes of the bags over 6-inch metal rods) and then places them in boxes. Discuss how this process and its output might variously fall into the cells 1, 2, 3, or 4 in problem 8.

10. Consider again the **Plastic Packaging** case of problem 9.

    (a) Who is the immediate customer of the hole-punching process?

    (b) Is it possible for the hole-punching process to produce hole locations with small variation and yet still produce a poor quality bag? Why or why not?

    (c) After observing that 100 out of 100 sampled bags fit over the two 6-inch wickets, an analyst might conclude that the hole-punching process needs no improvement. Is this thinking correct? Why or why not?

(d) Hsiao, Linse, and McKay used statistical methodologies consistent with steps 1, 2, 3, and 4 of the Six-Step Process-Oriented Quality Assurance Cycle and detected unacceptable variation in hole location. Would it be advisable to pursue step 6 in Table 1.1 in an attempt to improve the hole-punching process? Why or why not?

11. **Hose Skiving.** Siegler, Heches, Hoppenworth, and Wilson applied the Six-Step Process-Oriented Quality Assurance Cycle to a skiving operation. Skiving consists of taking rubber off the ends of steel-reinforced hydraulic hose so that couplings may be placed on these ends. A crimping machine tightens the couplings onto the hose. If the skived length or diameter are not as designed, the crimping process can produce an unacceptable finished hose.

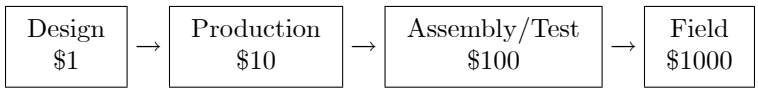
(a) What two variables did the investigators identify as directly related to product quality?

(b) Which step in the Six-Step Cycle was probably associated with identifying these two variables as important?

(c) The analysts applied steps 3 and 4 of the Six-Step Cycle and found that for a particular production line, aim and variation in skive length were satisfactory. (Unfortunately, outside diameter data were not available, so study of the outside diameter variable was not possible.) In keeping with the doctrine of continual improvement, steps 5 and 6 were considered. Was this a good idea? Why or why not?

12. Engineers at an aircraft engine manufacturer have identified several "givens" regarding cost of quality problems. Two of these are "Making it right the first time is always cheaper than doing it over" and "Fixing a problem at the source is always cheaper than fixing it later." Describe how the Six-Step Process-Oriented Quality Assurance Cycle in Table 1.1 relates to the two givens.

13. A common rule of thumb for the cost of quality problems is the "rule of 10." This rule can be summarized as follows (in terms of the dollar cost required to fix a nonconforming item):



Cost history of nonconforming parts for an aircraft engine manufacturer has been roughly as follows:

Nonconforming Item Found	Cost to Find and Fix
At production testing	$200
At final inspection	$260
At company rotor assembly	$20,000
At company assembly teardown	$60,000
In customer's airplane	$200,000
At unscheduled engine removal	$1,200,00

(a) For each step following "at production testing," calculate the ratios of "costs to find and fix" to "cost to find and fix at production testing."

(b) How do the ratios in (a) compare to the rule of 10 summarized in the four-box schematic?

(c) What does your response to (b) suggest about implementation of step 3 of the Six-Step Cycle of Table 1.1?

14. The following quotes are representative of some engineering attitudes toward quality assurance efforts. "Quality control is just a police function." "The quality control people are the ones who come in and shoot the wounded." "Our machinists will do what's easiest for them, so we'll start out with really tight engineering specifications on that part dimension."

(a) How might an engineer develop such attitudes?

(b) How can quality engineers personally avoid these attitudes and work to change them in others?

15. **Brush Ferrules.** Adams, Harrington, Heemstra, and Snyder did a quality improvement project concerned with the manufacture of some paint brushes. Bristle fibers are attached to a brush handle with a so-called "ferrule." If the ferrule is too thin, bristles fall out. If the ferrule is too thick, brush handles are damaged and can fall apart. At the beginning of the study there was some evidence that bristle fibers were falling out. "Crank position," "slider position," and "dwell time" are three production process variables that may affect ferrule thickness.

(a) What feature should analysts measure on each brush in this kind of problem?

(b) Suggest how an engineer might evaluate whether the quality problem is due to poor conformance or to poor design.

(c) From the limited information given above, what seems to have motivated the investigation?

(d) The students considered plotting the variable identified in (a) versus the time at which the corresponding brush was produced. One of the analysts

suggested first sorting the brushes according to the different crank position, slider position, and dwell time combinations, then plotting the variable chosen in (a) versus time of production *on separate graphs*. The others argued that no insight into the problem would be gained by having separate graphs for each combination. What point of view do you support? Defend your answer.

16. **Window Frames.** Christenson, Hutchinson, Mechem, and Theis worked with a manufacturing engineering department in an effort to identify cause(s) of variation and possibly reduce the amount of offset in window frame corner joints. (Excessive offset had previously been identified as the most frequently reported type of window nonconformity.)

    (a) How might the company have come to know that excessive offset in corner joints was a problem of prime importance?

    (b) What step in the Six-Step Cycle corresponds to your answer in (a)?

    (c) The team considered the following six categories of factors potentially contributing to unacceptable offset: 1) measurements, 2) materials, 3) workers, 4) environment, 5) methods, 6) machines. Suggest at least one possible cause in each of these categories.

    (d) Which step in the Six-Step Cycle of Table 1.1 is most clearly related to the kind of categorization of factors alluded to in part (c)?

17. **Machined Steel Slugs.** Harris, Murray, and Spear worked with a plant that manufactures steel slugs used to seal a hole in a certain type of casting. The group's first task was to develop and write up a standard operating procedure for data collection on several critical dimensions of these slugs. The slugs are turned on a South Bend Turrett Lathe using 1018 cold rolled steel bar stock. The entire manufacturing process is automated by means of a CNC (computer numerical control) program and only requires an operator to reload the lathe with new bar stock. The group attempted to learn about the CNC lathe program. It discovered it was possible for the operator to change the finished part dimensions by adjusting the offset on the lathe.

    (a) What benefit is there to having a standard data collection procedure in this context?

    (b) Why was it important for the group to learn about the CNC lathe program? Which step of the Six-Step Cycle is directly affected by their knowledge of the lathe program?

18. **Cut-Off Machine.** Wade, Keller, Sharp, and Takes studied factors affecting tool life for carbide cutting inserts. The group discovered that "feed rate" and "stop delay" were two factors known by production staff to affect tool life. Once a tool wears a prescribed amount, the tool life is over.

(a) What steps might the group have taken to independently verify that feed rate and stop delay impact tool life for carbide cutting inserts?

(b) What is the important response variable in this problem?

(c) How would you suggest that the variable in (b) be measured?

(d) Suggest why increased tool life might be attractive to customers using the inserts.

19. **Potentiometers.** Chamdani, Davis, and Kusumaptra worked with personnel from a potentiometer assembly plant to improve the quality of finished trimming potentiometers. The fourteen wire springs fastened to the potentiometer rotor assemblies (produced elsewhere) were causing short circuits and open circuits in the final potentiometers. Engineers suspected that the primary cause of the problems was a lack of symmetry on metal strips holding these springs. Of concern was the distance from one edge of the metal strip to the first spring and the corresponding distance from the last spring to the other end of the strip.

    (a) Suggest how the assembly plant might have discovered the short and open circuits.

    (b) Suggest how the plant producing the rotor assemblies perhaps became aware of the short and open circuits (the production plant doesn't test every rotor assembly). (Hint: Think about one of the three important emphases of modern quality philosophy. How does your response relate to the Six-Step Cycle in Table 1.1?)

    (c) If "lack of symmetry" is the cause of quality problems, what should henceforth be recorded for each metal strip inspected?

    (d) Based on your answer to (c), what measurement value corresponds to perfect symmetry?

20. "Empowerment" is a term frequently heard in today's organizations in relation to process improvement. Empowerment concerns moving decision-making authority in an organization down to the lowest appropriate levels. Unfortunately, the concept is sometimes employed only until a mistake is made, then a severe reprimand occurs and/or the decision-making privilege is moved back up to a higher level.

    (a) Name two things that are lacking in an approach to quality improvement like that described above. (Hint: Consider decision-making resulting from empowerment as a process.)

    (b) How does real, effective (and consistent) empowerment logically fit in the Six-Step Quality Improvement Cycle?

21. **Lab Carbon Blank.** The following data were provided by L. A. Currie of the National Institute of Standards and Technology (NIST). The data are preliminary

and exploratory, but real. The unit of measure is "instrument response" and is approximately equal to one microgram of carbon. (That is, 5.18 corresponds to 5.18 instrument units of carbon and about 5.18 micrograms of carbon.) The responses come from consecutive tests on "blank" material generated in the lab.

Test Number	1	2	3	4	5	6	7
Measured Carbon	5.18	1.91	6.66	1.12	2.79	3.91	2.87
Test Number	8	9	10	11	12	13	14
Measured Carbon	4.72	3.68	3.54	2.15	2.82	4.38	1.64

(a) Plot measured carbon content versus order of measurement.

(b) The data are ordered in time, but (as it turns out) time intervals between measurements were not equal (an appropriate plan for data collection was not necessarily in place). What feature of the plot in (a) might still have meaning?

(c) If one treats the measurement of lab-generated blank material as repeat measurements of a single blank, what does a trend on a plot like that in (a) suggest regarding variation of the measurement process? (Assume the plot is made from data equally spaced in time and collected by a single individual.)

(d) Make a frequency histogram of these data with categories 1.00–1.99, 2.00–2.99, etc.

(e) What could be missed if only a histogram was made (and one didn't make a plot like that in (a)) for data like these?

CHAPTER 2

Statistics and Measurement

Good measurement is fundamental to quality assurance. That which cannot be measured cannot be guaranteed to a customer. If Brinell hardness 220 is needed for certain castings and one has no means of reliably measuring hardness, there is no way to provide the castings. So most successful companies devote substantial resources to the development and maintenance of good measurement systems. In this chapter, we consider some basic concepts of measurement and discuss a variety of statistical tools aimed at quantifying and improving the effectiveness of measurement.

The chapter begins with an exposition of basic concepts and introduction to probability modeling of measurement error. Then elementary one- and two-sample statistical methods are applied to measurement problems in Section 2.2. Section 2.3 considers how slightly more complex statistical methods can be used to quantify the importance of sources of variability in measurement. Then Section 2.4 discusses studies conducted to evaluate the sizes of unavoidable measurement variation and variation in measurement chargeable to consistent differences between how operators use a measurement system. Section 2.6 considers how measurement precision effects one's ability to detect differences between process conditions. Finally, the chapter concludes with a brief section on contexts where "measurements" are go/no-go calls on individual items.

2.1 Basic Concepts in Metrology and Probability Modeling of Measurement

Metrology is the science of measurement. Measurement of many physical quantities (like lengths from inches to miles and weights from ounces to tons) is so commonplace

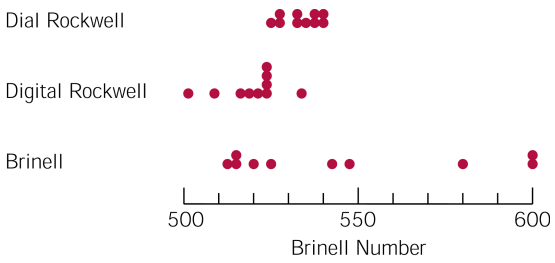


FIGURE 2.1. Brinell hardness measurements made on three different machines

that we think little about basic issues involved in metrology. But often engineers are forced by circumstances to leave the world of off-the-shelf measurement technology and devise their own instruments. And frequently because of externally imposed quality requirements for a product, one must ask "Can we even measure that?" Then the fundamental issues of **validity**, **precision**, and **accuracy** come into focus.

Definition 4 *A measurement or measuring method is said to be **valid** if it usefully or appropriately represents the feature of the measured object or phenomenon that is of interest.*

Definition 5 *A measurement system is said to be **precise** if it produces small variation in repeated measurement of the same object or phenomenon.*

Definition 6 *A measurement system is said to be **accurate** (or sometimes **unbiased**) if on average it produces the true or correct values of quantities of interest.*

Validity is the first concern when developing a measurement method. Without it, there is no point in proceeding to consider precision or accuracy. The issue is whether a method of measurement will faithfully portray the quantity of interest. When developing a new pH meter, one wants a device that will react to changes in acidity, not to changes in temperature of the solution being tested or to changes in the amount of light incident on the container holding the solution. When looking for a measure of customer satisfaction with a new model of automobile, one needs to consider those things that are important to customers. (For example, number of warranty service calls per vehicle is probably a more valid measure of customer satisfaction or aggravation with a new car than warranty dollars spent per vehicle by the manufacturer.)

Precision of measurement has to do with getting similar values every time a particular measurement is done. A bathroom scale that can produce any number between 150 lb and 160 lb when one gets on it repeatedly is really not very useful. After establishing that a measurement system produces valid measurements, consistency of those measurements is needed. Figure 2.1 portrays some hardness measurements made by a group of students (Blad, Sobotka, and Zaug) on a single metal specimen with three different hardness testers. The figure shows that the Dial Rockwell tester produced the most consistent results and would therefore be termed the most precise.

Precision is largely an intrinsic property of a measurement method or device. There is not really any way to "adjust" for poor precision or to remedy it except to 1) overhaul or replace measurement technology or to 2) average multiple measurements. In this latter regard, the reader should be familiar with the fact from elementary statistics that if $y_1, y_2, \ldots, y_n$ can be thought of as independent measurements of the same quantity, each with some mean μ and standard deviation σ, then the sample mean, $\overline{y}$, has expected or average value μ and standard deviation $\sigma/\sqrt{n}$. So people sometimes rely on multiple measurements and averaging to reduce an unacceptable precision of individual measurement (quantified by σ) to an acceptable precision of average measurement (quantified by $\sigma/\sqrt{n}$).

But even validity and precision together don't tell the whole story regarding the usefulness of real-world measurements. This can be illustrated by again considering Figure 2.1. The Dial Rockwell tester is apparently the most precise of the three testers. But it is *not* obvious from the figure what the truth is about "the" real Brinell hardness of the specimen. That is, the issue of accuracy remains. Whether any of the three testers produces essentially the "right" hardness value on average is not clear. In order to assess that, one needs to reference the testers to an accepted standard of hardness measurement.

The task of comparing a measurement method or device to a standard one and, if necessary, working out conversions that will allow the method to produce "correct" (converted) values on average is called **calibration.** In the United States, the National Institute of Standards and Technology (NIST) is responsible for maintaining and disseminating consistent standards for calibrating measurement equipment. One could imagine (if the problem were important enough) sending the students' specimen to NIST for an authoritative hardness evaluation and using the result to calibrate the testers represented in Figure 2.1. Or more likely, one might test some other specimens supplied by NIST as having known hardnesses, and use those to assess the accuracy of the testers in question (and guide any recalibration that might be needed).

An analogy that is sometimes helpful in remembering the difference between accuracy and precision of measurement is that of target shooting. Accuracy in target shooting has to do with producing a pattern centered on the bull's eye (the ideal). Precision has to do with producing a tight pattern (consistency). Figure 2.2 on page 36 illustrates four possibilities for accuracy and precision in target shooting.

Probability theory provides a helpful way to describe measurement error/variation. If a fixed quantity x called the **measurand** is to be measured with error (as all real-world quantities are) one might represent what is actually observed as

$$y = x + \epsilon \tag{2.1}$$

where ϵ is a random variable, say with mean δ and standard deviation $\sigma_{\text{measurement}}$. Model (2.1) says that the mean of what is observed is

$$\mu_y = x + \delta \ . \tag{2.2}$$

If $\delta = 0$, the measurement of x is accurate or unbiased. If δ is not 0, it is called the **measurement bias.** The standard deviation of y is (for fixed x) the standard deviation of ϵ, $\sigma_{\text{measurement}}$. So $\sigma_{\text{measurement}}$ quantifies measurement precision in model

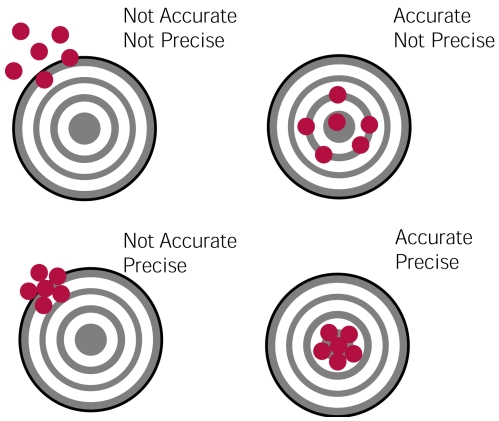


FIGURE 2.2. Measurement/target-shooting analogy

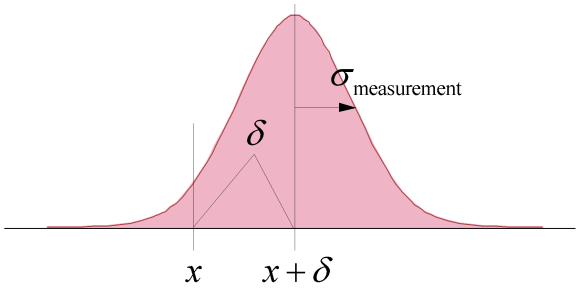


FIGURE 2.3. The distribution of a measurment y of a quantity x where measurment bias is δ and standard deviation of measurment is $\sigma_{\text{measurement}}$

(2.1). Figure 2.3 pictures the probability distribution of y and the elements x, δ, and $\sigma_{\text{measurement}}$.

Device "Linearity"

Ideally, δ is 0 (and it is the work of calibration to attempt to make it 0). At a minimum, measurement devices are designed to have a **linearity** property. This means that over the range of measurands a device will normally be used to evaluate, if its bias is not 0, it is at least constant (i.e. δ does not depend upon x). This is illustrated in Figure 2.4 (where we assume that the vertical and horizontal scales are the same).

Thinking in terms of model (2.1) is especially helpful when the measurand x itself is subject to variation. For instance, when parts produced on a machine have varying diameters x, one might think of model (2.1) as applying separately to each individual part diameter. But then in view of the reality of manufacturing variation, it makes sense to think of diameters as random, say with mean μ_x and standard deviation σ_x, independent of the measurement errors. This combination of assumptions then implies

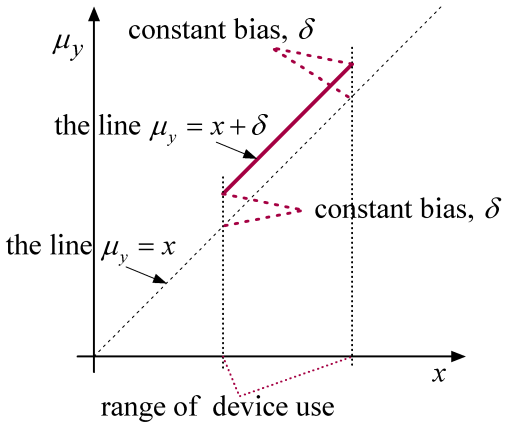


FIGURE 2.4. Measuement device "linearity" is bias constant in the measurand

(for a linear device) that the mean of what is observed is

$$\mu_y = \mu_x + \delta \tag{2.3}$$

and the standard deviation of what is observed is

$$\sigma_y = \sqrt{\sigma_x^2 + \sigma_{\text{measurement}}^2}\,. \tag{2.4}$$

Standard Deviation of Observations Subject to Measurement Error

A nonzero δ is still a measurement bias, but now observed variation across parts is seen to include one component due to variation in x and another due to measurement error. The relationships (2.3) and (2.4) between the distributions of measurement error (ϵ) and item-to-item variation in the measurand (x) and the distribution of the observed measurements (y) are pictured in Figure 2.5.

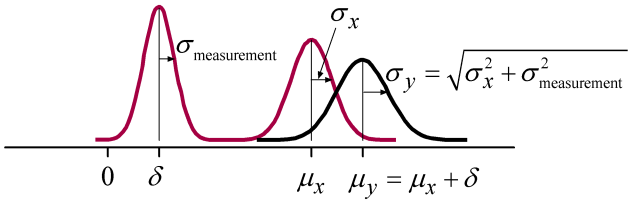


FIGURE 2.5. Random measurement error (maroon) and part variation (maroon) combine to produce observed variation (black)

Left-to-right on Figure 2.5, the two distributions in maroon represent measurement error (with bias $\delta > 0$) and measurand variation, that combine to produce variation in y represented by the distribution in black. It is the middle (maroon) distribution of x that is of fundamental interest and the figure indicates that measurement error will both

tend to shift location of that distribution and flatten it in the creation of the distribution of y. It is only this last distribution (the black one) that can be observed directly, and only when both δ and $\sigma_{\text{measurement}}$ are negligible (close to 0) are the distributions of x and y essentially the same.

Observe that equation (2.4) implies that

$$\sigma_x = \sqrt{\sigma_y^2 - \sigma_{\text{measurement}}^2}\ .$$

This suggests a way of estimating σ_x alone. If one has (single) measurements y for several parts that produce a sample standard deviation s_y, and several measurements on a single part that produce a sample standard deviation s, then a plausible estimator of σ_x is

Estimator of Process or Part Variation Excluding Measurement Error

$$\hat{\sigma}_x = \sqrt{\max\left(0, s_y^2 - s^2\right)}\ . \tag{2.5}$$

In the next sections, we will explore the use of reasoning like this, formulas like (2.5), and the application of elementary confidence interval methods to quantify various aspects of measurement precision and bias.

Section 2.1 Exercises

1. In a calibration study one compares outputs of a measurement device to "known" or "standard" values. What purpose does this serve?

2. **Pellet Densification**. Crocfer, Downey, Rixner, and Thomas studied the densification of Nd_2O_3. Pellets of this material were fired at 1400°C for various lengths of time and the resulting densities measured (in g/cc). In this context, what are the measurand (x), y, ϵ, and δ?

3. Suppose that in the context of problem 2, five pellets were fired and their densities were each measured using a single device. Further, assume the measurement device has constant bias. How many measurands (x's), y's, ϵ's, and δ's are there in this setting?

4. In the study of problem 2, the purpose was to evaluate the *effect* of firing on pellet density. Each of the pellets fired had different original densities (that were not recorded). Does the measurement protocol described in problem 2 provide data that track what is of primary interest, i.e. does it produce a valid measure of firing effect? What additional data should have been collected? Why?

5. In the context of problem 2, the density of a single pellet was repeatedly measured five times using a single device. How many measurands (x's), y's, ϵ's, and δ's are there in this setting?

6. In the context of problem 2 suppose that the standard deviation of densities from repeated measurements of the same pellet with the same device is $\sqrt{2.0}$. Suppose further that the standard deviation of actual densities one pellet to the next (the standard deviation of measurands) is $\sqrt{1.4}$. What should one then expect for a standard deviation of measured density values pellet to pellet?

7. Consider the five pellets mentioned in problem 3. Density measurements similar to the following were obtained by a single operator using a single piece of equipment with a standard protocol under fixed physical circumstances:

$$6.5, 6.6, 4.9, 5.1, \text{ and } 5.4\ .$$

   (a) What is the sample standard deviation of the $n = 5$ density measurements?

   (b) In the notation of this section, which of σ_y, σ_x or $\sigma_{\text{measurement}}$ is legitimately estimated by your sample standard deviation in (a)?

8. Again consider the five pellets of problem 3 and the five density values recorded in problem 7.

   (a) Compute the average measured density.

   (b) Assuming an appropriate model and using the notation of this section, what does your sample average estimate?

2.2 Elementary One- and Two-Sample Statistical Methods and Measurement

Elementary statistics courses provide basic inference methods for means and standard deviations based on one and two normal samples. (See, for example, Sections 6.3 and 6.4 of Vardeman and Jobe's *Basic Engineering Data Collection and Analysis*.) In this section use elementary one- and two-sample confidence interval methods to study in the simplest contexts possible 1) how measurement error influences what can be learned from data and 2) how basic properties of that measurement error can be quantified. Subsequent sections will introduce more complicated data structures and statistical methods, but the basic modeling ideas and conceptual issues can most easily be understood by first addressing them without unnecessary (and tangential) complexity.

2.2.1 One-Sample Methods and Measurement Error

"Ordinary" confidence interval formulas based on a model that says that $y_1, y_2, \ldots, y_n$ are a sample from a normal distribution with mean μ and standard deviation σ are

Confidence Limits for a Normal Mean

$$\overline{y} \pm t\frac{s}{\sqrt{n}} \text{ for estimating } \mu \tag{2.6}$$

and

Confidence Limits for a Normal Standard Deviation

$$\left(s\sqrt{\frac{n-1}{\chi^2_{\text{upper}}}}, s\sqrt{\frac{n-1}{\chi^2_{\text{lower}}}}\right) \text{ for estimating } \sigma\ . \tag{2.7}$$

These are mathematically straightforward, but little is typically said in basic courses about the practical meaning of the parameters μ and σ. So a first point to make here is that sources of physical variation (and in particular, sources of measurement error and item-to-item variation) interact with data collection plans to give practical meaning to "μ" and "σ." This in turn governs what of practical importance can be learned from application of formulas like (2.6) and (2.7).

Two Initial Applications

Figures 2.6 and 2.7 are schematic representations of two different ways that a single "sample" of n observed values y might arise. These are

1. repeat measurements on a single measurand made using the same device, and
2. single measurements made on multiple measurands coming from a stable process made using the same device.

This Book's Use of the Word "Device"

Notice that henceforth we will use the language "device" as shorthand for a fixed combination of physical measurement equipment, operator identity, measurement procedure, and surrounding physical circumstances (like time of day, temperature, etc.). We will also use the shorthand "y_i's $\sim$ ind(μ, σ)" for the model statement that observations are independent with mean μ and standard deviation σ. And in schematics like Figures 2.6 and 2.7, the rulers will represent generic measurement devices, the spheres generic measurands, and the factories generic processes.

The case represented in Figure 2.6 also corresponds to Figure 2.3 (where "measurement" variation is simply that inherent in reuse of the device to evaluate a given measurand). The case represented in Figure 2.7 also corresponds to Figure 2.5 (where again "measurement" variation is variation inherent in the "device" and now real part-to-part variation is represented by σ_x). Consider what formulas (2.6) and (2.7) provide in the two situations.

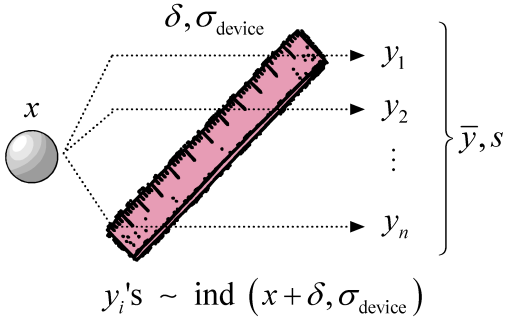


FIGURE 2.6. A single sample derived from n repeat measurements made with a fixed device on a single measurand

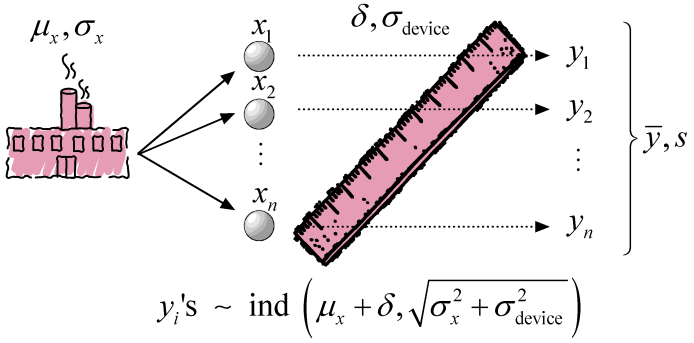


FIGURE 2.7. A single sample derived from single measurements made with a fixed device on each of n different measurands from a phsyically stable process

First, if as in Figure 2.6 n **repeat measurements of a single measurand**, $y_1, y_2, \ldots, y_n$, have sample mean $\overline{y}$ and sample standard deviation s, applying the t confidence interval for a mean, one gets an inference for

$$x + \delta = \text{measurand plus bias} .$$

So

1. in the event that the measurement device is known to be well-calibrated (one is sure that $\delta = 0$, there is no systematic error), the limits $\overline{y} \pm ts/\sqrt{n}$ based on $\nu = n - 1$ df are limits for x, and
2. in the event that what is being measured is a *standard* **for which** x **is known**, one may use the limits
$$(\overline{y} - x) \pm t\frac{s}{\sqrt{n}}$$
(once again based on $\nu = n - 1$ df) to estimate the device bias, δ.

Further, applying the χ^2 confidence interval for a standard deviation, one has an inference for the size of the device "noise," σ_{device}.

Next consider what can be inferred from **single measurements made on n different measurands** $y_1, y_2, \ldots, y_n$ from a stable process with sample mean $\overline{y}$ and sample standard deviation s as illustrated in Figure 2.7. Here

1. the limits $\overline{y} \pm ts/\sqrt{n}$ (for t based on $\nu = n - 1$ df) are limits for

$$\mu_x + \delta = \text{the mean of the distribution of true values plus bias ,}$$

and

2. the quantity s estimates $\sigma_y = \sqrt{\sigma_x^2 + \sigma_{\text{device}}^2}$, that we met first in display (2.4) and have noted really isn't of fundamental interest. So there is little point in direct application of the χ^2 confidence limits (2.7) in this context.

Example 7 ***Measuring Concentration.*** *Below are $n = 5$ consecutive concentration measurements made by a single analyst on a single physical specimen of material using a particular assay machine (the real units are not available, so for sake of example, let's call them "moles per liter," mol/l).*

$$1.0025, .9820, 1.0105, 1.0110, .9960$$

These have mean $\overline{y} = 1.0004\,\text{mol}/\text{l}$ and $s = .0120\,\text{mol}/\text{l}$. Consulting a χ^2 table like Table A.3 using $\nu = 5 - 1 = 4$ df, we can find 95% confidence limits for σ_{device} (the size of basic measurement variability) as

$$.0120\sqrt{\frac{4}{11.143}} \text{ and } .0120\sqrt{\frac{4}{.484}} \text{ i.e. } .0072\,\text{mol}/\text{l} \text{ and } .0345\,\text{mol}/\text{l}\,.$$

(One moral here is that ordinary small sample sizes give very wide confidence limits for a standard deviation.) Consulting a t table like Table A.4 also using 4 df, we can find 95% confidence limits for the measurand plus instrument bias ($x + \delta$) to be

$$1.0004 \pm 2.776\frac{.0120}{\sqrt{4}} \text{ i.e. } 1.0004\,\text{mol}/\text{l} \pm .0167\,\text{mol}/\text{l}\,.$$

Note that if the measurements in question were done on a standard material "known" to have actual concentration $1.0000\,\text{mol}/\text{l}$, these limits then correspond to limits for device bias of

$$0.0004\,\text{mol}/\text{l} \pm .0167\,\text{mol}/\text{l}\,.$$

Finally, suppose that subsequently samples from $n = 20$ different batches are analyzed and $\overline{y} = .9954$ and $s_y = .0300$. The 95% t confidence limits

$$.9954 \pm 2.093\frac{.0300}{\sqrt{20}} \text{ i.e. } .9954 \pm .0140$$

are for $\mu_x + \delta$, the process mean plus any device bias/systematic error.

Application to a Sample Consisting of Single Measurements of a Single Measurand Made Using Multiple Devices (From a Large Population of Such Devices)

The two cases illustrated by Figures 2.6 and 2.7 do not begin to exhaust the ways that the basic formulas (2.6) and (2.7) can be applied. We present two more applications of the one-sample formulas, beginning with an application where single measurements of a single measurand are made using multiple devices (from a large population of such devices).

There are contexts in which an organization has many "similar" measurement devices that could potentially be used to do measuring. In particular, a given piece of equipment might well be used by any of a large number of operators. Recall that we are using the word "device" to describe a particular combination of equipment, people, procedures, etc. used to produce a measurement. So, in this language, different operators with a fixed piece of equipment are different "devices." A way to compare these devices would be to use some (say n of them) to measure a single measurand. This is illustrated in Figure 2.8.

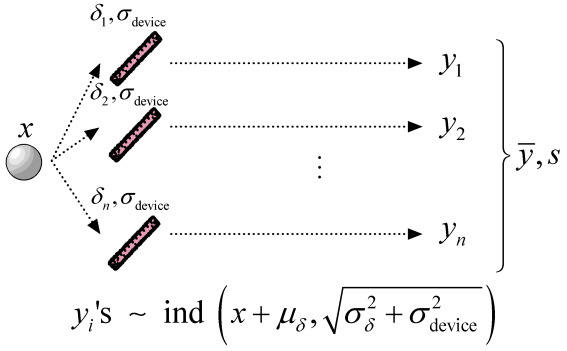


FIGURE 2.8. A single sample consisting of n single measurements of a fixed measurand made with each of n devices (from a large population of such devices with a common precision)

In this context, a measurement is of the form

$$y = x + \epsilon \, ,$$

where $\epsilon = \delta + \epsilon^*$, for δ the (randomly selected) bias of the device used and ϵ^* a measurement error with mean 0 and standard deviation σ_{device} (representing a repeat measurement variability for any one device). So one might write

$$y = x + \delta + \epsilon^* \, .$$

Thinking of x as fixed and δ and ϵ^* as independent random variables (δ with mean μ_δ, the average device bias, and standard deviation σ_δ measuring variability in device biases) the laws of mean and variance from elementary probability then imply that

$$\mu_y = x + \mu_\delta + 0 = x + \mu_\delta \tag{2.8}$$

and

$$\sigma_y = \sqrt{0 + \sigma_\delta^2 + \sigma_{\text{device}}^2} = \sqrt{\sigma_\delta^2 + \sigma_{\text{device}}^2} \tag{2.9}$$

as indicated on Figure 2.8. The theoretical average measurement is the measurand plus the average bias and the variability in measurements comes from both variation in device biases and the intrinsic imprecision of any particular device.

In a context where a schematic like Figure 2.8 represents a study where several operators each make a measurement on the same item using a fixed piece of equipment, the quantity

$$\sqrt{\sigma_\delta^2 + \sigma_{\text{device}}^2}$$

Repeatability and Reproducibility

is a kind of overall measurement variation that is sometimes called "$\sigma_{\text{R\&R}}$," the first "R" standing for **repeatability** and referring to σ_{device} (a variability for fixed operator on the single item) and the second "R" standing for **reproducibility** and referring to σ_δ (a between-operator variability).

With μ_y and σ_y identified in displays (2.8) and (2.9), it is clear what the one sample confidence limits (2.6) and (2.7) estimate in this context. Of the two, interval (2.7) for "σ" is probably most important, since σ_y is interpretable in the context of an R&R study, while μ_y typically has little practical meaning. It is another question (that we will address in future sections with more complicated methods) how one might go about separating the two components of σ_y to assess the relative sizes of repeatability and reproducibility variation.

Application to a Sample Consisting of Differences in Measurements on Multiple Measurands Made Using Two Devices (Assuming Device Linearity)

Another way to create a single sample of numbers is this. With two devices available and n different measurands, one might measure each once with both devices and create n differences between device 1 and device 2 measurements. This is a way of potentially comparing the two devices and is illustrated in Figure 2.9.

In this context, a difference is of the form

$$d = y_1 - y_2 = (x + \epsilon_1) - (x + \epsilon_2) = \epsilon_1 - \epsilon_2$$

and (again applying the laws of mean and variance from elementary probability) it follows that

$$\mu_d = \delta_1 - \delta_2 \text{ and } \sigma_d = \sqrt{\sigma_{\text{device1}}^2 + \sigma_{\text{device2}}^2}$$

as indicated on Figure 2.9. So applying the t interval for a mean (2.6), the limits

$$\bar{d} \pm t\frac{s}{\sqrt{n}}$$

provide a way to estimate $\delta_1 - \delta_2$, the difference in device biases.

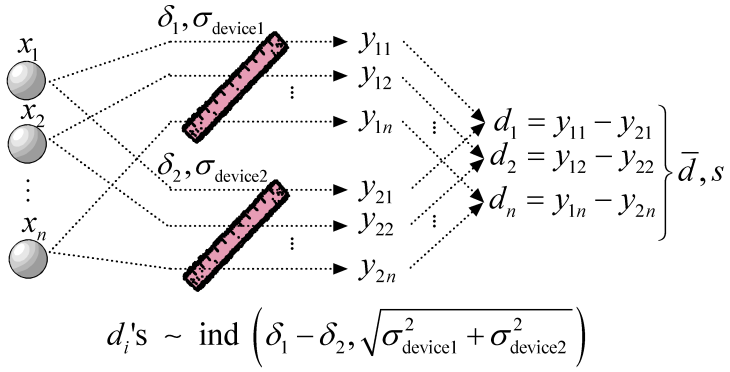


FIGURE 2.9. A single sample consisting of n differences of single measurements of n measurands made using 2 devices (assuming device linearity)

2.2.2 Two-Sample Methods and Measurement Error

Parallel to the one-sample formulas are the two-sample formulas of elementary statistics. These are based on a model that says that

$$y_{11}, y_{12}, \ldots, y_{1n_1} \text{ and } y_{21}, y_{22}, \ldots, y_{2n_2}$$

are independent samples from normal distributions with respective means μ_1 and μ_2 and respective standard deviations σ_1 and σ_2. In this context, the so-called "Satterthwaite approximation" gives limits

$$\bar{y}_1 - \bar{y}_2 \pm \hat{t}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ for estimating } \mu_1 - \mu_2\,, \tag{2.10}$$

Confidence Limits for a Differrence in Normal Means

where appropriate "approximate degrees of freedom" for $\hat{t}$ are

$$\hat{\nu} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}}\,. \tag{2.11}$$

Satterthwaite Degrees of Freedom for Formula (2.10)

(This method is one that you may not have seen in an elementary statistics course, where often only methods valid when one assumes that $\sigma_1 = \sigma_2$ are presented. We use

this method not only because it requires less in terms of model assumptions than the more common formula, but also because we will have other uses for the Satterthwaite idea in this chapter, so it might as well be met first in this simple context.) It turns out that $\min((n_1-1),(n_2-1)) \leq \hat{\nu}$, so that a simple conservative version of this method uses degrees of freedom

Conservative Simplification of Formula (2.11)

$$\hat{\nu}^* = \min((n_1-1),(n_2-1)) \ . \tag{2.12}$$

Further, in the two-sample context, there are elementary confidence limits

Confidence Limits for a Ratio of Two Normal Standard Deviations

$$\frac{s_1}{s_2} \cdot \frac{1}{\sqrt{F_{(n_1-1),(n_2-1),\text{upper}}}} \quad \text{and} \quad \frac{s_1}{s_2} \cdot \frac{1}{\sqrt{F_{(n_1-1),(n_2-1),\text{lower}}}} \quad \text{for } \frac{\sigma_1}{\sigma_2} \tag{2.13}$$

(and be reminded that $F_{(n_1-1),(n_2-1),\text{lower}} = 1/F_{(n_2-1),(n_1-1),\text{upper}}$ so that standard F tables like Table A.5 giving only upper percentage points can be employed).

Application to Two Samples Consisting of Repeat Measurements of a Single Measurand Made Using Two Different Devices

One way to create "two samples" of measurements is to measure the same item repeatedly with two different devices. This possibility is illustrated in Figure 2.10.

Direct application of the two-sample confidence interval formulas here shows that the two-sample Satterthwaite approximate t interval (2.10) provides limits for

$$\mu_1 - \mu_2 = (x+\delta_1) - (x+\delta_2) = \delta_1 - \delta_2$$

(the difference in device biases), while the F interval (2.13) provides a way of comparing device standard deviations σ_{device1} and σ_{device2} through direct estimation of

$$\frac{\sigma_{\text{device1}}}{\sigma_{\text{device2}}} \ .$$

This data collection plan thus provides for straightforward comparison of the basic characteristics of the two devices.

Example 8 ***Measuring Styrofoam "Packing Peanut" Size.*** *In an in-class measurement exercise, two students used the same caliper to measure the "size" of a single Styrofoam "packing peanut" according to a class-standard measurement protocol. Some*

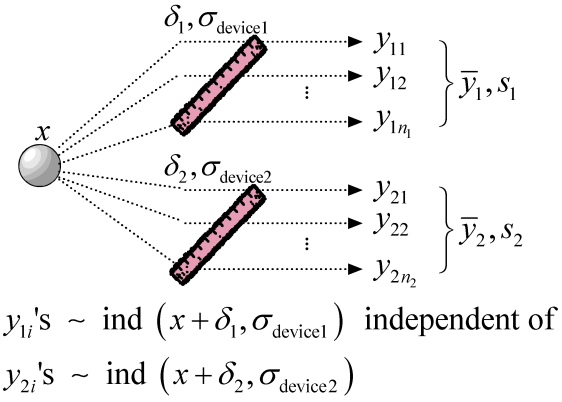


FIGURE 2.10. Two samples consisting of n_1 and n_2 measurements of a single measurand with two devices

summary statistics from their work follow.

Student 1	*Student 2*
$n_1 = 4$	$n_2 = 6$
$\overline{y}_1 = 1.42\,\text{cm}$	$\overline{y}_2 = 1.44\,\text{cm}$
$s_1 = .20\,\text{cm}$	$s_2 = .40\,\text{cm}$

In this context, the difference in the two measurement "devices" is the difference in "operators" making the measurements. Consider quantifying how this difference affects measurement.

To begin, note that from formula (2.11)

$$\hat{\nu} = \frac{\left(\dfrac{(.20)^2}{4} + \dfrac{(.40)^2}{6}\right)^2}{\dfrac{(.20)^4}{(4-1)(4)^2} + \dfrac{(.40)^4}{(6-1)(6)^2}} \approx 7.7$$

or using the more conservative display (2.12) one gets

$$\hat{\nu}^* = \min\left((4-1),(6-1)\right) = 3$$

So (rounding the first of these down to 7*) one should use either* 7 *or* 3 *degrees of freedom with formula (2.10). For sake of example, using* $\hat{\nu}^* = 3$ *degrees of freedom, consulting Table A.4, the upper 2.5% point of the* t *distribution with* 3 *df is* 3.182*. So 95% confidence limits for the difference in biases for the two operators using this caliper are*

$$1.42 - 1.44 \pm 3.182\sqrt{\frac{(.20)^2}{4} + \frac{(.40)^2}{6}}$$

i.e.

$$-.02\,\text{cm} \pm .61\,\text{cm}$$

The apparent difference in biases is small in comparison to the imprecision associated with that difference.

Then, since from Table A.5 the upper 2.5% point of the $F_{3,5}$ *distribution is* 7.764 *and the upper 2.5% point of the* $F_{5,3}$ *distribution is* 14.885*, 95% confidence limits for the ratio of standard deviations of measurement for the two operators are*

$$\frac{.20}{.40} \cdot \frac{1}{\sqrt{7.764}} \quad \textit{and} \quad \frac{.20}{.40} \cdot \sqrt{14.885}$$

i.e.

$$.19 \quad \textit{and} \quad 1.93$$

Since this interval covers values both smaller and larger than 1.00*, there is in the limited information available here no clear indicator of which of these students is the most consistent in his or her use of the caliper in this measuring task.*

Application to Two Samples Consisting of Single Measurements Made With Two Devices On Multiple Measurands From a Stable Process (Only One Device Being Used for a Given Measurand)

Comparing Devices When Measurement is Destructive

There are quality assurance contexts in which measurement is **destructive** (and cannot be repeated for a single measurand) and nevertheless one needs to somehow compare two different devices. In such situations, the only thing that can be done is to take items from some large pool of items or from some stable process and (probably after randomly assigning them one at a time to one or the other of the devices) measure them and try to make comparisons based on the resulting samples. This possibility is illustrated in Figure 2.11. This is a schematic for two samples consisting of single measurements made with two devices on multiple measurands from a stable process (only one device used for a given measurand).

Direct application of the two-sample Satterthwaite approximate t interval (2.10) provides limits for

$$\mu_1 - \mu_2 = (\mu_x + \delta_1) - (\mu_x + \delta_2) = \delta_1 - \delta_2$$

(the difference in device biases). So, in even in contexts where measurement is destructive, it is possible to compare device biases. It's worth contemplating, however, the difference between the present scenario and the immediately preceding one (represented by Figure 2.10).

The measurements y in Figure 2.10 are less variable than are the measurements y here in Figure 2.11. This is evident in the standard deviations shown on the figures and follows from the fact that in the present case (unlike the previous one) measurements are affected by unit-to-unit/measurand-to-measurand variation. So all else being equal,

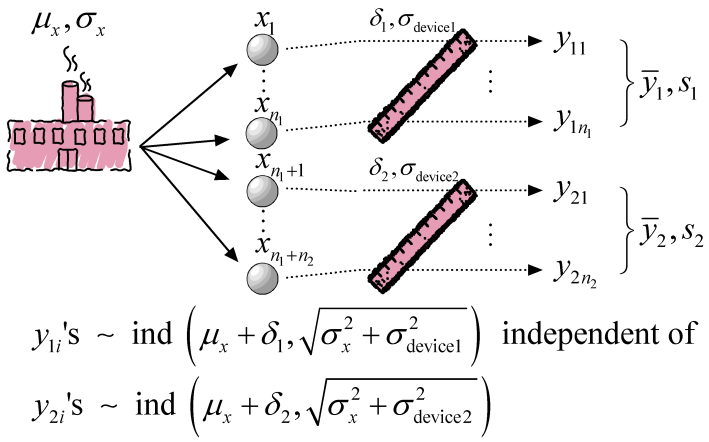


FIGURE 2.11. Two samples consisting of single measurments made on $n_1 + n_2$ measurands from a stable process, n_1 with device 1 and n_2 with device 2

one should expect limits (2.10) applied in the present context to be wider/less informative than when applied to data collected as in the last application. That should be in accord with intuition. One should expect to be able to learn more useful to comparing devices when the same item(s) can be remeasured than when it (they) can not be remeasured.

Notice that if the F limits (2.13) are applied here, one winds up with only an indirect comparison of σ_{device1} and σ_{device2}, since all that can be easily estimated (using the limits (2.13)) is the ratio

$$\frac{\sqrt{\sigma_x^2 + \sigma_{\text{device1}}^2}}{\sqrt{\sigma_x^2 + \sigma_{\text{device2}}^2}}$$

and NOT the (more interesting) ratio $\sigma_{\text{device1}}/\sigma_{\text{device2}}$.

Application to Two Samples Consisting of Repeat Measurements Made With One Device On Two Measurands

A basic activity of quality assurance is the comparison of nominally identical items. Accordingly, another way to create two samples is to make repeated measurements on two measurands with a single device. This is illustrated in Figure 2.12 on page 50.

In this context,

$$\mu_1 - \mu_2 = (x_1 + \delta) - (x_2 + \delta) = x_1 - x_2$$

so that application of the two-sample Satterthwaite approximate t interval (2.10) provides limits for the *difference in the measurands*, and a direct way of comparing the measurands. The device bias affects both samples in the same way and "washes out" when one takes a difference. (This, of course, assumes that the device is linear, i.e. that the bias is constant.)

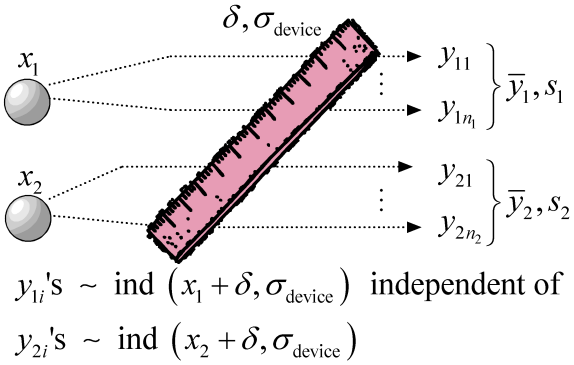


FIGURE 2.12. Two samples consisting of repeat measurments made with one device on two measurands

Application to Two Samples Consisting of Single Measurements Made Using a Single Device on Multiple Measurands Produced by Two Stable Processes

Another basic activity of quality assurance is the comparison of nominally identical *processes*. Accordingly, another way to create two samples is to make single measurements on samples of measurands produced by two processes. This possibility is illustrated in Figure 2.13.

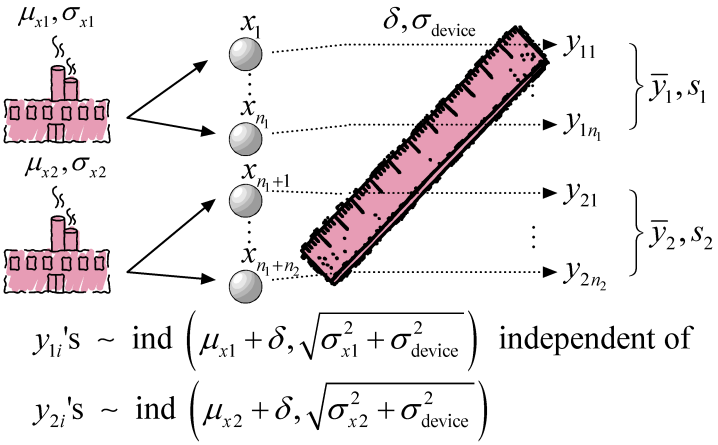


FIGURE 2.13. Two samples consisting of single measurements made using a single device on multiple measurands produced by two stable processes

In this context,

$$\mu_1 - \mu_2 = (\mu_{x1} + \delta) - (\mu_{x2} + \delta) = \mu_{x1} - \mu_{x2}$$

so that application of the two-sample Satterthwaite approximate t interval (2.10) provides limits for the *difference in the process mean measurands* and hence a direct way of comparing the processes. Again, the device bias affects both samples in the same way and "washes out" when one takes a difference (still assuming that the device is linear, i.e. that the bias is constant).

If the F limits (2.13) are applied here, one winds up with only an indirect comparison of σ_{x1} and σ_{x2}, since what can be easily estimated is the ratio

$$\frac{\sqrt{\sigma_{x1}^2 + \sigma_{\text{device}}^2}}{\sqrt{\sigma_{x2}^2 + \sigma_{\text{device}}^2}}$$

and not the practically more interesting σ_{x1}/σ_{x2}.

Section 2.2 Exercises

1. Consider again the **Pellet Densification** case of problem 7 in Section 2.1. Suppose the five data values $6.5, 6.6, 4.9, 5.1,$ and 5.4 were measured densities for a single pellet produced by five different operators using the same piece of measuring equipment (or by the same operator using five different pieces of equipment—the two scenarios are conceptually handled in the same way). Use the notation of this section ($x, \delta, \mu_\delta, \sigma_\delta,$ and σ_{device}) below.

   (a) What does the sample average of these five data values estimate?

   (b) What does the sample standard deviation of these five data values estimate?

   (c) Which of the two estimates in (a) and (b) is probably more important? Why?

2. Return again to the context of problem 7 of Section 2.1. Suppose the original set of five data values $6.5, 6.6, 4.9, 5.1,$ and 5.4 was obtained from five different pellets by operator 1 using piece of equipment 1. Using a second piece of equipment, operator 1 recorded densities $6.6, 5.7, 5.9, 6.2,$ and 6.3 for the same five pellets. So, for pellet 1, "device 1" produced measurement 6.5 and "device 2" produced 6.6; for pellet 2, "device 1" produced measurement 6.6 and "device 2" produced 5.7, and so on.

   (a) Give the five differences in measured densities (device 1 minus device 2). Calculate the sample average difference. What does this estimate? (Hint: Consider δ's .)

   (b) Calculate the sample standard deviation of the five differences (device 1 minus device 2). What does this estimate? (Hint: Consider the σ_{device}'s.)

   (c) Find 90% confidence limits for the average difference in measurements from the two devices.

3. Suppose the two sets of five measurements referred to in problems 1 and 2 actually came from one pellet, i.e., operator 1 measured the same pellet five times with piece of equipment 1 and then measured that same pellet five times with piece of equipment 2.

   (a) Find a 95% confidence interval for the ratio of the two device standard deviations ($\sigma_{\text{device1}}/\sigma_{\text{device2}}$). What do your limits indicate about the consistency of measurements from device 1 compared to that of measurements from device 2?

   (b) Find a 95% two-sample Satterthwaite approximate t interval for the difference in the two device averages (device 1 minus device 2). If your interval were to include 0, what would you conclude regarding device biases 1 and 2?

4. Consider now the same ten data values referred to in problems 2 and 3, but a different data collection plan. Suppose the first five data values were measurements on five different pellets by operator 1 using piece of equipment 1 and the second set of data values was for another set of pellets by operator 1 using piece of equipment 2. Assume both sets of pellets came from the same physically stable process.

   (a) What does the sample standard deviation from the first set of five data values estimate?

   (b) What does the sample standard deviation from the second set of five data values estimate?

   (c) What does the difference in the two sample average densities estimate?

5. Reflect on problems 3 and 4. Which data-taking approach is better for estimating the difference in device biases? Why?

6. In the same **Pellet Densification** context considered in problems 1 through 5, suppose one pellet was measured five times by operator 1 and a different pellet was measured five times by operator 1 (the same physical equipment was used for the entire set of 10 observations). What is estimated by the difference in the two sample averages?

7. Once again in the context of problems 1 through 6, suppose the first five data values were measurements on five different pellets made by operator 1 using piece of equipment 1 and the second five were measurements of a different set of pellets by operator 1 using piece of equipment 1. Assume the two sets of pellets come from different firing methods (method 1 and method 2). Assume the two firing processes are physically stable.

   (a) Find the two-sample Satterthwaite approximate t interval (method 1 minus method 2) for the difference in the process mean measurands.

(b) In words, what does the interval in (a) estimate? In symbols, what does the interval in (a) estimate?

(c) With this approach to data taking, can either device bias be estimated directly? Why or why not?

8. Still in the context of problems 1 through 7, density measurements similar to the values $6.5, 6.6, 4.9, 5.1,$ and 5.4 were obtained for five different pellets by a single operator using a single piece of measuring equipment under a standard protocol and fixed physical circumstances. Use the t confidence interval for a mean and give 95% confidence limits for the mean of the distribution of true densities plus measurement bias.

9. Suppose the five measurements in problem 8 are repeat measurements from only one pellet, not from five different pellets.

(a) Use the χ^2 confidence limits for a standard deviation (from elementary statistics) and give a 95% confidence interval for $\sigma_{\text{measurement}}$.

(b) Use the t confidence interval formula for a mean from elementary statistics and give 95% confidence limits for the (single) true pellet density plus measurement bias.

2.3 Some Intermediate Statistical Methods and Measurement

Through reference to familiar elementary one- and two-sample methods of statistical inference, Section 2.2 illustrated the basic insight that:

> How sources of physical variation interact with a data collection plan governs what of practical importance can be learned from a data set, and in particular, how measurement error is reflected in the data set.

In this section we consider some computationally more complicated statistical methods and what they provide in terms of quantification of the impact of measurement variation on quality assurance data.

2.3.1 A Simple Method for Separating Process and Measurement Variation

In Section 2.1 we essentially observed that

1. repeated measurement of a single measurand with a single device allows one to estimate device variability, and

2. single measurements made on multiple measurands from a stable process allow one to estimate a combination of process and measurement variability,

and remarked that these facts suggest formula (2.5) as a way to estimate a process standard deviation alone. Our first objective in this section is to elaborate a bit on this thinking.

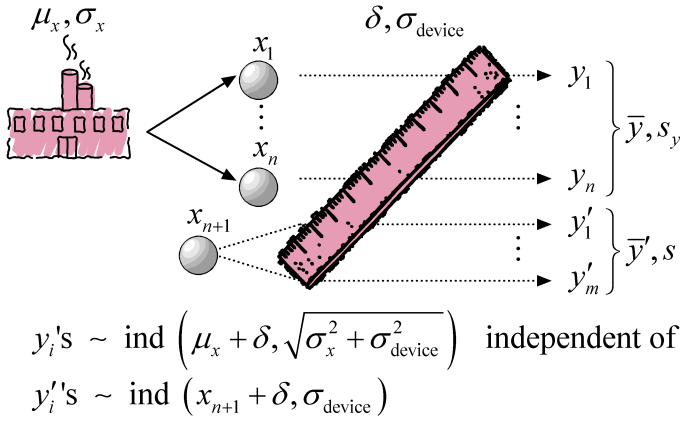


FIGURE 2.14. Schematic of a data collection plan that allows evaluation of σ_x without inflation by measurement variation

Figure 2.14 is a schematic of a data collection plan that combines elements 1 and 2 above. Here we use the notation y for the single measurements on n items from the process and the notation y' for the m repeat measurements on a single measurand. The sample standard deviation of the y's, s_y, is a natural empirical approximation for $\sigma_y = \sqrt{\sigma_x^2 + \sigma_{\text{device}}^2}$ and the sample standard deviation of the y''s, s, is a natural empirical approximation for σ_{device}. That suggests that one estimate the process standard deviation with

Estimator of Process Standard Deviation not Inflated by Measurement Variability

$$\hat{\sigma}_x = \sqrt{\max\left(0, s_y^2 - s^2\right)} \tag{2.14}$$

as indicated in display (2.5). (The maximum of 0 and $s_y^2 - s^2$ under the root is there simply to ensure that one is not trying to take the square root of a negative number in the rare case that s exceeds s_y.) $\hat{\sigma}_x$ is not only a sensible single number estimate of σ_x, but can also be used to make approximate confidence limits for the process standard deviation. The so-called Satterthwaite approximation suggests that one use

$$\hat{\sigma}_x\sqrt{\frac{\hat{\nu}}{\chi^2_{\text{upper}}}} \quad \text{and} \quad \hat{\sigma}_x\sqrt{\frac{\hat{\nu}}{\chi^2_{\text{lower}}}} \tag{2.15}$$

Satterthwaite Approximate Confidence Limits for a Process Standard Deviation

as limits for σ_x, where appropriate approximate degrees of freedom $\hat{\nu}$ to be used finding χ^2 percentage points are

$$\hat{\nu} = \frac{\hat{\sigma}_x^4}{\dfrac{s_y^4}{n-1} + \dfrac{s^4}{m-1}} \tag{2.16}$$

Satterthwaite Approximate df for Use With Limits (2.15)

Example 9 ***(Example 7 Revisited.)*** *In Example 7, we considered $m = 5$ measurements made by a single analyst on a single physical specimen of material using a particular assay machine that produced $s = .0120\,\text{mol/l}$. Subsequently, specimens from $n = 20$ different batches were analyzed and $s_y = .0300\,\text{mol/l}$. Using formula (2.14), an estimate of real process standard deviation uninflated by measurement variation is*

$$\hat{\sigma}_x = \sqrt{\max\left(0, (.0300)^2 - (.0120)^2\right)} = .0275\,\text{mol/l}$$

and this value can used to make confidence limits. By formula (2.16) approximate degrees of freedom are

$$\hat{\nu} = \frac{(.0275)^4}{\frac{(.0300)^4}{19} + \frac{(.0120)^4}{4}} = 11.96\,.$$

So rounding down to $\hat{\nu} = 11$, since the upper 2.5% point of the χ^2_{11} distribution is 21.920 and the lower 2.5% point is 3.816, by formula (2.15) approximate 95% confidence limits for the real process standard deviation (σ_x) are

$$.0275\sqrt{\frac{11}{21.920}} \text{ and } .0275\sqrt{\frac{11}{3.816}}\,,$$

i.e.

$$.0195\,\text{mol/l} \text{ and } .0467\,\text{mol/l}\,.$$

2.3.2 One-Way Random Effects Models and Associated Inference

One of the basic models of intermediate statistical methods is the so-called "one-way random effects model" for I samples of observations

$$\begin{array}{c} y_{11}, y_{12}, \ldots, y_{1n_1} \\ y_{21}, y_{22}, \ldots, y_{2n_2} \\ \vdots \\ y_{I1}, y_{I2}, \ldots, y_{In_I} \end{array}$$

One-Way Random Effects Model

This model says that the observations may be thought of as

$$y_{ij} = \mu_i + \epsilon_{ij}$$

where the ϵ_{ij} are independent normal random variables with mean 0 and standard deviation σ, while the I values μ_i are independent normal random variables with mean μ and standard deviation σ_μ (independent of the ϵ's). (One can think of I means μ_i drawn at random from a normal distribution of μ_i's, and subsequently observations y generated from I different normal populations with those means and a common standard deviation.) In this model, the three parameters are σ (the "within group" standard deviation), σ_μ (the "between group" standard deviation), and μ (the overall mean). The squares of the standard deviations are called "variance components" since for any particular observation, the laws of expectation and variance imply that

$$\mu_y = \mu + 0 = \mu \text{ and } \sigma_y^2 = \sigma_\mu^2 + \sigma^2$$

(i.e. σ_μ^2 and σ^2 are components of the variance of y).

Two quality assurance contexts where this model can be helpful are where

1. multiple measurands from a stable process are each measured multiple times using the same device, and
2. a single measurand is measured multiple times using multiple devices.

These two scenarios and the accompanying parameter values are illustrated in Figures 2.15 and 2.16.

There are well-established (but not altogether simple) methods of inference associated with the one-way random effects model, that can be applied to make confidence intervals for the model parameters (and inferences of practical interest in metrological applications). Some of these are based on so-called ANOVA methods and the one-way ANOVA identity that says that with

$$\overline{y}_i = \frac{1}{n_i}\sum_j y_{ij}, n = \sum_i n_i, \text{ and } \overline{y} = \frac{1}{n}\sum_i n_i \overline{y}_i,$$

it is the case that

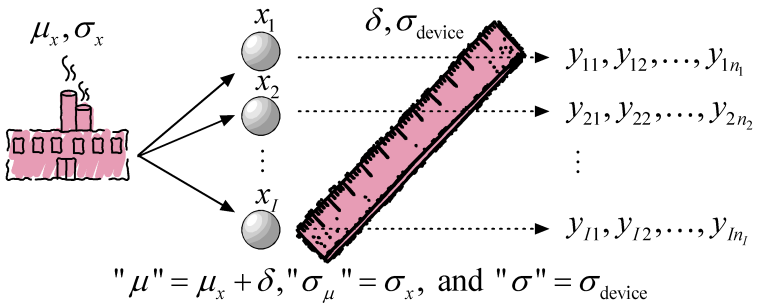


FIGURE 2.15. Multiple measurands from a stable process each measured multiple times using the same device

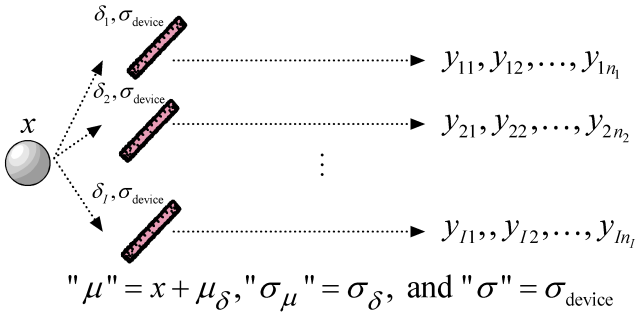


FIGURE 2.16. A single measurand measured multiple times using multiple devices

One-Way ANOVA Identity

$$\sum_{i,j} (y_{ij} - \overline{y})^2 = \sum_i n_i (\overline{y}_i - \overline{y})^2 + \sum_{i,j} (y_{ij} - \overline{y}_i)^2 \tag{2.17}$$

or in shorthand "sum of squares" notation

One-Way ANOVA Identity in Sum of Squares Notation

$$SSTot = SSTr + SSE \tag{2.18}$$

$SSTot$ is a measure of overall raw variability in the whole data set. $SSTot$ is $n-1$ times the overall sample variance computed ignoring the boundaries between samples. SSE is a measure of variability left unaccounted for after taking account of the sample boundaries, and is a multiple of a weighted average of the I sample variances. $SSTr$ is a measure of variation in the sample means $\overline{y}_{i\cdot}$ and is most simply thought of as the difference $SSTot - SSE$. The "sums of squares" SSE and $SSTr$ have respective associated degrees of freedom $n-I$ and $I-1$. The ratios of sums of squares to their degrees of freedom are called "mean squares" and symbolized as

$$MSE = \frac{SSE}{n-I} \text{ and } MSTr = \frac{MSTr}{I-1}\,. \tag{2.19}$$

Confidence limits for the parameter σ^2 of the one-way random effects model can be built on the error mean square. A single-number estimate of σ is

One-Way ANOVA Estimator of σ

$$\hat{\sigma} = \sqrt{MSE} \tag{2.20}$$

and confidence limits for σ are

One-Way ANOVA-based Confidence Limits for σ

$$\hat{\sigma}\sqrt{\frac{n-I}{\chi^2_{\text{upper}}}} \quad \text{and} \quad \hat{\sigma}\sqrt{\frac{n-I}{\chi^2_{\text{lower}}}} \tag{2.21}$$

where the appropriate degrees of freedom are $\nu = n - I$. Further, in the case that all n_i's are the same, i.e. $n_i = m$ for all i, the Satterthwaite approximation can be used to make fairly simple approximate confidence limits for σ_μ. That is, a single number estimator of σ_μ is

$$\hat{\sigma}_\mu = \sqrt{\frac{1}{m}\max\left(0, MSTr - MSE\right)}\,, \tag{2.22}$$

One-Way ANOVA-based Estimator for σ_μ

and with approximate degrees of freedom

$$\hat{\nu} = \frac{m^2 \cdot \hat{\sigma}_\mu^4}{\dfrac{MSTr^2}{I-1} + \dfrac{MSE^2}{n-I}} \tag{2.23}$$

Satterthwaite Approximate df for Use With Limits (2.24)

approximate confidence limits for σ_μ are

$$\hat{\sigma}_\mu\sqrt{\frac{\hat{\nu}}{\chi^2_{\text{upper}}}} \quad \text{and} \quad \hat{\sigma}_\mu\sqrt{\frac{\hat{\nu}}{\chi^2_{\text{lower}}}}\,. \tag{2.24}$$

One-Way ANOVA-based Confidence Limits for σ_μ

Operationally, the mean squares implicitly defined in displays (2.17) through (2.19) are rarely computed "by hand." And given that statistical software is going to be used, rather than employ the methods represented by formulas (2.20) through (2.24), more efficient methods of confidence interval estimation can be used. High quality statistical software (like the open source command line driven `R` package or the commercial menu driven `JMP` package) implements the best known methods of estimation of the parameters σ, σ_μ, and μ (based not on ANOVA methods, but instead on computationally more difficult REML methods) and prints out confidence limits directly.

Example 10 ***Part Hardness.*** *Below are $m = 2$ hardness values (in mm) measured on each of $I = 9$ steel parts by a single operator at a farm implement manufacturer.*

Part	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
	3.30	3.20	3.20	3.25	3.25	3.30	3.15	3.25	3.25
	3.30	3.25	3.30	3.30	3.30	3.30	3.20	3.20	3.30

This is a scenario of the type illustrated in Figure 2.15. Either working "by hand" with formulas (2.17) through (2.19) or reading directly off a report from a statistical package

$$MSE = .001389 \text{ and } MSTr = .003368$$

So using formulas (2.20) and (2.21) (here $n = mI = 18$ so that error degrees of freedom are $n - I = 18 - 9 = 9$) 95% confidence limits for σ_{device} ($= \sigma$ here) are

$$\sqrt{.001389}\sqrt{\frac{9}{19.023}} \text{ and } \sqrt{.001389}\sqrt{\frac{9}{2.700}}$$

i.e.

$$.026\,\text{mm} \text{ and } .068\,\text{mm}$$

Further, using formulas (2.22) through (2.24), Satterthwaite degrees of freedom for $\hat{\sigma}_\mu$ are

$$\hat{\nu} = \frac{\left(2^2\right)\left(\frac{1}{2}\left(.003368 - .001389\right)\right)^2}{\frac{\left(.003368\right)^2}{9-1} + \frac{\left(.001389\right)^2}{18-9}} \approx 2.4$$

and rounding down to 2 degrees of freedom, approximate 95% confidence limits for σ_x ($= \sigma_\mu$ here) are

$$\sqrt{\frac{1}{2}\left(.003368 - .001389\right)}\sqrt{\frac{2}{7.378}} \text{ and } \sqrt{\frac{1}{2}\left(.003368 - .001389\right)}\sqrt{\frac{2}{.051}}$$

i.e.

$$.016\,\text{mm} \text{ and } .197\,\text{mm}$$

The JMP package (using REML methods instead of the Satterthwaite approximation based on ANOVA means squares) produces limits for σ_x

$$0\,\text{mm} \text{ and } \sqrt{.0027603} = .053\,\text{mm}$$

These more reliable limits at least confirm that the simpler methods "get into the right ballpark" in this example.

What is clear from this analysis is that this is a case where part-to-part variation in hardness (measured by σ_x) is small enough and poorly determined enough in comparison to basic measurement noise (measured by σ_{device} estimated as $.03726 = \sqrt{.001389}$) that it is impossible to really tell its size.

Example 11 ***Paper Weighing.*** *Below are $m = 3$ measurements of the weight (in g) of a single $20\,\text{cm} \times 20\,\text{cm}$ piece of $20\,\text{lb}$ bond paper made by each of $I = 5$ different technicians using a single balance.*

Operator	***1***	***2***	***3***	***4***	***5***
	3.481	3.448	3.485	3.475	3.472
	3.477	3.472	3.464	3.472	3.470
	3.470	3.470	3.477	3.473	3.474

This is a scenario of the type illustrated in Figure 2.16 and further illustrates the concepts of repeatability (fixed device) variation and reproducibility (here, device-to-device, i.e. operator-to-operator) variation first discussed on page 44. Use of the JMP

statistical package (and REML estimation) with these data produces 95% confidence limits for the two standard deviations σ_δ ($= \sigma_\mu$ here) and σ_{device} ($= \sigma$ here). These place

$$0 < \sigma_\delta < \sqrt{4.5 \times 10^{-5}} = .0067\,\text{g}$$

and

$$.0057\,\text{g} = \sqrt{3.2 \times 10^{-5}} < \sigma_{device} < \sqrt{.0002014} = .0142\,\text{g}$$

with 95% confidence. This is a case where repeatability variation is clearly larger than reproducibility (operator-to-operator) variation in weight measuring. If one doesn't like the overall size of measurement variation, it appears that some fundamental change in equipment or how it is used will be required. Simple training of the operators aimed at making how they use the equipment more uniform (and reduction of differences between their biases) has far less potential to improve measurement precision.

Section 2.3 Exercises

1. **Fiber Angle**. Grunig, Hamdorf, Herman, and Potthof studied a carpet-like product. Fiber angles (to the backing) were of interest. Operator 1 obtained the values $19, 20, 20$, and 23 (in degrees) from four measurements of fiber angle for a single specimen. This same operator then measured fiber angles once each for three other specimens of the "carpet" and obtained the values $20, 15$ and 23.

   (a) Using the methods of this section, give an estimate of the specimen-to-specimen standard deviation of fiber angle.

   (b) Give the appropriate "approximate degrees of freedom" associated with your estimate from (a). Then find a 95% confidence interval for the specimen-to-specimen fiber angle standard deviation.

2. Continue with the **Fiber Angle** case of problem 1. Operator 2 obtained the fiber angle measurements $20, 25, 17$, and 22 from the first specimen mentioned in problem1 and operator 3 obtained the values $20, 19, 15$, and 16. (Fiber angle for the same specimen was measured four times by each of the three operators.) As before, all measurements were in degrees. The data summaries below are from use of the JMP statistical package with these $n = 12$ measurements of fiber angle for this specimen. Use them to answer (a) through (c). (The estimates and confidence intervals in the second table are for variances, not standard deviations. You will need to take square roots to get inferences for standard deviations.)

ANOVA Table

Source	SS	df	MS
Operator	28.66	2	14.33
Error	60	9	6.66
Total	88.66	11	

REML Variance Component Analysis

Random Effect	VarComponent	95% lower	95% upper
Operator	1.92	−5.27	9.11
Error	6.66	3.15	22.22

(a) Give an appropriate single number estimate of $\sigma_{\text{repeatability}}$. Determine 95% confidence limits for device (repeatability) standard deviation, $\sigma_{\text{repeatability}}$.

(b) From the computer output, give the appropriate estimate of $\sigma_{\text{reproducibility}}$. Give 95% confidence limits for $\sigma_{\text{reproducibility}}$.

(c) Based on your answers to (a) and (b), where would you focus measurement improvement efforts?

3. Continuing with the **Fiber Angle** case, in addition to the repeat measurements $19, 20, 20$, and 23 made by operator 1 on specimen 1, this person also measured angles on 2 other specimens. Four angle measurements on specimen 2 were $15, 17, 20$, and 20 and four angle measurements on specimen 3 were $23, 20, 22$, and 20. The data summaries below are from use of the JMP statistical package with these $n = 12$ measurements for these three specimens. Use them to answer (a) through (c). (The estimates and confidence intervals in the second table are for variances, not standard deviations. You will need to take square roots to get inferences for standard deviations.)

ANOVA Table

Source	SS	df	MS
Specimen	23.17	2	11.58
Error	33.75	9	3.75
Total	56.92	11	

REML Variance Component Analysis

Random Effect	VarComponent	95% lower	95% upper
Specimen	1.96	3.78	7.69
Error	3.75	1.77	12.5

(a) Give an appropriate single number estimate of σ_{device}. Determine 95% confidence limits for device variation, σ_{device}.

(b) From the computer output, give an appropriate estimate of σ_x. Give 95% confidence limits for σ_x.

(c) Based on your answers to (a) and (b), does it seem possible to determine fiber angle for a fixed specimen with acceptable precision? (Hint: Consider the sizes of the estimated σ_{device} and σ_x.)

2.4 Gauge R&R Studies

We have twice made some discussion of "gauge R&R," first on page 44 in the context of comparison of two operators and then in Example 11, where three operators were involved. In both cases, only a single part (or measurand) was considered. In a typical industrial gauge R&R study, each of J operators uses the same gauge or measurement system to measure each of I parts (common to all operators) a total of m different times. Variation in measurement typical of that seen in the m measurements for a particular operator on a particular part is called the **repeatability** variation of the gauge. Variation which can be attributed to differences between the J operators is called **reproducibility** variation of the measurement system.

This section considers the analysis of such full-blown gauge R&R studies involving a total of mIJ measurements. We begin with a discussion of the two-way random effects model that is commonly used to support analyses of gauge R&R data. Then primarily for ease of exposition and making connections to common analyses of gauge R&R studies, we discuss some range-based statistical methods. Finally, we provide what are really superior analyses, based on ANOVA calculations.

2.4.1 Two-Way Random Effects Models and Gauge R&R Studies

Typical industrial gauge R&R data are conveniently thought of as laid out in the cells of a table with I rows corresponding to parts and J columns corresponding to operators.

Example 12 ***Gauge R&R for a 1-Inch Micrometer Caliper.*** *Heyde, Kuebrick, and Swanson conducted a gauge R&R study on a certain micrometer caliper as part of a class project. Table 2.1 shows the data that the $J = 3$ (student) operators obtained, each making $m = 3$ measurements of the heights of $I = 10$ steel punches.*

Notice that even for a given punch/student combination, measured heights are not exactly the same. Further, it is possible to verify that averaging the 30 measurements

TABLE 2.1. Measured Heights of 10 Steel Punches in 10^{-3} Inch

	Student 1	**Student 2**	**Student 3**
Punch 1	496, 496, 499	497, 499, 497	497, 498, 496
Punch 2	498, 497, 499	498, 496, 499	497, 499, 500
Punch 3	498, 498, 498	497, 498, 497	496, 498, 497
Punch 4	497, 497, 498	496, 496, 499	498, 497, 497
Punch 5	499, 501, 500	499, 499, 499	499, 499, 500
Punch 6	499, 498, 499	500, 499, 497	498, 498, 498
Punch 7	503, 499, 502	498, 499, 499	500, 499, 502
Punch 8	500, 499, 499	501, 498, 499	500, 501, 499
Punch 9	499, 500, 499	500, 500, 498	500, 499, 500
Punch 10	497, 496, 496	500, 494, 496	496, 498, 496

made by student 1, a mean of about $.49853\,\text{in}$ *is obtained, while corresponding means for students 2 and 3 are respectively about* $.49813\,\text{in}$ *and* $.49840\,\text{in}$. *Student 1 may tend to measure slightly higher than students 2 and 3. That is, by these rough "eyeball" standards, there is some hint in these data of both repeatability and reproducibility components in the overall measurement imprecision.*

To this point in our discussions of R&R, we have not involved more than a single measurand. Effectively, we have confined attention to a single row of a table like Table 2.1. Standard industrial gauge R&R studies treat multiple parts as a way of making sure that reliability of measurement doesn't vary unacceptably across parts. So here we consider the kind of multiple-part case represented in Table 2.1.

The model most commonly used in this context is the so-called "two-way random effects model" that can be found in many intermediate-level statistical methods texts. (See, for example, Section 8.4 of Vardeman's *Statistics for Engineering Problem Solving*.) Let

$$y_{ijk} = \text{the } k\text{th measurement made by operator } j \text{ on part } i\ .$$

Two-Way Random Effects Model

The model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}\ , \tag{2.25}$$

where μ is an (unknown) constant, the α_i are normal random variables with mean 0 and variance σ^2_α, the β_j are normal random variables with mean 0 and variance σ^2_β, the $\alpha\beta_{ij}$ are normal random variables with mean 0 and variance $\sigma^2_{\alpha\beta}$, the ϵ_{ijk} are normal random variables with mean 0 and variance σ^2, and all of the α's, β's, $\alpha\beta$'s, and ϵ's are independent. In this model, the unknown constant μ is an average (over all possible operators and all possible parts) measurement, the α's are (random) effects of different parts, the β's are (random) effects of different operators, the $\alpha\beta$'s are (random) joint effects peculiar to particular part×operator combinations, and the ϵ's are (random) measurement errors. The variances $\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha\beta}$, and σ^2 are called "variance components" and their sizes govern how much variability is seen in the measurements y_{ijk}.

Consider a hypothetical case with $I = 2$, $J = 2$, and $m = 2$. Model (2.25) says that there is a normal distribution with mean 0 and variance σ^2_α from which α_1 and α_2 are drawn. And there is a normal distribution with mean 0 and variance σ^2_β from which β_1 and β_2 are drawn. And there is a normal distribution with mean 0 and variance $\sigma^2_{\alpha\beta}$ from which $\alpha\beta_{11}$, $\alpha\beta_{12}$, $\alpha\beta_{21}$, and $\alpha\beta_{22}$ are drawn. And there is a normal distribution with mean 0 and variance σ^2 from which eight ϵ's are drawn. Then these realized values of the random effects are added to produce the eight measurements as indicated in Table 2.2.

Repeatability Standard Deviation in the Two-Way Model

Either directly from equation (2.25) or as illustrated in Table 2.2, according to the two-way random effects model the only differences between measurements for a fixed part×operator combination are the measurement errors ϵ. And the variability of these

TABLE 2.2. Measurements in a Hypothetical Gauge R&R Study

	Operator 1	**Operator 2**
Part 1	$y_{111} = \mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \epsilon_{111}$ $y_{112} = \mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \epsilon_{112}$	$y_{121} = \mu + \alpha_1 + \beta_2 + \alpha\beta_{12} + \epsilon_{121}$ $y_{122} = \mu + \alpha_1 + \beta_2 + \alpha\beta_{12} + \epsilon_{122}$
Part 2	$y_{211} = \mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \epsilon_{211}$ $y_{212} = \mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \epsilon_{212}$	$y_{221} = \mu + \alpha_2 + \beta_2 + \alpha\beta_{22} + \epsilon_{221}$ $y_{222} = \mu + \alpha_2 + \beta_2 + \alpha\beta_{22} + \epsilon_{222}$

is governed by the parameter σ. That is, σ is a measure of repeatability variation in this model, and one objective of an analysis of gauge R&R data is to estimate it.

Then, if one looks at a fixed "part i" (row i), the quantity $\mu + \alpha_i$ is common across the row. In the context of a gauge R&R study this can be interpreted as the value of the ith measurand (these vary across parts/rows because the α_i vary). Then, still for a fixed part i, it is the values $\beta_j + \alpha\beta_{ij}$ that vary column/operator to column/operator. so in this gauge R&R context, this quantity functions as a kind of *part-i-specific operator bias*. (More on the qualifier "part-i-specific" in a bit.) According to model (2.25), the variance of $\beta_j + \alpha\beta_{ij}$ is $\sigma_\beta^2 + \sigma_{\alpha\beta}^2$, so an appropriate measure of reproducibility variation in this model is

Reproducibility Standard Deviation in the Two-Way Model

$$\sigma_{\text{reproducibility}} = \sqrt{\sigma_\beta^2 + \sigma_{\alpha\beta}^2}\,. \tag{2.26}$$

According to the model, this is the standard deviation that would be experienced by many operators making a single measurement on the same part *assuming that there is no repeatability component to the overall variation*. Another way to say the same thing is to recognize this quantity as the standard deviation that would be experienced computing with long-run average measurements for many operators on the same part. That is, the quantity (2.26) is a measure of variability in operator bias for a fixed part in this model.

As long as one confines attention to a single row of a standard gauge R&R study, the one-way random effects model and analysis of Section 2.3 are relevant. The quantity $\sigma_{\text{reproducibility}}$ here is exactly σ_δ from application of the one-way model to a single-part gauge R&R study. (And the present σ is exactly σ_{device}.) What is new and at first perhaps a bit puzzling is that in the present context of multiple parts and display (2.26), the reproducibility variation has two components, σ_β and $\sigma_{\alpha\beta}$. This is because for a given part i, the model says that bias for operator j has both components β_j and $\alpha\beta_{ij}$. The model terms $\alpha\beta_{ij}$ allow "operator bias" to change part-to-part/measurand-to-measurand (an issue that simply doesn't arise in the context of a single part study). As such, they are *a measure of non-linearity* (bias non-constant in the measurand) in the overall measurement system. Two-way data like those in Table 2.1 allow one to estimate all of $\sigma_{\text{reproducibility}}$, σ_β and $\sigma_{\alpha\beta}$, and all else being equal, cases where the $\sigma_{\alpha\beta}$ component of $\sigma_{\text{reproducibility}}$ is small are preferable to those where it is large.

The quantity

Combined R&R Standard Deviation

$$\sigma_{\text{R\&R}} = \sqrt{\sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2} = \sqrt{\sigma_{\text{reproducibility}}^2 + \sigma^2} \tag{2.27}$$

is the standard deviation implied by the model (2.25) for many operators each making a single measurement on the same part. That is, quantity (2.27) is a measure of the combined imprecision in measurement attributable to *both* repeatability and reproducibility sources. And one might think of

$$\frac{\sigma^2}{\sigma^2_{\text{R\&R}}} = \frac{\sigma^2}{\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2} \quad \text{and} \quad \frac{\sigma^2_{\text{reproducibility}}}{\sigma^2_{\text{R\&R}}} = \frac{\sigma^2_\beta + \sigma^2_{\alpha\beta}}{\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2} \tag{2.28}$$

as the fractions of total measurement variance due respectively to repeatability and reproducibility. If one can produce estimates of σ and $\sigma_{\text{reproducibility}}$, estimates of these quantities (2.27) and (2.28) follow in straightforward fashion.

It is common to treat some multiple of $\sigma_{\text{R\&R}}$ (often the multiplier is six, but sometimes 5.15 is used) as a kind of uncertainty associated with a measurement made using the gauge or measurement system in question. And when a gauge is being used to check conformance of a part dimension or other measured characteristic to engineering specifications (say, some lower specification L and some upper specification U) this multiple is compared to the spread in specifications. **Specifications** U and L are numbers set by product design engineers that are supposed to delineate what is required of a measured dimension in order that the item in question be functional. The hope is that measurement uncertainty is at least an order of magnitude smaller than the spread in specifications. Some organizations go so far as to call the quantity

Engineering Specifications

Gauge Capability Ratio

$$GCR = \frac{6\sigma_{\text{R\&R}}}{U - L} \tag{2.29}$$

a **gauge capability (or precision-to-tolerance) ratio,** and require that it be no larger than .1 (and preferably as small as .01) before using the gauge for checking conformance to such specifications. (In practice, one will only have an estimate of $\sigma_{\text{R\&R}}$ upon which to make an empirical approximation of a gauge capability ratio.)

2.4.2 Range-Based Estimation

Because range-based estimation (similar to, but not exactly the same as what follows) is in common use for the analysis of gauge R&R studies and is easy to describe, we will treat it here. In the next sub-section, better methods based on ANOVA calculations (and REML methods) will be presented.

Consider first the estimation of σ. Restricting attention to any particular part$\times$operator combination, say part i and operator j, model (2.25) says that observations obtained for that combination differ only by independent normal random measurement error with mean 0 and variance σ^2. That suggests that a measure of variability for the ij sample might be used as the basis of an estimator of σ. Historical precedent and ease of computation suggest measuring variability using a range (instead of a sample standard deviation or variance).

So let R_{ij} be the range of the m measurements on part i by operator j. The expected value of the range of a sample from a normal distribution is a constant (depending upon

m) times the standard deviation of the distribution being sampled. The constants are well known and called d_2. (We will write $d_2(m)$ to emphasize their dependence upon m and note that values of $d_2(m)$ are given in Table A.1.) It then follows that

$$\mathrm{E}R_{ij} = d_2(m)\sigma,$$

which in turn suggests that the ratio

$$\frac{R_{ij}}{d_2(m)}$$

is a plausible estimator of σ. Better yet, one might average these over all $I \times J$ part×operator combinations to produce the range-based estimator of σ,

$$\hat{\sigma}_{\text{repeatability}} = \frac{\overline{R}}{d_2(m)} \; . \tag{2.30}$$

Range-Based Estimator for Repeatability Standard Deviation

Example 13 ***(Example 12 continued.)*** *Subtracting the smallest measurement for each part×operator combination in Table 2.1 from the largest for that combination, one obtains the ranges in Table 2.3 on page 68. The 30 ranges in Table 2.3 have mean* $\overline{R} = 1.9$. *From Table A.1,* $d_2(3) = 1.693$. *So using expression (2.30) an estimate of* σ, *the repeatability standard deviation for the caliper used by the students, is*

$$\hat{\sigma}_{\text{repeatability}} = \frac{\overline{R}}{d_2(3)} = \frac{1.9}{1.693} = 1.12 \times 10^{-3} \text{ in.}$$

(Again, this is an estimate of the (long-run) standard deviation that would be experienced by any particular student measuring any particular punch many times.)

Consider now the standard deviation (2.26) representing the reproducibility portion of the gauge imprecision. It will be convenient to have some additional notation. Let

$$\overline{y}_{ij} = \text{the (sample) mean measurement made on part } i \text{ by operator } j \tag{2.31}$$

and

$$\begin{aligned} \Delta_i &= \max_j \overline{y}_{ij} - \min_j \overline{y}_{ij} \\ &= \text{the range of the mean measurements made on part } i \; . \end{aligned}$$

Notice that with the obvious notation for the sample average of the measurement errors ϵ, according to model (2.25)

$$\overline{y}_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \overline{\epsilon}_{ij} \; .$$

TABLE 2.3. Ranges of 30 Part× Operator Samples of Measured Punch Heights

Punch	Student 1	Student 2	Student 3
1	3	2	2
2	2	3	3
3	0	1	2
4	1	3	1
5	2	0	1
6	1	3	0
7	4	1	3
8	1	3	2
9	1	2	1
10	1	6	2

Thus, for a fixed part i these means $\overline{y}_{ij}$ vary only according to independent normal random variables $\beta_j + \alpha\beta_{ij} + \overline{\epsilon}_{ij}$ that have mean 0 and variance $\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2/m$. Thus their range, Δ_i, has mean

$$\mathrm{E}\Delta_i = d_2(J)\sqrt{\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2/m}\,.$$

This suggests $\Delta_i/d_2(J)$, or better yet, the average of these over all parts i, $\overline{\Delta}/d_2(J)$, as an estimator of $\sqrt{\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2/m}$. This in turn suggests that one can estimate $\sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2/m$ with $(\overline{\Delta}/d_2(J))^2$. Then remembering that $\overline{R}/d_2(m) = \hat{\sigma}_{\text{repeatability}}$ is an estimator of σ, an obvious estimator of $\sigma^2_\beta + \sigma^2_{\alpha\beta}$ becomes

$$\left(\frac{\overline{\Delta}}{d_2(J)}\right)^2 - \frac{1}{m}\left(\frac{\overline{R}}{d_2(m)}\right)^2. \tag{2.32}$$

The quantity (2.32) is meant to approximate $\sigma^2_\beta + \sigma^2_{\alpha\beta}$, which is nonnegative. But the estimator (2.32) can on occasion give negative values. When this happens, it is sensible to replace the negative value by 0, and thus expression (2.32) by

$$\max\left(0, \left(\frac{\overline{\Delta}}{d_2(J)}\right)^2 - \frac{1}{m}\left(\frac{\overline{R}}{d_2(m)}\right)^2\right). \tag{2.33}$$

So finally, an estimator of the reproducibility standard deviation can be had by taking the square root of expression (2.33). That is, one may estimate the quantity (2.26) with

Range-Based Estimator for Reproducibility Standard Deviation

$$\hat{\sigma}_{\text{reproducibility}} = \sqrt{\max\left(0, \left(\frac{\overline{\Delta}}{d_2(J)}\right)^2 - \frac{1}{m}\left(\frac{\overline{R}}{d_2(m)}\right)^2\right)}. \tag{2.34}$$

TABLE 2.4. Part×Operator Means and Ranges of Such Means for the Punch Height Data

Punch (i)	$\overline{y}_{i1}$	$\overline{y}_{i2}$	$\overline{y}_{i3}$	Δ_i
1	497.00	497.67	497.00	.67
2	498.00	497.67	498.67	1.00
3	498.00	497.33	497.00	1.00
4	497.33	497.00	497.33	.33
5	500.00	499.00	499.33	1.00
6	498.67	498.67	498.00	.67
7	501.33	498.67	500.33	2.67
8	499.33	499.33	500.00	.67
9	499.33	499.33	499.67	.33
10	496.33	496.67	496.67	.33

Example 14 ***(Examples 12 and 13 continued.)*** *Table 2.4 organizes* $\overline{y}_{ij}$ *and* Δ_i *values for the punch height measurements of Table 2.1. Then* $\overline{\Delta} = 8.67/10 = .867$*, and since* $J = 3, d_2(J) = d_2(3) = 1.693$*. So using equation (2.34),*

$$\begin{aligned}\hat{\sigma}_{\text{reproducibility}} &= \sqrt{\max\left(0, \left(\frac{.867}{1.693}\right)^2 - \frac{1}{3}\left(\frac{1.9}{1.693}\right)^2\right)}\,, \\ &= \sqrt{\max(0, -.158)}\,, \\ &= 0\,.\end{aligned}$$

This calculation suggests that this is a problem where σ *appears to be so large that the reproducibility standard deviation cannot be seen above the intrinsic "noise" in measurement conceptualized as the repeatability component of variation. Estimates of the ratios (2.28) based on* $\hat{\sigma}_{\text{repeatability}}$ *and* $\hat{\sigma}_{\text{reproducibility}}$ *would attribute fractions* 1 *and* 0 *of the overall variance in measurement to respectively repeatability and reproducibility.*

2.4.3 ANOVA-Based Estimation

The formulas of the previous sub-section are easy to discuss and use, but they are not at all the best available. Ranges are not the most effective tools for estimating normal standard deviations. And the range-based methods have no corresponding way for making confidence intervals. More effective (and computationally more demanding) statistical tools are available and we proceed to discuss some of them.

An $I \times J \times m$ data set of y_{ijk}'s like that produced in a typical gauge R&R study is often summarized in a so-called two-way ANOVA table. Table 2.5 on page 70 is a generic version of such a summary. Any decent statistical package will process a gauge R&R data set and produce such a summary table. As in one-way ANOVA, "mean squares" are essentially sample variances (squares of sample standard deviations). MSA is essentially a sample variance of part averages, MSB is essentially a sample variance of

TABLE 2.5. A Generic Gauge R&R Two-Way ANOVA Table

Source	SS	df	MS
Part	SSA	$I-1$	$MSA = SSA/(I-1)$
Operator	SSB	$J-1$	$MSB = SSB/(J-1)$
Part×Operator	$SSAB$	$(I-1)(J-1)$	$MSAB = SSAB/(I-1)(J-1)$
Error	SSE	$IJ(m-1)$	$MSE = SSE/IJ(m-1)$
Total	$SSTot$	$IJm-1$	

operator averages, MSE is an average of within-cell sample variances, $MSTot$ isn't typically calculated, but is a grand sample variance of all observations.

For purposes of being clear (and not because they are typically used for "hand calculation") we provide formulas for sums of squares. With cell means $\overline{y}_{ij}$ as in display (2.31) define row and column averages and the grand average of these

$$\overline{y}_{i\cdot} = \frac{1}{J}\sum_j \overline{y}_{ij} \text{ and } \overline{y}_{\cdot j} = \frac{1}{I}\sum_i \overline{y}_{ij} \text{ and } \overline{y}_{\cdot\cdot} = \frac{1}{IJ}\sum_{ij} \overline{y}_{ij}\ .$$

Then the sums of squares are

$$\begin{aligned}
SSTot &= \sum_{ijk}\left(y_{ijk}-\overline{y}_{\cdot\cdot}\right)^2\ , \\
SSE &= \sum_{ijk}\left(y_{ijk}-\overline{y}_{ij}\right)^2\ , \\
SSA &= mJ\sum_i\left(\overline{y}_{i\cdot}-\overline{y}_{\cdot\cdot}\right)^2\ , \\
SSB &= mI\sum_j\left(\overline{y}_{\cdot j}-\overline{y}_{\cdot\cdot}\right)^2\ , \text{ and} \\
SSAB &= m\sum_{ij}\left(\overline{y}_{ij}-\overline{y}_{i\cdot}-\overline{y}_{\cdot j}+\overline{y}_{\cdot\cdot}\right)^2 \\
&= SSTot-SSE-SSA-SAB
\end{aligned}$$

Corresponding degrees of freedom and mean squares are

$$\begin{aligned}
dfE &= (m-1)IJ \text{ and } MSE = SSE/(m-1)IJ\ , \\
dfA &= I-1 \text{ and } MSA = SSA/(I-1)\ , \\
dfB &= J-1 \text{ and } MSB = SSB/(J-1)\ , \text{ and} \\
dfAB &= (I-1)(J-1) \text{ and } MSAB = SSAB/(I-1)(J-1)\ .
\end{aligned}$$

Example 15 ***In-Class Gauge R&R Study.*** *The data in Table 2.6 were collected in an in-class gauge R&R exercise where $I = 4$ Styrofoam packing peanuts were measured for size (in* in*) by $J = 3$ students $m = 2$ times apiece using the same inexpensive caliper. The* JMP *statistical package produces the sums of squares*

$$\begin{aligned}
SSA &= .00241250, SSB = .00080833, SSAB = .00072500, \\
SSE &= .00035000, \text{ and } SSTot = .00429583
\end{aligned}$$

TABLE 2.6. Data from a Small In-Class Gauge R&R Study

	Operator 1	**Operator 2**	**Operator 3**
Part 1	.52, .52	.54, .53	.55, .55
Part 2	.56, .55	.54, .54	.55, .56
Part 3	.57, .56	.55, .56	.57, .57
Part 4	.55, .55	.54, .55	.56, .55

for these data that can be used as raw material for making important inferences for the R&R study based on model (2.25). Corresponding means squares are

$$\begin{aligned} MSE &= .00035000/(2-1)(4)(3) = .00002917\,, \\ MSA &= .00241250/(4-1) = .00080417\,, \\ MSB &= .00080833/(3-1) = .00040417\,, \text{ and} \\ MSAB &= .00072500/(4-1)(3-1) = .00012083\,. \end{aligned}$$

High quality statistical software (like JMP or R) will automatically produce REML-based estimates and confidence intervals for the variance components $\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha\beta}$, and σ^2. As the quantities $\sigma^2_{\text{reproducibility}}$ and $\sigma^2_{\text{R\&R}}$ are a bit specialized (being of interest in our R&R application of the two-way random effects model, but not in other common applications) inferences for them are not automatically available. It is possible, but usually not convenient, to use the output of REML analyses to make inferences for these more specialized quantities. So here we will provide formulas for ANOVA-based estimators of $\sigma, \sigma_{\text{reproducibility}}$, and $\sigma_{\text{R\&R}}$ and appropriate Satterthwaite approximate degrees of freedom for making confidence limits. (Where readers know how to obtain REML-based estimates and intervals, our recommendation is to use them in preference to ANOVA-based estimators that follow.)

Single number estimators for the quantities of most interest in a gauge R&R study are

$$\hat{\sigma}_{\text{repeatability}} = \hat{\sigma} = \sqrt{MSE}\,, \tag{2.35}$$

ANOVA-Based Estimator for Repeatability Standard Deviation

$$\hat{\sigma}_{\text{reproducibility}} = \sqrt{\max\left(0, \frac{MSB}{mI} + \frac{(I-1)}{mI}MSAB - \frac{1}{m}MSE\right)}\,, \tag{2.36}$$

ANOVA-Based Estimator for Reproducibility Standard Deviation

and

ANOVA-Based Estimator for $\sigma_{\text{R\&R}}$

$$\hat{\sigma}_{\text{R\&R}} = \sqrt{\frac{1}{mI}MSB + \frac{I-1}{mI}MSAB + \frac{m-1}{m}MSE}\,. \tag{2.37}$$

Confidence limits based on any of these estimators are of the generic form (already used several times in this chapter)

Generic Confidence Limts for a Standard Deviation

$$\text{"}\hat{\sigma}\text{"}\sqrt{\frac{\text{"}\hat{\nu}\text{"}}{\chi^2_{\text{upper}}}} \qquad \text{and} \qquad \text{"}\hat{\sigma}\text{"}\sqrt{\frac{\text{"}\hat{\nu}\text{"}}{\chi^2_{\text{lower}}}} \tag{2.38}$$

where "$\hat{\sigma}$" is one of the estimators, "$\hat{\nu}$" is a corresponding (exact or "Satterthwaite approximate") degrees of freedom, and the χ^2 percentage points are based on "$\hat{\nu}$." So it only remains to record formulas for appropriate degrees of freedom. These are

Degrees of Freedom for Use With Formulas (2.35) and (2.38)

$$\nu_{\text{repeatability}} = IJ\,(m-1)\ , \tag{2.39}$$

Degrees of Freedom for Use With Formulas (2.36) and (2.38)

$$\begin{aligned}\hat{\nu}_{\text{reproducibility}} &= \frac{\hat{\sigma}^4_{\text{reproducibility}}}{\frac{\left(\frac{MSB}{mI}\right)^2}{J-1} + \frac{\left(\frac{(I-1)MSAB}{mI}\right)^2}{(I-1)(J-1)} + \frac{\left(\frac{MSE}{m}\right)^2}{IJ(m-1)}} \\ &= \frac{\hat{\sigma}^4_{\text{reproducibility}}}{\frac{1}{m^2}\left(\frac{MSB^2}{I^2(J-1)} + \frac{(I-1)MSAB^2}{I^2(J-1)} + \frac{MSE^2}{IJ(m-1)}\right)}\ ,\end{aligned} \tag{2.40}$$

and

Degrees of Freedom for Use With Formulas (2.37) and (2.38)

$$\hat{\nu}_{\text{R\&R}} = \frac{\hat{\sigma}^4_{\text{R\&R}}}{\frac{\left(\frac{MSB}{mI}\right)^2}{J-1} + \frac{\left(\frac{(I-1)MSAB}{mI}\right)^2}{(I-1)(J-1)} + \frac{\left(\frac{(m-1)MSE}{m}\right)^2}{IJ(m-1)}}$$
$$= \frac{\hat{\sigma}^4_{\text{R\&R}}}{\frac{1}{m^2}\left(\frac{MSB^2}{I^2(J-1)} + \frac{(I-1)MSAB^2}{I^2(J-1)} + \frac{(m-1)MSE^2}{IJ}\right)} . \quad (2.41)$$

Formulas (2.36), (2.40), (2.37) and (2.41) are tedious (but hardly impossible) to use with a pocket calculator. But a very small program, MathCAD worksheet, or spreadsheet template can be written to evaluate the estimates of standard deviations and approximate degrees of freedom from the sums of squares, m, I, and J.

Example 16 ***(Example 15 continued.)*** *A two-way random effects analysis of the data of Table 2.6 made using the JMP statistical package produces REML-based confidence limits of*

$$0 \text{ and } \sqrt{.0001359}, \text{ i.e., } 0\,\text{in} \text{ and } .012\,\text{in} \text{ for } \sigma_\beta$$

and

$$0 \text{ and } \sqrt{.0001152}, \text{ i.e., } 0\,\text{in} \text{ and } .011\,\text{in} \text{ for } \sigma_{\alpha\beta} .$$

There is thus at least the suggestion that a substantial part of the reproducibility variation in the data of Table 2.6 is a kind of non-constant bias on the part of the student operators measuring the peanuts.

Using formulas (2.35), (2.36), and (2.37) it is possible to verify that in this problem

$$\begin{aligned} \hat{\sigma}_{\textit{repeatability}} &= \hat{\sigma} = .005401\,\text{in} , \\ \hat{\sigma}_{\textit{reproducibility}} &= .009014\,\text{in} , \text{ and} \\ \hat{\sigma}_{\textit{R\&R}} &= .011\,\text{in} . \end{aligned}$$

Using formulas (2.39), (2.40), and (2.41) these have corresponding degrees of freedom

$$\begin{aligned} \nu_{\textit{repeatability}} &= (4)(3)(2-1) = 12 , \\ \hat{\nu}_{\textit{reproducibility}} &= 4.04 , \text{ and} \\ \hat{\nu}_{\textit{R\&R}} &= 7.45 . \end{aligned}$$

So (rounding degrees of freedom down in the last two cases) using the limits (2.38), 95% confidence limits for $\sigma_{\textit{repeatability}}$ are

$$.005401\sqrt{\frac{12}{23.337}} \text{ and } .005401\sqrt{\frac{12}{4.404}}$$

i.e.

$$.0039\,\text{in} \;\; \textit{and} \;\; .0089\,\text{in}\,,$$

approximate 95% confidence limits for $\sigma_{\textit{reproducibility}}$ *are*

$$.009014\sqrt{\frac{4}{11.143}} \;\; \textit{and} \;\; .009014\sqrt{\frac{4}{.484}}$$

i.e.

$$.0054\,\text{in} \;\; \textit{and} \;\; .0259\,\text{in}\,,$$

and approximate 95% confidence limits for $\sigma_{\textit{R\&R}}$ *are*

$$.011\sqrt{\frac{7}{16.013}} \;\; \textit{and} \;\; .011\sqrt{\frac{7}{1.690}}$$

i.e.

$$.0073\,\text{in} \;\; \textit{and} \;\; .0224\,\text{in}\,.$$

These intervals show that none of these standard deviations are terribly well-determined (degrees of freedom are small and intervals are wide). If better information is needed, more data would have to be collected. But there is at least some indication that $\sigma_{\textit{repeatability}}$ *and* $\sigma_{\textit{reproducibility}}$ *are roughly of the same order of magnitude. The caliper used to make the measurements was a fairly crude one, and there were detectable differences in the way the student operators used that caliper.*

Suppose, for sake of example, that engineering requirements on these Styrofoam peanuts were that they be of size $.50\,\text{in} \pm .05\,\text{in}$. *In such a context, the gauge capability ratio (2.29) could be estimated to be between*

$$\frac{6\,(.0073)}{.65-.55} = .44 \;\; \textit{and} \;\; \frac{6\,(.0224)}{.65-.55} = 1.34$$

These values are not small. (See again the discussion on page 66.) This measurement "system" is not really adequate to check conformance to even these crude $\pm .05\,\text{in}$ *product requirements.*

Some observations regarding the planning of a gauge R&R study are in order at this point. The precisions with which one can estimate σ, $\sigma_{\text{reproducibility}}$, and $\sigma_{\text{R\&R}}$ obviously depend upon I, J, and m. Roughly speaking, precision of estimation of σ is governed by the product $(m-1)IJ$, so increasing any of the "dimensions" of the data array will improve estimation of repeatability. However, it is primarily J that governs the precision with which $\sigma_{\text{reproducibility}}$ and $\sigma_{\text{R\&R}}$ can be estimated. Only by increasing the number of operators in a gauge R&R study can one substantially improve the estimation of reproducibility variation.

While this fact about the estimation of reproducibility is perfectly plausible, its implications are not always fully appreciated (or at least not kept clearly in mind) by quality assurance practitioners. For example, many standard gauge R&R data collection forms allow for at most $J = 3$ operators. But three is a *very* small sample size

when it comes to estimating a variance or standard deviation. So although the data in Table 2.1 are perhaps more or less typical of many R&R data sets, the small ($J = 3$) number of operators evident there should *not* be thought of as in any way ideal. To get a really good handle on the size of reproducibility variation, many more operators would be needed.

Section 2.4 Exercises

1. Consider again the situation of problem 3 of the Section 2.3 exercises and the data from the **Fiber Angle** case used there. (Operator 1 measured fiber angle for 3 different specimens 4 times each.) Recast that scenario into the two-way framework of this section.

   (a) Give the values of I, J, and m.

   (b) Find a range-based estimate of σ_{device} .

   (c) Find a range-based estimate of σ_x.

2. Based only on the data of problem 3 of the Section 2.3 exercises, can $\sigma_{\text{reproducibility}}$ be estimated? Why or why not?

3. Consider again the situation of problems 2 and 1 of the Section 2.3 exercises and the data from the **Fiber Angle** case used there. (Fiber angle for specimen 1 was measured 4 times by each of operators 1, 2, and 3.) Recast that scenario into the two-way framework of this section.

   (a) Give the values of I, J, and m.

   (b) Find a range-based estimate of $\sigma_{\text{repeatability}}$.

   (c) Find a range-based estimate of $\sigma_{\text{reproducibility}}$.

   (d) Based only on the data considered here, can σ_x be estimated? Why or why not?

4. **Washer Assembly**. Sudam, Heimer and Mueller studied a clothes washer base assembly. Two operators measured the distance from one edge of a washer base assembly to an attachment. For a single base assembly, the same distance was measured four times by each operator. This was repeated on 3 different base assemblies. The target distance was 13.320 with an upper specification of $U = 13.42$ and a lower specification of of $L = 13.22$. A standard gauge R & R study was conducted and data like those below were obtained. (Units are 10^{-1} in.)

	Operator 1	**Operator 2**
Part 1	13.285, 13.284, 13.283, 13.282	13.284, 13.288, 13.287, 13.283
Part 2	13.298, 13.293, 13.291, 13.291	13.297, 13.292, 13.292, 13.293
Part 3	13.357, 13.356, 13.354, 13.356	13.355, 13.354, 13.352, 13.357

(a) What were the values of I, J, and m in this study?

(b) Based on the ANOVA table for the data given below, find the estimates for $\sigma_{\text{repeatability}}$, $\sigma_{\text{reproducibility}}$, and $\sigma_{\text{R\&R}}$.

(c) Give 95% confidence limits for $\sigma_{\text{repeatability}}$, $\sigma_{\text{reproducibility}}$ and $\sigma_{\text{R\&R}}$.

(d) Find 95% confidence limits for the GCR. (Hint: Use the last of your answers to (c).)

ANOVA Table

Source	**SS**	**df**	**MS**
Part	.0236793	2	.0118396
Operator	.0000007	1	.0000007
Part×Operator	.0000106	2	.0000053
Error	.0000895	18	.0000050
Total	.0237800	23	

2.5 Simple Linear Regression and Calibration Studies

Calibration is an essential activity in the qualification and maintenance of measurement devices. In a calibration study, one uses a measurement device to produce measurements on "standard" specimens with (relatively well-) "known" values of measurands, and sees how the measurements compare to the known values. If there are systematic discrepancies between what is known to be true and what the device reads, a conversion scheme is created to (in future use of the device) adjust what is read to something that is hopefully closer to the (future) truth. A slight extension of "regression" analysis (curve fitting) as presented in an elementary statistics course is the relevant statistical methodology in making this conversion. (See, for example, Section 9.1 of Vardeman and Jobe's *Basic Engineering Data Collection and Analysis*.) This section discusses exactly how regression analysis is used in calibration.

Calibration studies employ true/gold-standard-measurement values of measurands x and "local" measurements y. (Strictly speaking, y need not even be in the same units as x.) Regression analysis can provide both "point conversions" and measures of uncertainty (the latter through inversion of "prediction limits"). The simplest version of this is where observed measurements are approximately linearly related to measurands, i.e.

$$y \approx \beta_0 + \beta_1 x$$

This is "linear calibration." The standard statistical model for such a circumstance is

Simple Linear Regression Model

$$y = \beta_0 + \beta_1 x + \epsilon \tag{2.42}$$

for a normal error ϵ with mean 0 and standard deviation σ. (σ describes how much y's vary for a fixed x, and in the present context typically amounts to a repeatability standard deviation.) This model can be pictured as in Figure 2.17.

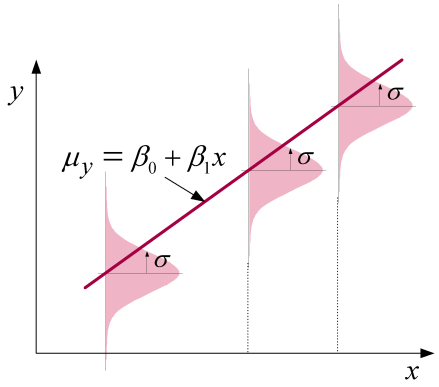


FIGURE 2.17. A schematic of the usual simple linear regression model (2.42)

For n data pairs (x_i, y_i), simple linear regression methodology allows one to make confidence intervals and tests associated with the model, and *prediction limits for a new measurement* y_{new} *associated with a new measurand*, x_{new}. These are of the form

Prediction Limits for y_{new} in SLR

$$(b_0 + b_1 x_{\text{new}}) \pm t s_{\text{LF}} \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{new}} - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}} \tag{2.43}$$

where the least squares line is $\hat{y} = b_0 + b_1 x$ and s_{LF} (a "line-fitting" sample standard deviation) is an estimate of σ derived from the fit of the line to the data. Any good statistical package will compute and plot these limits as functions of x_{new} along with a least squares line through the data set.

Example 17 ***Measuring Cr^{6+} Concentration With a UV-vis Spectrophotometer.*** *The data below were taken from a Web page of the School of Chemistry at the University of Witwatersrand developed and maintained by Dr. Dan Billing. They are measured absorbance values, y, for $n = 6$ solutions with "known" Cr^{6+} concentrations, x (in mg/l), from an analytical lab.*

x	0	1	2	4	6	8
y	.002	.078	.163	.297	.464	.600

Figure 2.18 on page 78 is a plot of these data, the corresponding least squares line, and the prediction limits (2.43).

What is here of most interest about simple linear regression technology is what it says about calibration and measurement in general. Some applications of inference methods based on the model (2.42) to metrology are the following.

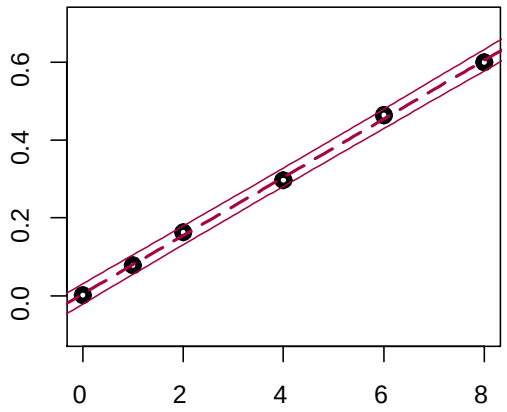


FIGURE 2.18. Scatterplot of the Cr^{6+} Concentration calibration data, least squares line, and prediction limits for y_{new}

1. From a simple linear regression output,

$$s_{\text{LF}} = \sqrt{MSE} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \text{"root mean square error"} \quad (2.44)$$

is an estimated repeatability standard deviation. One may make confidence intervals for $\sigma = \sigma_{\text{repeatability}}$ based on the estimate (2.44) using $\nu = n - 2$ degrees of freedom and limits

Confidence Limits for σ in Model (2.42)

$$s_{\text{LF}}\sqrt{\frac{n-2}{\chi^2_{\text{upper}}}} \quad \text{and} \quad s_{\text{LF}}\sqrt{\frac{n-2}{\chi^2_{\text{lower}}}}\,. \quad (2.45)$$

2. The least squares equation $\hat{y} = b_0 + b_1 x$ can be solved for x, giving

Coversion Formula for a Future Measurment, y_{new}

$$\hat{x}_{\text{new}} = \frac{y_{\text{new}} - b_0}{b_1} \quad (2.46)$$

as a way of estimating a new "gold-standard" value (a new measurand, x_{new}) from a measured local value, y_{new}.

3. One can take the prediction limits (2.43) for y_{new} and "turn them around" to get confidence limits for the x_{new} corresponding to a measured local y_{new}. This provides a defensible way to set "error bounds" on what y_{new} indicates about x_{new}.

4. In cases (unlike Example 17) where y and x are in the same units, confidence limits for the slope β_1 of the simple linear regression model

$$b_1 \pm t \frac{s_{\text{LF}}}{\sqrt{\sum (x_i - \bar{x})^2}} \tag{2.47}$$

Confidence Limits for β_1 in Model (2.42)

provide a way of investigating the constancy of bias (linearity of the measurement device in the sense introduced on page 36). That is, when x and y are in the same units, $\beta_1 = 1.0$ is the case of constant bias. If confidence limits for β_1 fail to include 1.0, there is clear evidence of device nonlinearity.

Example 18 ***(Example 17 continued.)*** *Use of the* JMP *statistical package with the data of Example 17 produces*

$$y = .0048702 + .0749895x \text{ with } s_{\text{LF}} = .007855 \,.$$

We might expect a local (y) repeatability standard deviation of around .008 *(in the y absorbance units). In fact, 95% confidence limits for σ can be made (using $n - 2 = 4$ degrees of freedom and formula (2.45)) as*

$$.007855\sqrt{\frac{4}{11.143}} \text{ and } .007855\sqrt{\frac{4}{.484}} \,,$$

i.e.

$$.0047 \text{ and } .0226 \,.$$

Making use of the slope and intercept of the least squares line, a conversion formula for going from y_{new} to x_{new} is (as in display (2.46))

$$\hat{x}_{new} = \frac{y_{new} - .0048702}{.0749895} \,,$$

So, for example, a future measured absorbance of $y_{new} = .20$ suggests a concentration of

$$\hat{x}_{new} = \frac{.20 - .0048702}{.0749895} = 2.60\,\text{mg/l} \,.$$

Finally, Figure 2.19 on page 80 is a modification of Figure 2.18 that illustrates how the plotted prediction limits (2.43) provide both 95% predictions for a new measurement on a fixed/known measurand and 95% confidence limits on a new measurand,

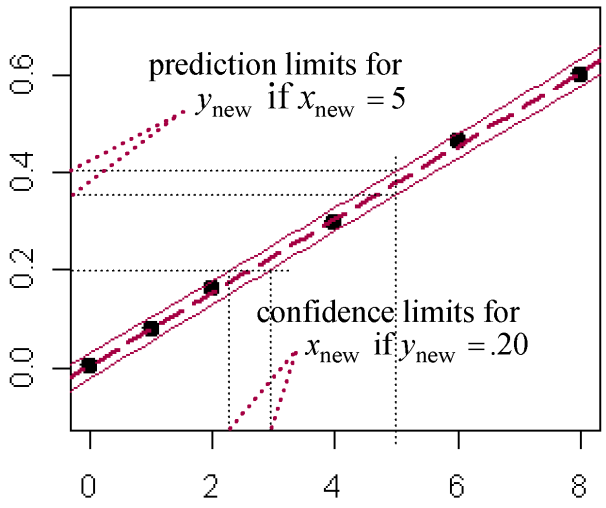


FIGURE 2.19. Confidence limits for x_{new} based on an observed y_{new} (and prediction limits (2.43))

having observed a particular measurement. Reading from the figure, one is "95% sure" that a future observed absorbance of .20 *comes from a concentration between*

$$2.28\,\text{mg/l} \quad \textit{and} \quad 2.903\,\text{mg/l}\,.$$

Example 19 ***A Check on Device "Linearity."*** *A calibration data set due to John Mandel compared* $n = 14$ *measured values* y *for a single laboratory to corresponding consensus values* x *for the same specimens derived from multiple labs. (The units are not available, but were the same for* x *and* y *values.) A simple linear regression analysis of the data pairs produced*

$$b_1 = .882 \quad \textit{and} \quad \frac{s_{\text{LF}}}{\sqrt{\sum (x_i - \bar{x})^2}} = .012$$

so that (using the upper 2.5% point of the t_{12} *distribution,* 2.179*, and formula (2.47)) 95% confidence limits for* β_1 *are*

$$.882 \pm 2.179\,(.012)$$

or

$$.882 \pm .026\,.$$

A 95% confidence interval for β_1 clearly does not include 1.0. So bias for the single laboratory was not constant. (The measurement "device" was not linear in the sense discussed on page 36.)

Section 2.5 Exercises

1. $n = 14$ polymer specimens of *known* weights, x, were weighed and the measured weights, y, recorded. The following table contains the data. (All units are g's.)

x	1	1	3	3	5	5	7
y	1.10	.95	2.98	3.01	5.02	4.99	6.97
x	7	10	10	12	12	14	14
y	7.10	10.03	9.99	12.00	11.98	14.10	14.00

   (a) Find the least squares line $\hat{y} = b_0 + b_1 x$ for these data.

   (b) Find the estimated repeatability standard deviation corresponding to your regression analysis.

   (c) Find 95% confidence limits for the y repeatability standard deviation based on your answer to (b).

2. In the context of problem 1, suppose a new specimen is measured as having a weight of 6.10 g.

   (a) Find the "calibrated weight," $\hat{x}$, corresponding to this new specimen based on your regression analysis.

   (b) Find 95% confidence limits for the slope of the relationship between measured and actual weight (β_1). Does the device used to produce the y measurements have constant bias (is it "linear")? Why or why not?

3. Based on your regression analysis in problem 1, find 95% prediction limits for the next measured weight for a new specimen with standard known weight of 8 g.

4. Would it be wise to use the above regression analyses to adjust a measured specimen weight of $y_{\text{new}} = .2$ g? Why or why not?

2.6 Measurement Precision and the Ability to Detect a Change or Difference

Estimating the repeatability and reproducibility of a gauge or measurement system amounts to finding a sensible standard deviation ("sigma") to associate with it. Where only one operator is involved, the repeatability measure, σ, is appropriate. Where multiple operators are going to be used, the measure $\sigma_{\mathrm{R\&R}}$ defined in display (2.27) is more germane. After an appropriate sigma has been identified, it can form the basis of judgments concerning the adequacy of the gauge or system for specific purposes. For example, looking at measures like the gauge capability ratio (2.29) is a way of judging (based on an appropriate sigma) the adequacy of a gauge for the purpose of checking conformance to a set of engineering specifications. In this section, we consider the slightly different matter of the adequacy of a gauge or measurement system for the purpose of detecting a change or difference.

The problem of determining whether "something has changed"/"there is a difference" is fundamental to engineering and technology. For example, in the context of process monitoring, engineers need to know whether process parameters (e.g., the mean widget diameter being produced by a particular lathe) are at standard values or have changed. And, in evaluating whether two machines are producing similar output, one needs to assess whether product characteristics from the two machines are the same or are consistently different. And, for example, when using hazardous materials in manufacturing, engineers need to compare chemical analyses for current environmental samples to analyses for "blank" samples, looking for evidence that important quantities of toxic materials have escaped a production process and thus increased their ambient level from some "background" level.

For the remainder of this section, let $\sigma_{\mathrm{measurement}}$ stand for an appropriate standard deviation for describing the precision of some measurement system. (Depending upon the context this could be σ or $\sigma_{\mathrm{R\&R}}$ from a gauge R&R study.) We will investigate the impact of $\sigma_{\mathrm{measurement}}$ on one's ability to detect change or difference through consideration of the distribution of the statistic

$$\overline{y}_{\mathrm{new}} - \overline{y}_{\mathrm{old}} \,, \tag{2.48}$$

where $\overline{y}_{\mathrm{new}}$ is the sample mean of n_{new} independent measurements taken on a particular "new" object and $\overline{y}_{\mathrm{old}}$ is the sample mean of n_{old} independent measurements taken on a particular "old" object. This is much like the context of Section 2.2.2, and Figure 2.20 (which is a slight modification of Figure 2.12) illustrates the situation under consideration. The related issue issue of how to think about situations where particular *objects* being measured are themselves of interest only to the extent that they represent "new" and "old" *conditions* that produced them will be considered briefly at the end of this section.

This discussion will allow the possibility that the information on the "old" object is

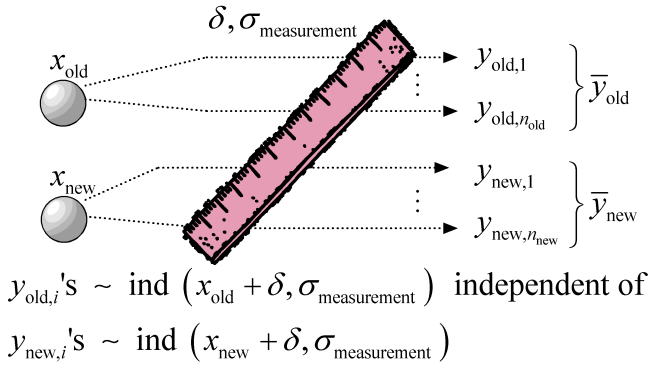


FIGURE 2.20. n_{old} measurements on an "old" measurand and n_{new} measurements on a "new" measurand made with a single system with (constant) bias δ and measurement standard deviation $\sigma_{\text{measurement}}$

strong enough that n_{old} can be thought of as being essentially infinite, and thus $\overline{y}_{\text{old}}$ essentially equal to the (long-run) mean of old observations, $x_{\text{old}} + \delta$. If $x_{\text{new}} + \delta$ is the long-run mean (expected value) of the new observations and the new and old measurements are independent, the random variable in display (2.48) has mean

Mean for the Random Variable $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$

$$\text{E}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) = (x_{\text{new}} + \delta) - (x_{\text{old}} + \delta) = x_{\text{new}} - x_{\text{old}} \ , \tag{2.49}$$

and variance

Variance for the Random Variable $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$

$$\text{Var}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) = \sigma^2_{\text{measurement}} \left(\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}} \right) \ . \tag{2.50}$$

When the information on the old object is strong enough to consider n_{old} to be essentially infinite, expression (2.50) reduces to

$$\text{Var}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) = \sigma^2_{\text{measurement}} \left(\frac{1}{n_{\text{new}}} \right) \ . \tag{2.51}$$

In the event that it is sensible to think of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ as normally distributed (either because repeat measurements are themselves approximately normally distributed or because both n_{new} and n_{old} are large) one then has the picture of the distribution of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ given in Figure 2.21 on page 84. If $x_{\text{new}} = x_{\text{old}}$, then the normal curve in the figure is centered at 0. If $x_{\text{new}} \neq x_{\text{old}}$ the normal curve in the figure is centered at the nonzero difference between the new and old measurands.

Example 20 ***Chemical Analysis for Benzene.*** *An appropriate standard deviation for characterizing an industrial laboratory's precision in a particular analysis for the benzene content of samples of a particular type is* $\sigma_{\text{measurement}} = .03\mu\,\text{g}/\,\text{l}$. *Suppose that in order to determine whether the amount of benzene in a particular environmental sample exceeds that in a particular similar "blank" sample (supposedly containing*

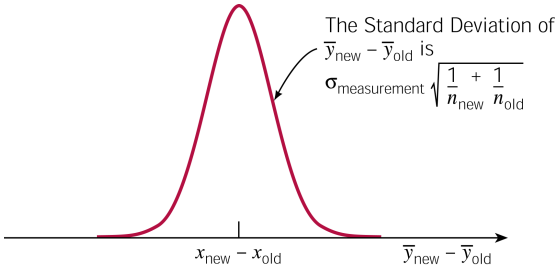


FIGURE 2.21. The distribution of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$

only background levels of the substance), the environmental sample will be analyzed $n_{\text{new}} = 1$ *time and its measured content compared to the mean measured content from* $n_{\text{old}} = 5$ *analyses of a single blank sample. Then from equations (2.49) and (2.50) the random variable* $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ *has mean*

$$\mathrm{E}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) = x_{\text{new}} - x_{\text{old}} \ ,$$

and standard deviation

$$\sqrt{\mathrm{Var}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}})} = \sigma_{\text{measurement}}\sqrt{\frac{1}{1} + \frac{1}{5}} = .03\sqrt{\frac{6}{5}} = .033\mu\,\mathrm{g/l}\ .$$

Figure 2.22 shows a corresponding probability distribution for the difference $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ *assuming that repeat measurements of a given sample are normally distributed.*

Note, by the way, that if the information on the blank sample was essentially perfect and equation (2.51) was relevant, the standard deviation associated with $\overline{y}_{\text{new}} - \mu_{\text{old}}$ *would be* $.030\mu\mathrm{g/l}$*, not much smaller than the* $.033\mu\mathrm{g/l}$ *value found here.*

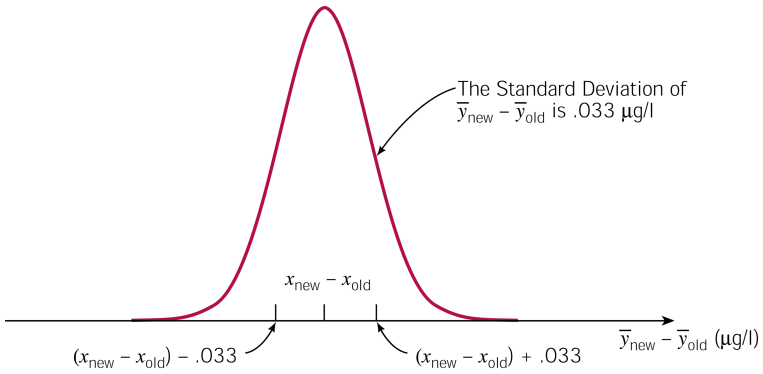


FIGURE 2.22. Distribution of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ in the benzene analysis example

The picture of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ given in equations (2.49) and (2.50) and Figure 2.21 forms the basis for several common ways of evaluating the adequacy of a measurement technique to characterize a change or difference. One simple-minded rule of thumb often employed by analytical chemists is that a difference $x_{\text{new}} - x_{\text{old}}$ needs to be on the order of 10 times the standard deviation of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ before it can be adequately characterized by a measurement process with standard deviation $\sigma_{\text{measurement}}$ using sample sizes n_{new} and n_{old}. For example, in the benzene analyses of Example 20, with sample sizes $n_{\text{new}} = 1$ and $n_{\text{old}} = 5$, this rule of thumb says that only increases in real benzene content on the order of at least

$$10 \times .033 = .33 \mu \,\text{g}/\,\text{l}$$

can be reliably characterized. This somewhat ad hoc (but nevertheless popular) guideline amounts to a requirement that one's "signal-to-noise ratio" for determination of a difference (ratio of mean to standard deviation) be at least 10 before being comfortable with the resulting precision.

A second approach to using the picture of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ given in equations (2.49) and (2.50) and Figure 2.21 to describe one's ability to detect a difference between measurands involves some ideas from hypothesis testing. In interpreting an observed value of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$, one might require that it be of a certain minimum magnitude before declaring that there is clearly a difference between new and old objects. For the sake of concreteness, suppose for the rest of this section that one is concerned about detecting an increase in response. That is, suppose one is interested in detecting the possibility that $x_{\text{new}} - x_{\text{old}} > 0$. It then makes sense to set some **critical limit, L_{c}**, and to only declare that there has been a change (that there is a difference) if

$$\overline{y}_{\text{new}} - \overline{y}_{\text{old}} > L_{\text{c}} \,. \tag{2.52}$$

If one wishes to limit the probability of a "false positive" (i.e., a type I error) L_{c} should be large enough that the eventuality (2.52) occurs rarely when in fact $x_{\text{new}} = x_{\text{old}}$. For example, if one may assume that $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ is essentially normally distributed, it is possible to use the fact that when $x_{\text{new}} = x_{\text{old}}$ the variable

$$\frac{\overline{y}_{\text{new}} - \overline{y}_{\text{old}}}{\sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}}$$

is standard normal, to set a value for L_{c}. From a normal table, one may pick a number z_1 so that for standard normal Z, $P[Z > z_1] = \alpha$, for α a small number of one's choosing. Then setting

Critical Value for Normal $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ if an Increase in Response Is to Be Detected

$$L_{\text{c}} = z_1 \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}\,, \tag{2.53}$$

the probability of a false positive is no more than α.

Once one has established a critical value L_{c} (using equation (2.53) or otherwise) it is then reasonable to ask what is the probability of detecting a change of a given size, or equivalently what size change can be reliably detected. Again, assuming that $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ is essentially normally distributed as in Figure 2.21, it is possible to answer this question. That is, using the normal model, with

z-score for the Critical Value L_{c}

$$z_2 = \frac{L_{\text{c}} - (x_{\text{new}} - x_{\text{old}})}{\sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}}\,, \tag{2.54}$$

the probability (depending upon $x_{\text{new}} - x_{\text{old}}$) of declaring that there has been a change (that there is a difference) is

Probability of Declaring That Response Has Increased (for Normal $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$)

$$\gamma = P[Z > z_2]\,, \tag{2.55}$$

(for Z again standard normal). Or by rewriting equation (2.54), for z_2 chosen so that γ in display (2.55) is large, one can solve for the size of change in measurand required to produce a large (at least γ) probability of detection, namely

Change in Measurand with γ Probability of Detection (for Normal $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$)

$$x_{\text{new}} - x_{\text{old}} = L_{\text{c}} - z_2 \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}\,. \tag{2.56}$$

Notice that in display (2.56) z_2 is typically negative so that this difference in measurands $x_{\text{new}} - x_{\text{old}}$ is then typically larger than L_{c}.

In analytical chemistry, the value of $x_{\text{new}} - x_{\text{old}}$ required to produce a large probability of detecting an increase in measurand is given a special name.

Definition 21 *For a given standard deviation of measurement $\sigma_{\text{measurement}}$, sample sizes n_{new} and n_{old}, critical value L_{c} and desired (large) probability γ, the* ***lower limit of detection, L_{d},*** *of a measurement protocol is the smallest difference in measurands $x_{\text{new}} - x_{\text{old}}$ producing*

$$P[\overline{y}_{\text{new}} - \overline{y}_{\text{old}} > L_{\text{c}}] \geq \gamma.$$

When $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ is normal and z_2 is chosen according to display (2.55), equation (2.56) implies that

Lower Limit of Detection (for Normal $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$)

$$L_{\text{d}} = L_{\text{c}} - z_2 \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}\,. \tag{2.57}$$

(where again, z_2 is typically negative so that this L_{d} is then typically larger than L_{c}). And although there is no requirement that L_{c} be set according to equation (2.53), when this is used to control the chance of a false positive, equations (2.53) and (2.57) combine to produce

Lower Limit of Detection if Equation (2.53) is Used to Set the Critical Value and Equation (2.55) is Used

$$L_{\text{d}} = (z_1 - z_2)\, \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}} \tag{2.58}$$

Example 22 ***(Example 20 continued.)*** *Consider again the benzene analysis example. Suppose that it is desirable to limit the probability of producing a "false positive" (a declaration that the environmental sample contains more benzene than the blank sample when in fact there is no real difference in the two) to no more than $\alpha = .10$. Consulting a standard normal table, for Z standard normal, $P[Z > 1.282] = .10$. So, using equation (2.53), an appropriate critical value is*

$$L_{\text{c}} = 1.282(.030)\sqrt{\frac{1}{1} + \frac{1}{5}} = .042.$$

Should one wish to evaluate the probability of detecting an increase in real benzene content of size $x_{\text{new}} - x_{\text{old}} = .02\mu\,\text{g}/\text{l}$ using this critical value, equations (2.54) and (2.55) show that with

$$z_2 = \frac{.042 - .02}{.030\sqrt{\frac{1}{1} + \frac{1}{5}}} = .67,$$

the probability is only about

$$P[Z > z_2] = P[Z > .67] = .2514.$$

There is a substantial (75%) chance of failing to identify a .02μ g/ l increase in benzene content beyond that resident in the blank sample. This unpleasant fact motivates the question "How big does the increase in benzene concentration need to be in order to have a large (say 95%) chance of seeing it above the measurement noise?" Since consultation of the standard normal table shows that

$$P[Z > -1.645] = .95,$$

equations (2.55) and (2.57) imply that for $\gamma = .95$

$$L_{\mathrm{d}} = x_{\mathrm{new}} - x_{\mathrm{old}} = .042 - (-1.645)(.030)\sqrt{\frac{1}{1} + \frac{1}{5}} = .096\mu\,\mathrm{g/\,l}$$

is (in the language of Definition 21) the lower limit of detection in this situation. A real increase in benzene content must be of at least this size for there to be a large (95%) chance of "seeing" it through the measurement noise.

The particular example used in the foregoing discussion of the implications of equations (2.49) and (2.50) and Figure 2.21 is from analytical chemistry. But the basic method illustrated is perfectly general and could, for example, be equally well applied to the consideration of the implications of measurement precision for one's ability to detect a difference in diameters of two particular parts turned on a lathe.

It is, however, an extremely important distinction that the discussion thus far in this section has been phrased in terms of detecting a difference between two particular objects and *not* between processes or populations standing behind those objects. Example 20 concerns comparison of a particular environmental sample and a particular blank sample. It does *not* directly address the issue of how the population of environmental samples from a site of interest compares to a population of blanks. Similarly, in a manufacturing context, comparisons based on the foregoing material would concern two particular measured parts, not the process conditions operative when those parts were made. The point here is that only measurement variation has been taken into account, and not object-to-object variation for processes or populations the measured objects might represent. (And unless there is no "within-population" or "process" variation, detection of a difference between an old and a new object is not the same as detection of a difference between old and new population or process means.)

The problem of comparing two processes or populations while admitting the reality of measurement noise was first raised in Section 2.2.2. Figure 2.23 is a slight modification of Figure 2.13 from that section and represents the scenario under consideration.

In this second context, the difference in sample means (2.48) has mean

$$\mathrm{E}(\overline{y}_{\mathrm{new}} - \overline{y}_{\mathrm{old}}) = (\mu_{x\mathrm{new}} + \delta) - (\mu_{x\mathrm{old}} + \delta) = \mu_{x\mathrm{new}} - \mu_{x\mathrm{old}}\ ,$$

the difference in process or condition mean measurands. The variance of the difference

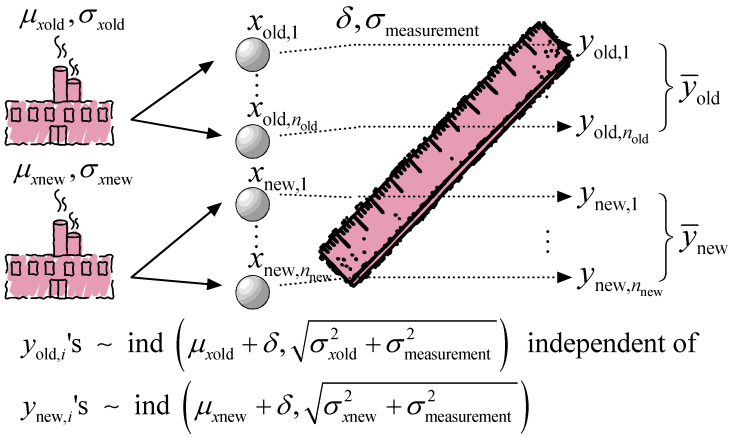


FIGURE 2.23. n_{old} measurements on an "old" process and n_{new} measurements on a "new" process made with a single system with (constant) bias δ and measurement standard deviation $\sigma_{\text{measurement}}$

is

$$\begin{aligned}\text{Var}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) &= \frac{\sigma^2_{x\text{new}} + \sigma^2_{\text{measurement}}}{n_{\text{new}}} + \frac{\sigma^2_{x\text{old}} + \sigma^2_{\text{measurement}}}{n_{\text{old}}} \\ &= \sigma^2_{\text{measurement}}\left(\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}\right) + \frac{\sigma^2_{x\text{new}}}{n_{\text{new}}} + \frac{\sigma^2_{x\text{old}}}{n_{\text{old}}}\end{aligned}$$

and (of course) item-to-item/measurand-to-measurand variability is reflected in $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$. In the event that the "old" and "new" processes have comparable values of σ_x, the formulas and language in this section can be reinterpreted to allow application to the problem of detecting changes in a process or population mean, by replacing $\sigma_{\text{measurement}}$ with

$$\sqrt{\sigma_x^2 + \sigma^2_{\text{measurement}}}$$

Otherwise, a separate development of formulas is required. For example, if in the context of Example 20, blank samples are more homogeneous than are field samples from a particular site, an analysis parallel to that in this section but based on the two different values for σ_x will be needed for application to the problem of comparing a site mean benzene level to a blank mean level.

Section 2.6 Exercises

1. An analyst can tolerate only a .01 (1%) chance of falsely concluding that a measurand associated with a field sample exceeds that for a blank (should those actually be the same).

(a) In the context of the discussion of this section, what value of z_1 is then appropriate?

(b) The analyst was also interested in identifying the smallest positive difference in measurands ($x_{\text{field}} - x_{\text{blank}}$) for which there is probability of only .05 of incorrectly concluding the difference is zero or negative. What value should this person use for z_2?

(c) Using your values for z_1 and z_2 from (a) and (b) above and equation (2.58), what is the lower limit of detection for this scenario if only single measurements are made on both the field sample and the blank? (Your answer will be a multiple of $\sigma_{\text{measurement}}$.)

2. **Dimethyl Phenanthrene (DMP4) Atmospheric Blank.** Currie, et al. presented the paper "Impact of the Chemical and Isotopic Blank on the Interpretation of Environmental Radiocarbon Results" at the International Radiocarbon Conference in Glasgow, Scotland, August 1994. Their presentation included discussion of six measurements of the DMP4 content of an atmospheric field filter blank. (DMP4 is a product of softwood pyrolysis.) The six responses in nanograms (10^{-9} g) were

$$8.25,\ 7.30,\ 7.27,\ 6.54,\ 6.75 \text{ and } 7.32.$$

Suppose here that one analyst measured the DMP4 field filter blank all six times with the same instrument.

(a) Find the sample standard deviation of these values. In terms of the gauge R&R material of Section 2.4, what does this sample standard deviation estimate? In parts (b), (c) and (d) below, use this estimate as if it were the true value of $\sigma_{\text{measurement}}$.

(b) Find a critical limit, L_{c}, if the probability of a false positive detection is to be 10% and both a field sample and a (new) blank are to be measured once.

(c) Find a lower limit of detection where the probability of a false positive detection is 10% and the probability of a false negative is to be 5% for a difference in measurands at the lower limit of detection. (Suppose as in part (b) that both a field sample and a new blank are to be measured once.)

(d) Continuing under the assumption that a single individual made all six measurements, what must be true about the measurement method if the sample standard deviation in (a) is reasonably thought of as estimate of "$\sigma_{\text{measurement}}$"? (Consider the ideas of Section 2.4.)

2.7 R&R Considerations for Go/No-Go Inspection

Ideally, observation of a process results in quantitative measurements. But there are some contexts in which all that is determined is whether an item or process condition is of one of two types, that we will for the present call "conforming" and "non-conforming." It is, for example, common to check the conformance of machined metal parts to some engineering requirements via the use of a "go/no-go gauge." (A part is conforming if a critical dimension fits into the larger of two check fixtures and does not fit into the smaller of the two.) And it is common to task human beings with making visual inspections of manufactured items and producing a "OK/Not-OK" call on each.

Engineers are sometimes then called upon to apply the qualitative "repeatability" and "reproducibility" concepts of metrology to such Go/No-Go or "0/1" contexts. One wants to separate some measure of overall inconsistency in 0/1 "calls" on items into pieces that can be mentally charged to inherent inconsistency in the equipment or method, and the remainder that can be charged to differences between how operators use it. Exactly how to do this is presently not well-established. The best available statistical methodology for this kind of problem is more complicated than can be presented here (involving so-called "generalized linear models" and random effects in these). What we *can* present is a rational way of making point estimates of what might be termed repeatability and reproducibility components of variation in 0/1 calls. (These are based on reasoning similar to that employed in Section 2.4.2 to find correct range-based estimates in usual measurement R&R contexts.) We then remind the reader of elementary methods of estimating differences in population proportions and point out their relevance in the present situation.

2.7.1 Some Simple Probability Modeling

To begin, think of coding a "non-conforming" call as "1" and a "conforming" call as "0," and having J operators each make m calls on a fixed part. Suppose that J operators have individual probabilities $p_1, p_2, \ldots, p_J$ of calling the part "non-conforming" on any single viewing, and that across m viewings

$$X_j = \text{the number of non-conforming calls among the } m \text{ made by operator } j$$

is Binomial (m, p_j). We'll assume that the p_j are random draws from some population with mean π and variance v**.**

The quantity

$$p_j\,(1-p_j)$$

is a kind of "per call variance" associated with the declarations of operator j, and might serve as a kind of repeatability variance for that operator. (Given the value of p_j, elementary probability theory says that the variance of X_j is $mp_j\,(1-p_j)$.) The biggest problem here is that unlike what is true in the usual case of gauge R&R for

measurements, this variance is not constant across operators. But its expected value, namely

$$\begin{aligned} \mathrm{E}\left(p_j\left(1-p_j\right)\right) &= \pi - \mathrm{E}p_j^2 \\ &= \pi - \left(v+\pi^2\right) \\ &= \pi\left(1-\pi\right) - v \end{aligned}$$

can be used as a sensible measure of variability in conforming/non-conforming classifications chargeable to repeatability sources. The variance v serves as a measure of reproducibility variance. This ultimately points to

$$\pi\left(1-\pi\right)$$

as the "total R&R variance" here. That is, we make definitions for 0/1 contexts

R&R Variance for One Part in a 0/1 Context

$$\sigma^2_{\text{R\&R}} = \pi\left(1-\pi\right) \tag{2.59}$$

Repeatability Variance for One Part in a 0/1 Context

$$\sigma^2_{\text{repeatability}} = \pi\left(1-\pi\right) - v \tag{2.60}$$

and

Reproducibility Variance for One Part in a 0/1 Context

$$\sigma^2_{\text{reproducibility}} = v \tag{2.61}$$

2.7.2 Simple R&R Point Estimates for 0/1 Contexts

Still thinking of a single fixed part, let

$$\hat{p}_j = \frac{\text{the number of "non-conforming" calls made by operator } j}{m} = \frac{X_j}{m}$$

and define the (sample) average and (sample) variance of these,

$$\bar{\hat{p}} = \frac{1}{J}\sum_{j=1}^{J}\hat{p}_j \text{ and } s^2_{\hat{p}} = \frac{1}{J-1}\sum_{j=1}^{J}\left(\hat{p}_j - \bar{\hat{p}}\right)^2 .$$

It is possible to argue that

$$\mathrm{E}\bar{\hat{p}} = \pi$$

and that

$$\begin{aligned} \mathrm{E}s^2_{\hat{p}} &= \mathrm{Var}\hat{p}_j \\ &= \mathrm{Var}p_j + \mathrm{E}\frac{p_j\left(1-p_j\right)}{m} \\ &= \frac{m-1}{m}v + \frac{\pi\left(1-\pi\right)}{m} , \end{aligned}$$

so that

$$v = \frac{m}{m-1}\mathrm{E}s^2_{\hat{p}} - \frac{\pi\left(1-\pi\right)}{m-1} .$$

This line of reasoning suggests the simple estimators (still based on a single part)

Estimator of R&R Variance for a Single Part in a 0/1 Context

$$\hat{\sigma}^2_{\text{R\&R}} = \bar{\hat{p}}\left(1-\bar{\hat{p}}\right) \tag{2.62}$$

and

$$\hat{\sigma}^2_{\text{reproducibility}} = \max\left(0, \frac{1}{m-1}\left(m s^2_{\hat{p}} - \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)\right)\right) \tag{2.63}$$

On rare occasions, $s^2_{\hat{p}}$ will exceed $\bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$, leading to a value of $\hat{\sigma}^2_{\text{reproducibility}}$ in display (2.63) larger than $\hat{\sigma}^2_{\text{R\&R}}$ specified in display (2.62). In those cases we will reduce $\hat{\sigma}^2_{\text{reproducibility}}$ to $\hat{\sigma}^2_{\text{R\&R}} = \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$ and thus produce a final version of the estimator (2.63)

Estimator of Reproducibility Variance for a Single Part in a 0/1 Context

$$\hat{\sigma}^2_{\text{reproducibility}} = \min\left(\hat{\sigma}^2_{\text{R\&R}}, \max\left(0, \frac{1}{m-1}\left(m s^2_{\hat{p}} - \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)\right)\right)\right) . \tag{2.64}$$

Then from displays (2.62) and (2.64) an obvious estimator of the repeatability variance is

Estimator of Repeatability Variance for a Single Part in a 0/1 Context

$$\hat{\sigma}^2_{\text{repeatability}} = \hat{\sigma}^2_{\text{R\&R}} - \hat{\sigma}^2_{\text{reproducibility}} \tag{2.65}$$

Again, the estimators (2.62), (2.64), and (2.65) are based on a single part. Exactly what to do based on multiple parts (say I of them) is not completely obvious. But in order to produce a simple methodology, we will simply average estimates made one part at a time across multiple parts, presuming that parts in hand are sensibly thought of as a random sample of parts to be checked, and that this averaging is a reasonable way to combine information across parts.

In order for any of this to have a chance of working, m will need to be fairly large. The usual gauge R&R "$m = 2$ or 3" just isn't going to produce informative results in the present context. And in order for this to work in practice (so that an operator isn't just repeatedly looking at the same few parts over and over and remembering how he or she has called them in the past) a large value of I may also be needed.

TABLE 2.7. Hypothetical Results of Visual Inspection of 5 Parts by 3 Operators

	Operator 1	Operator 2	Operator 3	$\bar{\hat{p}}$	$\bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$	$s^2_{\hat{p}}$
Part 1	.2	.4	.2	.2667	.1956	.0133
Part 2	.6	.6	.7	.6333	.2322	.0033
Part 3	1.0	.8	.7	.8333	.1389	.0233
Part 4	.1	.1	.1	.1	.0900	0
Part 5	.1	.3	.3	.2333	.1789	.0133

TABLE 2.8. R&R Cacluations for the Hypothetical Visual Inspection Data

	$\hat{\sigma}^2_{\text{R\&R}} = \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$	$\hat{\sigma}^2_{\text{reproducibility}}$	$\hat{\sigma}^2_{\text{repeatability}}$
Part 1	.1956	0	.1956
Part 2	.2322	0	.2322
Part 3	.1389	.0105	.1284
Part 4	.0900	0	.0900
Part 5	.1789	0	.1789
Average	.1671	.0021	.1650

Example 23 ***A Simple Numerical Example.*** *For purposes of illustrating the formulas of this section, we will use a small numerical example due to Prof. Max Morris. Suppose that $I = 5$ parts are inspected by $J = 3$ operators, $m = 10$ times apiece, and that in Table 2.7 are sample fractions of "non-conforming" calls made by the operators and a few summary statistics.*

The entries in the next to last column of Table 2.7 are $\hat{\sigma}^2_{R\&R}$ values for the 5 parts. Estimates of $\hat{\sigma}^2_{reproducibility}$ are, for example, computed as for Part 1

$$\begin{aligned}\hat{\sigma}^2_{reproducibility} &= \max\left(0, \frac{1}{10-1}\left(10\left(.0133\right) - .1956\right)\right)\\ &= 0\end{aligned}$$

leaving estimates of $\hat{\sigma}^2_{repeatability}$ computed as for Part 1

$$\begin{aligned}\hat{\sigma}^2_{repeatability} &= .1956 - 0\\ &= .1956\end{aligned}$$

The whole set of estimates and their averages are collected in Table 2.8.

Then, for example, a fraction of only

$$\frac{.0021}{.1671} = 1.3\%$$

of the inconsistency in conforming/non-conforming calls seen in the original data seems to be attributable to clear differences in how the operators judge the parts (differences in the binomial "success probabilities" p_j). Rather, the bulk of the variance seems to be attributable to unavoidable binomial variation. The p's are not close enough to either 0 or 1 to make the calls tend to be consistent. So the variation seen in the $\hat{p}$'s in a given row is not clear evidence of large operator differences.

Of course, we need to remember that the computations above are on the variance (and not standard deviation) scale. On the (more natural) standard deviation scale, reproducibility variation

$$\sqrt{.0021} = .05$$

and repeatability variation

$$\sqrt{.1650} = .41$$

are not quite so strikingly dissimilar.

2.7.3 Application of Inference Methods for the Difference in Two Binomial "p's"

The question of whether call rates for two operators on the same part are really detectably different brings up the elementary statistics topic of estimating the difference in two binomial parameters, say p_1 and p_2. A common elementary large sample approximate confidence interval for $p_1 - p_2$ has endpoints

$$\hat{p}_1 - \hat{p}_2 \pm z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\,.$$

But, as it turns out, this formula can fail badly if either p is extreme or either n is small. So we will use a slight modification that is more reliable, namely

Confidence Limts for $p_1 - p_2$

$$\hat{p}_1 - \hat{p}_2 \pm z\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}} \tag{2.66}$$

where

Values to Use in Formula (2.66)

$$\tilde{p}_i = \frac{n_i\hat{p}_i + 2}{n_i + 4} \tag{2.67}$$

That is, under the square root of the usual formula one essentially replaces the $\hat{p}$ values with $\tilde{p}$ values derived by adding 2 "successes" in 4 "additional trials" to the counts used to make up the $\hat{p}$ values. (This has the effect of making the standard large sample interval a bit wider and correcting the problem that for small sample sizes and extreme values of p it can fail to hold its nominal confidence level.)

Example 24 ***(Example 23 continued.)*** *Consider again Part 1 from Example 23, and in particular consider the question of whether Operator 1 and Operator 2 have clearly different probabilities of calling that part non-conforming on a single call. With* $\hat{p}_1 = .2$ *and* $\hat{p}_2 = .4$*, formula (2.67) says that*

$$\tilde{p}_1 = \frac{2+2}{10+4} = .2857 \text{ and } \tilde{p}_2 = \frac{4+2}{10+4} = .4286$$

so that using formula (2.66) approximate 95% confidence limits for the difference $p_1 - p_2$ *are*

$$.2 - .4 \pm 1.96\sqrt{\frac{.2857\,(1-.2857)}{10} + \frac{.4286\,(1-.4286)}{10}}$$

i.e.

$$-.2 \pm .49$$

These limits cover 0 *and there thus is no clear evidence in the* $\hat{p}_1 = .2$ *and* $\hat{p}_2 = .4$ *values (from the relatively small samples of sizes* $m = 10$*) that Operators 1 and 2 have different probabilities of calling Part 1 non-conforming.*

Section 2.7 Exercises

1. Suppose that 10 parts are inspected by 4 operators 16 times apiece. Each inspection determines whether or not the item is conforming. The counts in the table below correspond to the numbers of "non-conforming" calls out of 16 inspections.

	Operator 1	**Operator 2**	**Operator 3**	**Operator 4**
Part 1	10	11	11	10
Part 2	11	9	12	10
Part 3	8	8	9	7
Part 4	15	14	14	16
Part 5	12	14	11	12
Part 6	15	15	16	15
Part 7	14	11	14	12
Part 8	16	16	15	15
Part 9	13	15	14	15
Part 10	16	15	16	16

   (a) Using the data above, fill in the table below:

Part	$\bar{\hat{p}}$	$\bar{\hat{p}}(1-\bar{\hat{p}})$	$s^2_{\hat{p}}$
1			
2			
$\vdots$			
10			

(b) What is the fraction of inconsistency in conforming/non-conforming calls that can be attributed to clear differences in how the operators judged the parts (differences in the binomial "success probabilities" p_j)? (Make your answer on the variance scale.)

(c) What is the estimated reproducibility variation (on the standard deviation scale)?

(d) What is the estimated repeatability variation (on the standard deviation scale)?

(e) For part 10, give a 90% confidence interval for the difference (operator 1 minus operator 3) in probabilities of a non-conforming call. Does it appear the operators 1 and 3 have different probabilities of a non-conforming call on *any* one of the parts? Why?

2.8 Chapter Summary

This chapter has been concerned with how measurement error impacts what can be learned from empirical data. It presented some ideas from the probability modeling of measurement variation, and considered how the interpretation of elementary statistical inferences is affected by measurement error. Then a variety of more advanced statistical tools were discussed, because of their usefulness in quantifying, partitioning, and (in some cases) removing the effects of measurement variation in quality assurance and improvement projects.

2.9 Chapter 2 Exercises

1. Does a perfectly calibrated device return measurements of a measurand that are completely free of error? Explain.

2. Is a standard (an item with corresponding "known" measurand) needed in *both* device calibration and estimation of σ_{device}? If not, which requires a standard? Explain.

3. A measurement device may have a bias as large as 1 unit (in absolute value) and a device standard deviation as large as 1 unit. You measure x and observe $y = 10$. If you believe in the simple (normal) measurement model and want to report an interval you are "at least 99% sure" contains x, you should report what limits? (Hint: Before measurement, how far do you expect y to be from x with the indicated worst possible values of absolute bias and standard deviation? Interpret "99% sure" in "plus or minus 3 standard deviations" terms.)

4. The same axel diameter is measured $n_1 = 25$ times with device #1 and $n_2 = 25$ times with device #2, with resulting means and standard deviations $\overline{y}_1 = 2.001\,\text{in}, \overline{y}_2 = 2.004\,\text{in}, s_1 = .003\,\text{in}$, and $s_2 = .004\,\text{in}$. The upper 2.5% point of the $F_{24,24}$ distribution is about 2.27.

   (a) Give 95% confidence limits for the difference in device biases.

   (b) Give 95% confidence limits for the ratio of the two device standard deviations.

   (c) Is there a clear difference in device biases based on your interval in (a)? Why or why not?

   (d) Is there a clear difference in device standard deviations based on your interval in (b)? Why or why not?

5. Two different (physically stable) production lines produce plastic pop bottles. Suppose $n_1 = 25$ bottles from line #1 and $n_2 = 25$ bottles from line #2 are burst-tested on a single tester, with resulting means and standard deviations $\overline{y}_1 = 201\text{psi}, \overline{y}_2 = 202\text{psi}, s_1 = 3\text{psi}$, and $s_2 = 4\text{psi}$.

   (a) Give a 95% confidence interval for the difference between the mean burst strengths for lines #1 and #2 (line #1 minus line #2).

   (b) Give a 95% confidence interval for the ratio of burst strength standard deviations (line #1 divided by line #2). The upper 2.5% point of the $F_{24,24}$ distribution is about 2.27.

   (c) Is there a clear difference between mean burst strengths? Why or why not?

   (d) Is there a clear difference between the consistencies of burst strengths? Why or why not?

6. Using a single tester, a single metal specimen was tested for Brinell hardness 20 times with resulting sample standard deviation of hardness 10HB. Subsequently, 40 different specimens cut from the same ingot of steel have sample standard deviation of measured hardness 20HB (using the same tester).

   (a) Give 95% confidence limits for a "test variability" standard deviation.

   (b) Give approximate 95% confidence limits for a specimen-to-specimen standard deviation of actual Brinell hardness.

7. An ANOVA analysis of a gauge R&R data set produced $\hat{\sigma}_{\text{R\&R}} = 53$ (in appropriate units) and $\widehat{\nu}_{\text{R\&R}} = 3$. In these units, engineering specifications on a critical dimension of a machined steel part are $nominal \pm 200$. Give approximate 95% confidence limits for a GCR (gauge capability ratio) for checking conformance to these specifications.

8. 95% confidence limits for a particular gauge capability ratio are 6 to 8. What does this indicate about the usability of the gauge for checking conformance to the specifications under consideration?

9. Below is an analysis of variance table from a calibration study. The data were light intensities, y (in unspecified analyzer units) for specimens of *known* Riboflavin concentration x (in $\mu\text{g}/\text{ml}$).

ANOVA Table

Source	**SS**	**df**	**MS**
Model	10946.445	1	10946.445
Error	27.155	8	3.4
Total	10973.6	9	

Parameter estimates for the simple linear regression model were $b_0 = 6.4634$ and $b_1 = 129.1768$.

(a) Give a 95% confidence interval for a repeatability standard deviation for this analyzer.

(b) Suppose a new specimen with unknown concentration is analyzed and $y_{\text{new}} = 75$ is observed. Give a single number estimate of the concentration in that specimen.

10. The final step in the production of some glass vials is a visual inspection presently carried out by human inspectors. A particular single vial (marked in an "invisible" ink that can be seen only under ultraviolet light) known to be defective is repeatedly run through the inspection process among a large number of newly produced vials. In fact, each of 5 company inspectors sees that vial 10 times in a company study. Below are the rates at which that vial was identified as defective by the various operators ("1.0" means 100%.)

$$.6, .9, .9, 1.0, 1.0$$

(a) What two values of $\hat{p}$ reflect perfect consistency of "defective/non-defective" calls made by a particular inspector?

(b) What distribution models the number of correct "defective" calls made by a particular inspector?

(c) On the scale of (estimated) variances (not standard deviations), what is the fraction of overall variation seen in the "defective/non-defective" calls for this vial that should be attributed to operator-to-operator differences?

(d) Give 95% confidence limits for the long run difference in proportions of "defective" calls for the first operator (that made 6 out of 10 "defective"calls) and the last operator (who made all "defective"calls).

11. **Laser Metal Cutting.** Davis, Martin and Popinga used a Ytterbium Argon gas laser to make some cuts in 316 stainless steel. Using $95\,\text{MJ/pulse}$ and $20\,\text{Hz}$ settings on the laser and a $15.5\,\text{mm}$ distance to the steel specimens (set at a 45° angle to the laser beam) the students made cuts in specimens using 100, 500, and 1000 pulses. The measured depths of four different cuts (in machine

units) at each pulse level are given below (assume the same operator made all measurements and that repeatability variation is negligible here).

100 Pulses	**500 Pulses**	**1000 Pulses**
$7.4, 8.6, 5.6, 8.0$	$24.2, 29.5, 26.5, 23.8$	$33.4, 37.5, 35.9, 34.8$

(a) What is the response variable in this problem?

(b) Give the sample average values for the 100, 500, and 1000 pulse levels. Calculate the sample range for the data at each pulse level. Give estimates of the standard deviation of cut depth for each level of pulse, first based on the sample range and then using the sample standard deviation. (You will have two estimates for each of the three population standard deviations.)

(c) Assuming variability is the same for all three pulse levels, give an estimate of the common standard deviation based on the three sample ranges.

(d) The concepts of measurement validity, precision, and accuracy are discussed in Section 2.1. The analysts decided to report the average cut depth for the different pulse levels. This averaging can be thought of in terms of improving which of 1) validity, 2) precision, or 3) accuracy (over the use of any single measurement)? The concept of calibration is most closely associated with which of the three?

12. **Fiber Angle.** Grunig, Hamdorf, Herman, and Potthoff studied a carpet-like product. They measured the angle at which fibers were glued to a sheet of base material. A piece of finished product was obtained and cut into five sections. Each of the four team members measured the fiber angle eight times for each section. The results of their measuring are given below (in degrees above an undisclosed reference value). A corresponding ANOVA table is also given.

(a) Say what each term in the equation $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$ means in this problem (including the subscripts i, j, and k).

(b) Using ranges, estimate the repeatability and reproducibility standard deviations for angle measurement. Based on this analysis what aspect of the measuring procedure seems to need the most attention? Explain.

(c) Using ANOVA-based formulas, estimate the repeatability and reproducibility standard deviations for angle measurement. Is this analysis in essential agreement with that in part (b)? Explain.

(d) Using your answer to (c), give an estimate of the standard deviation that would be experienced by many analysts making a single measurement on the same angle (in the same section) assuming there is no repeatability component to the overall variation.

(e) Specifications on the fiber angle are $nominal \pm 5\,^{\circ}$. Estimate the gauge capability ratio using first ranges and then ANOVA-based estimates. Does it appear this measurement method is adequate to check conformance to the specifications? Why or why not?

Angle	Analyst 1	Analyst 2	Analyst 3	Analyst 4
1	19, 20, 20, 23	20, 25, 17, 22	20, 19, 15, 16	10, 10, 10, 5
	20, 20, 20, 15	23, 15, 23, 20	20, 19, 12, 14	5, 5, 5, 5
2	15, 17, 20, 20	15, 13, 5, 10	15, 20, 14, 16	10, 10, 10, 10
	10, 15, 15, 15	8, 8, 10, 12	13, 20, 15, 15	10, 15, 15, 10
3	23, 20, 22, 20	20, 23, 20, 20	15, 20, 22, 18	10, 10, 10, 15
	25, 22, 20, 23	23, 23, 22, 20	15, 20, 16, 20	15, 10, 10, 10
4	15, 16, 22, 15	20, 22, 18, 23	13, 13, 15, 20	5, 10, 10, 10
	15, 15, 22, 17	23, 23, 24, 20	11, 20, 13, 15	10, 10, 10, 10
5	20, 20, 22, 20	18, 20, 18, 23	10, 14, 17, 12	5, 10, 10, 10
	27, 17, 20, 15	20, 20, 18, 15	11, 10, 15, 10	10, 10, 10, 10

ANOVA Table

Source	SS	df	MS
Angle	390.913	4	97.728
Analyst	2217.15	3	739.05
Angle×Analyst	797.788	12	66.482
Error	971.75	140	6.941
Total	4377.6	159	

13. Refer to the **Fiber Angle** case in problem 12.

    (a) Is it preferable to have eight measurements on a given section by each analyst as opposed to, say, two measurements on a given section by each analyst? Why or why not?

    (b) For a given number of angle measurements per analyst×section combination, is it preferable to have 4 analysts instead of 2, 6, or 8? Why or why not?

    (c) When making angle measurements for a given section, does it matter if the angle at a fixed location on the piece is repeatedly measured, or is it acceptable (or even preferable?) for each analyst to measure at 8 different locations on the section? Discuss.

    (d) Continuing with (c), does it matter that the locations used on a given section varied analyst to analyst? Why or why not?

14. **Bolt Shanks.** A 1-inch micrometer is used by an aircraft engine manufacturer to measure the diameter of a body-bound bolt shank. Specifications on this dimension have been set with a spread of .002 in. Three operators and ten body-bound

bolt shanks were used in a gauge R&R study. Each bolt shank was measured twice by each operator (starting with part 1 and proceeding sequentially to part 10) to produce the data below (in inches). A corresponding ANOVA table is provided as well (SS's and MS's are in $10^{-6}\,\text{in}^2$.).

Part	Operator A	Operator B	Operator C
1	.3473	.3467	.3472
	.3473	.3465	.3471
2	.3471	.3465	.3471
	.3471	.3464	.3471
3	.3472	.3467	.3471
	.3472	.3464	.3471
4	.3474	.3470	.3473
	.3475	.3470	.3474
5	.3474	.3470	.3473
	.3474	.3470	.3473
6	.3472	.3463	.3471
	.3472	.3464	.3471
7	.3473	.3465	.3472
	.3473	.3469	.3471
8	.3474	.3470	.3473
	.3473	.3470	.3473
9	.3472	.3465	.3472
	.3472	.3466	.3471
10	.3474	.3470	.3474
	.3474	.3470	.3473

ANOVA Table for Diameter (1st set of Data)

Source	SS	df	MS
Part	1.3	9	.145
Operator	3.78	2	1.89
Part×Operator	.321	18	.0178
Error	.195	30	.0065
Total	5.601	59	

(a) Plot the bolt shank diameter measurements versus part number using a different plotting symbol for each operator. (You may wish to also plot part×operator means and connect consecutive ones for a given operator with line segments.) Discuss what your plot reveals about the measurement system.

(b) Find an ANOVA-based estimate of repeatability standard deviation..

(c) Find an ANOVA-based estimated standard deviation for measurement assuming there is no repeatability component of variation.

(d) Using your answers to (b) and (c), estimate the percent of total measurement variance due to repeatability.

(e) Using your answers to (b) and (c), estimate the percent of total measurement variance due to reproducibility.

(f) Discuss the relationship of your plot in (a) to your answers to (b) through (e).

(g) Find an ANOVA-based estimate of the gauge capability ratio. Is the measurement process acceptable for checking conformance to the specifications? Why or why not?

15. Refer to the **Bolt Shanks** case in problem 14. The data below are from three new operators with a different set of ten body-bound bolt shanks (numbered as part 11 through part 20). An appropriate ANOVA is also provided for these new data (units for the SS's and MS's are $10^{-6}\,\text{in}^2$).

Part	Operator D	Operator E	Operator F
11	.3694	.3693	.3693
	.3694	.3693	.3693
12	.3693	.3693	.3692
	.3693	.3692	.3692
13	.3698	.3697	.3697
	.3697	.3697	.3697
14	.3697	.3698	.3697
	.3696	.3697	.3697
15	.3694	.3695	.3695
	.3693	.3695	.3694
16	.3692	.3692	.3692
	.3693	.3692	.3691
17	.3696	.3695	.3695
	.3696	.3695	.3695
18	.3697	.3696	.3696
	.3696	.3696	.3696
19	.3697	.3696	.3695
	.3696	.3695	.3696
20	.3697	.3697	.3698
	.3697	.3698	.3697

ANOVA Table for Diameter (2nd set of Data)

Source	SS	df	MS
Part	2.08	9	.231
Operator	.016	2	.008
Part×Operator	.0873	18	.00485
Error	.07	30	.00233
Total	2.254	59	

(a) Answer (a) through (g) from problem 14 for these new data.

(b) Are your answers to (a) qualitatively different than those for problem 14? If your answer is yes, in what ways do the results differ, and what might be sources of the differences?

(c) Do conclusions from this R&R study indicate a more consistent measurement process for body-bound bolt shanks than those in problem 14? Why or why not?

16. **Transmission Gear Measurement.** Cummins, Rosario, and Vanek studied two gauges used to measure ring gear height and bevel gear height in the production of transmission differentials. (Ring gear height and bevel gear height determine the milling points for the customized transmission housings, creating the horizontal location in the housing and the "tightness" of the casing against the differential.) A test stand (hydraulically) puts a 1000 pound force on the differential. This force is used to keep the differential from free spinning while allowing spin with some force applied. A $3\,\text{in}$ Mitoya digital depth micrometer and a $6\,\text{in}$ Mitoya digital depth micrometer were used to make the measurements. Vanek used the $3\,\text{in}$ micrometer and took two ring gear height measurements on differential 8D4. Using the same $3\,\text{in}$ Mitoya micrometer, Cummins made two ring gear height measurements on the same part. Vanek then took two bevel gear height measurements with the $6\,\text{in}$ Mitoya micrometer on the same differential. Cummins followed with the same $6\,\text{in}$ micrometer and took two bevel gear height measurements on differential 8D4. This protocol was repeated two more times for the differential 8D4. The whole procedure was then applied to differential 31D4. The data follow. ANOVAs are given for both the ring gear data (SS and MS units are $10^{-4}\,\text{in}^2$) and the bevel gear data (SS and MS units are $10^{-5}\,\text{in}^2$).

Ring Gear Heights (inches) ($3\,\text{in}$ Mitoya Micrometer)

	Vanek	Cummins
8D4	1.88515	1.88470
	1.88515	1.88470
	1.88540	1.88380
	1.88530	1.88510
	1.88485	1.88435
	1.88490	1.88450
31D4	1.88365	1.88270
	1.88370	1.88295
	1.88330	1.88235
	1.88325	1.88235
	1.88270	1.88280
	1.88265	1.88260

Bevel Gear Heights (inches) ($6\,\text{in}$ Mitoya Micrometer)

	Vanek	Cummins
8D4	5.49950	5.49850
	5.49985	5.49945
	5.49975	5.49945
	5.50000	5.50005
	5.49930	5.50070
	5.49945	5.49945
31D4	5.49785	5.49700
	5.49775	5.49710
	5.49765	5.49615
	5.49750	5.49615
	5.49670	5.49595
	5.49680	5.49620

ANOVA Table for Ring Gear Height

Source	SS	df	MS
Differential	.219	1	.219
Operator	.021	1	.021
Differential×Operator	.0000042	1	.0000042
Error	.0249	20	.00124
Total	.2644	23	

ANOVA Table for Bevel Gear Height

Source	SS	df	MS
Differential	4.44	1	4.44
Operator	.148	1	.148
Differential×Operator	.124	1	.124
Error	.550	20	.02752
Total	5.262	23	

(a) Consider the ring gear heights measured with the $3\,\text{in}$ Mitoya micrometer. Give the values of m, I, and J.

(b) In the context of the ring gear height measurements, what do m, I, and J represent?

(c) Give an ANOVA-based estimated repeatability standard deviation for ring gear height measuring. Find a range-based estimate of this quantity.

(d) Give an ANOVA-based estimated reproducibility standard deviation for ring gear height measuring.

(e) The upper and lower specifications for ring gear heights are respectively $1.92\,\text{in}$ and $1.88\,\text{in}$. If the company requires the gauge capability ratio to be no larger than $.05$, does the $3\,\text{in}$ Mitoya micrometer, as currently used, seem to meet this requirement? Why or why not?

(f) Repeat (a) through (e) for bevel gear heights measured with the $6\,\text{in}$ Mitoya micrometer. Lower and upper specifications are respectively $5.50\,\text{in}$ and $5.53\,\text{in}$ for the bevel gear heights.

17. **Computer Locks.** Cheng, Lourits, Hugraha, and Sarief decided to study "tip diameter" for some computer safety locks produced by a campus machine shop. The team began its work with an evaluation of measurement precision for tip diameters. The following data are in inches and represent two diameter measurements for each of two analysts made on all 25 locks machined on one day. An appropriate ANOVA is also given. (The units for the SS's and MS's are $10^{-4}\,\text{in}^2$.)

Part	Lourits	Cheng
1	.375, .375	.374, .374
2	.375, .375	.377, .376
3	.375, .373	.374, .375
4	.375, .373	.375, .374
5	.374, .374	.374, .374
6	.374, .374	.374, .375
7	.374, .375	.375, .376
8	.374, .375	.374, .373
9	.374, .374	.375, .375
10	.374, .374	.374, .374
11	.375, .373	.374, .374
12	.375, .374	.376, .374
13	.376, .373	.373, .374
14	.373, .373	.379, .374
15	.372, .373	.374, .373
16	.373, .373	.374, .374
17	.373, .373	.374, .373
18	.373, .373	.373, .373
19	.373, .373	.376, .373
20	.373, .373	.373, .373
21	.374, .374	.374, .375
22	.375, .375	.374, .377
23	.375, .375	.376, .377
24	.376, .375	.376, .374
25	.374, .374	.374, .375

ANOVA Table for Diameter

Source	SS	df	MS
Part	.58	24	.0242
Operator	.0625	1	.0625
Part×Operator	.22	24	.00917
Error	.445	50	.0089
Total	1.3075	99	

(a) Organizations typically establish their own guidelines for interpreting the results of gauge R&R studies. One set of guidelines is below. ($6\hat{\sigma}_{\text{repeatability}}/(U-L)$ expressed as a percentage is sometimes called the "% gauge" for repeatability. $6\hat{\sigma}_{\text{reproducibility}}/(U-L)$ expressed as a percentage is sometimes called the "% gauge" for reproducibility.)

% gauge	Rating
33%	unacceptable
20%	marginal
10%	acceptable
2%	good
1%	excellent

Suppose that specifications for the lock tip diameters are $.375 \pm .002$ in. According to the guidelines above and using ANOVA-based estimates, how does the diameter measuring process "rate" (based on "% gauge" for repeatability and "% gauge" for reproducibility)? Why?

(b) Find expressions for $\overline{y}_{\text{operator1}}$ and $\overline{y}_{\text{operator2}}$ as functions of the model terms used in the equation $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$.

(c) Continuing with (b) and applying logic consistent with that used to develop equation (2.30), what does $|\overline{y}_{\text{operator1}} - \overline{y}_{\text{operator2}}|/d_2(2)$ estimate in terms of $\sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}^2$, and σ^2?

18. Refer to the **Computer Locks** case in problem 17. Consider the measurements made by Lourits. The sample average tip diameter for the ith randomly selected lock measured by Lourits can be written (holding only Lourits fixed) as

$$\overline{y}_{i\text{Lourits}} = \mu + \alpha_i + \beta_{\text{Lourits}} + \alpha\beta_{i\text{Lourits}} + \overline{\epsilon}_{i\text{Lourits}} \, .$$

(a) What is the random portion of $\overline{y}_{i\text{Lourits}}$?

(b) In terms of $\sigma^2, \sigma_\alpha^2, \sigma_\beta^2$, and $\sigma_{\alpha\beta}^2$, give the variance of your answer to part (a).

(c) Letting Γ be the range of the 25 variables $\overline{y}_{i\text{Lourits}}$, what does $\Gamma/d_2(25)$ estimate?

(d) Give the observed numerical value for $\Gamma/d_2(25)$ considered in part (c).

(e) In terms of $\sigma^2, \sigma_\alpha^2, \sigma_\beta^2$, and $\sigma_{\alpha\beta}^2$, what is the variance of (different) lock tip diameters as measured by a single operator (say Lourits) assuming there is no repeatability variation?

(f) In terms of $\sigma^2, \sigma_\alpha^2, \sigma_\beta^2$, and $\sigma_{\alpha\beta}^2$, what is the variance of (single) diameter measurements made on (different) lock tips made by the same operator (say Lourits)? (Hint: This is your answer to (e) plus the repeatability variance, σ^2.)

(g) Using the Lourits data, find a range-based estimate of the repeatability variance.

(h) Using the Lourits data, find a range-based estimate your answer to (e). (Hint: Use your answers for (d) and (g) appropriately.)

(i) Using the Lourits data, estimate your answer to (f). (Hint: Use your answers for (h) and (g) appropriately.)

19. **Implement Hardness.** Olsen, Hegstrom, and Casterton worked with a farm implement manufacturer on the hardness of a steel part. Before process monitoring and experimental design methodology were considered, the consistency of relevant hardness measurement was evaluated. Nine parts were obtained from a production line and three operators agreed to participate in the measuring process evaluation. Each operator made two readings on each of nine parts. The data below are in mm. An appropriate ANOVA is given (the units for the SS's and MS's are mm^2.)

Part	Operator A	Operator B	Operator C
1	3.30	3.25	3.30
	3.30	3.30	3.30
2	3.20	3.20	3.15
	3.25	3.30	3.30
3	3.20	3.20	3.25
	3.30	3.20	3.20
4	3.25	3.20	3.20
	3.30	3.25	3.20
5	3.25	3.10	3.20
	3.30	3.10	3.15
6	3.30	3.30	3.25
	3.30	3.20	3.20
7	3.15	3.10	3.15
	3.20	3.20	3.20
8	3.25	3.20	3.20
	3.20	3.20	3.25
9	3.25	3.20	3.30
	3.30	3.30	3.40

ANOVA Table for Hardness

Source	SS	df	MS
Part	.08833	8	.01104
Operator	.01778	2	.00889
Part×Operator	.04139	16	.00259
Error	.0575	27	.002130
Total	.205	59	

(a) Say what each term in equation (2.25) means in the context of this problem.

(b) What are the values of I, J, and m in this study?

(c) Give an ANOVA-based estimate of the repeatability standard deviation, σ.

(d) Give an ANOVA-based estimate of the reproducibility standard deviation, $\sigma_{\text{reproducibility}}$.

(e) Estimate the gauge capability ratio using the an ANOVA-based calculation if specifications on the hardness of this part are $nominal \pm 10\,\mathrm{mm}$.

(f) Using the corporate gauge rating table given in problem 17, rate the repeatability and the reproducibility of the hardness measurement method.

(g) Does it appear the current measuring process is adequate to check conformance to $nominal \pm 10\,\mathrm{mm}$ hardness specifications? Why or why not?

20. Refer to the **Implement Hardness** case in problem 19.

(a) Suppose each operator used a different gauge to measure hardness. How would this affect the interpretation of your calculations in exercise (2.19)?

(b) If it were known that measuring alters the part hardness in the vicinity of the point tested, how should this be addressed in a gauge R&R study?

(c) When an operator measures the same part two times in a row, it is likely the second measurement is "influenced" by the first in the sense that there is psychological pressure to produce a second measurement like the initial one. How might this affect results in a gauge R&R study? How could this problem be addressed/eliminated?

21. Is it important to include an evaluation of measuring processes early in a quality improvement effort? Why or why not?

22. Management tells engineers involved in a quality improvement project "We did a gauge R&R study last year and the estimated gauge capability ratio was $.005$. You don't need to redo the study." How should the engineers respond and why?

23. **Paper Weight.** Everingham, Hart, Hartong, Spears, and Jobe studied the top loading balance used by the Paper Science Department at Miami University, Oxford, Ohio. Two $20\,\mathrm{cm} \times 20\,\mathrm{cm}$ ($400\,\mathrm{cm}^2$) pieces of $20\,\mathrm{lb}$ bond paper were cut from several hundred feet of paper made in a departmental laboratory. Weights of the pieces obtained using the balance are given below in grams. The numbers in parentheses specify the order in which the measurements were made. (Piece 1 was measured 15 times, 3 times by each operator. That is, piece 1 was measured 1st by Spears, 2nd by Spears, 3rd by Hart,...,14th by Hartong, and lastly by Jobe.) Different orders were used for pieces 1 and 2, and both were determined using a random number generator. Usually, the upper specification minus the lower specification $(U - L)$ is about $4\,\mathrm{g}/\,\mathrm{m}^2$ for the density of this type of paper. An appropriate ANOVA is given below (units for the SS's and MS's are g^2).

Piece	Hartong	Hart	Spears	Everingham	Jobe
1	(14) 3.481	(3) 3.448	(1) 3.485	(13) 3.475	(10) 3.472
	(12) 3.477	(9) 3.472	(2) 3.464	(4) 3.472	(5) 3.470
	(7) 3.470	(6) 3.470	(11) 3.477	(8) 3.473	(15) 3.474
2	(1) 3.258	(13) 3.245	(7) 3.256	(6) 3.249	(11) 3.241
	(2) 3.254	(12) 3.247	(5) 3.257	(15) 3.238	(8) 3.250
	(3) 3.258	(9) 3.239	(10) 3.245	(14) 3.240	(4) 3.254

ANOVA Table for Weight

Source	SS	df	MS
Piece	.37386	1	.37386
Operator	.00061	4	.000152
Piece×Operator	.00013	4	.000032
Error	.00095	20	.000047
Total	.37555	29	

(a) What purpose is potentially served by randomizing the order of measurement as was done in this study?

(b) Give the table of operator×piece ranges, R_{ij}.

(c) Give the table of operator×piece averages, $\overline{y}_{ij}$.

(d) Give the ranges of the operator×piece means, Δ_i.

(e) Express the observed weight range determined by Spears for piece 2 in $\mathrm{g/m^2}$. (Note: $10^4\,\mathrm{cm}^2 = 1\,\mathrm{m}^2$.)

(f) Find a gauge repeatability rating based on ranges. (See part (a) of problem 17.) Pay attention to units.

(g) Find a gauge reproducibility rating based on ranges. (Again see part (a) of problem 17 and pay attention to units.)

(h) Calculate an estimated gauge capability ratio. Pay attention to units.

(i) What minimum value for $(U - L)$ would guarantee an estimated gauge capability ratio of at most .1?

(j) Using ANOVA-based estimates, answer (e)-(h).

(k) Using ANOVA-based estimates, give an exact 95% confidence interval for $\sigma_{\text{repeatability}}$. You units should be $\mathrm{g/\,m^2}$.

(l) Using the ANOVA-based estimates, give 95% approximate confidence limits for $\sigma_{\text{reproducibility}}$. Your units should be $\mathrm{g/\,m^2}$.

24. **Paper Thickness.** Everingham, Hart, Hartong, Spears, and Jobe continued their evaluation of the measuring equipment in the Paper Science Lab at Miami University by investigating the repeatability and reproducibility of the TMI automatic micrometer routinely used to measure paper thickness. The same two

$20\,\mathrm{cm} \times 20\,\mathrm{cm}$ pieces of $20\,\mathrm{lb}$ bond paper referred to in problem 23 were used in this study. But unlike measuring weight, measuring thickness alters the properties of the portion of the paper tested (by compressing it and thus changing the thickness). So, an 8×8 grid was marked on each piece of paper. The corresponding squares were labeled $1, 2, \ldots, 64$ left to right, top to bottom. Ten squares from a given piece were randomly allocated to each operator (50 squares from each piece were measured). Because so many measurements were to be made, only the "turn" for each analyst was determined randomly, and each operator made all 10 of his measurements on a given piece consecutively. A second randomization and corresponding order of measurement was made for piece 2. Hartong measured 3rd on piece 1 and 5th on piece 2, Hart was 1st on piece 1 and 3rd on piece 2, Spears was 5th and 4th, Everingham was 2nd and 2nd, and Jobe was 4th and 1st. The data follow (in mm). The numbers in parenthesis identify the squares (from a given piece) measured. (Thus, for piece 1, Hart began the measurement procedure by recording thicknesses for squares 51, 54, 18, $63, \ldots, 7$, then Everingham measured squares $33, 38, \ldots, 5$, etc. After the data for piece 1 were obtained, measurement on piece 2 began. Jobe measured squares $9, 3, \ldots, 22$ then Everingham measured squares $43, 21, \ldots, 57$, etc.) An appropriate ANOVA is also given (units for the SS's and MS's are mm^2).

Piece	Hartong	Hart	Spears	Everingham	Jobe
1	(14) .201	(51) .195	(48) .192	(33) .183	(43) .185
	(25) .190	(54) .210	(58) .191	(38) .189	(40) .204
	(17) .190	(18) .200	(15) .198	(36) .196	(49) .194
	(21) .194	(63) .203	(55) .197	(3) .195	(12) .199
	(53) .212	(20) .196	(44) .207	(59) .192	(29) .192
	(16) .209	(50) .189	(23) .202	(45) .195	(13) .193
	(47) .208	(31) .205	(64) .196	(41) .185	(56) .190
	(42) .192	(37) .203	(57) .188	(9) .193	(2) .195
	(22) .198	(34) .195	(26) .201	(62) .194	(8) .199
	(35) .191	(7) .186	(1) .181	(5) .194	(6) .197
2	(5) .188	(14) .186	(55) .177	(43) .179	(9) .191
	(16) .173	(24) .171	(51) .174	(21) .194	(3) .180
	(11) .188	(62) .178	(36) .184	(18) .187	(42) .194
	(47) .180	(34) .175	(12) .180	(39) .175	(50) .183
	(25) .178	(29) .183	(38) .179	(6) .173	(53) .181
	(15) .188	(10) .185	(41) .186	(7) .179	(17) .188
	(56) .166	(30) .190	(63) .183	(64) .171	(33) .188
	(26) .173	(40) .177	(45) .172	(54) .184	(23) .173
	(8) .175	(58) .184	(31) .174	(59) .181	(60) .180
	(52) .183	(13) .186	(2) .178	(57) .187	(22) .176

ANOVA Table for Thickness

Source	SS	df	MS
Piece	.00557	1	.00557
Operator	.00018	4	.000045
Piece×Operator	.00028	4	.00007
Error	.003986	90	.000044
Total	.010013	99	

(a) Say what each term in equation (2.25) means in the context of this problem.

(b) How is this study different from a "garden variety" gauge R&R study?

(c) Will the nonstandard feature of this study tend to increase, decrease, or have no effect on the estimate of the repeatability standard deviation? Why?

(d) Will the nonstandard feature of this study tend to increase, decrease, or have no effect on the estimated standard deviation of measurements from a given piece across many operators? Why?

(e) Give the ANOVA-based estimated standard deviation of paper thickness measurements for a fixed piece×operator combination, i.e., approximate the repeatability standard deviation assuming that square-to-square variation is negligible.

(f) Give the ANOVA-based estimated standard deviation of thicknesses measured on a fixed piece across many operators. (The quantity being estimated should include but not be limited to variability for a fixed piece×operator combination.) That is, approximate the reproducibility standard deviation assuming square-to-square variation is negligible.

(g) What percent of the overall measurement variance is due to repeatability? What part is due to reproducibility?

25. **Paper Burst Strength.** An important property of finished paper is the force ($\mathrm{lb}/\mathrm{in}^2$) required to burst or break through it. Everingham, Hart, Hartong, Spears, and Jobe investigated the repeatability and reproducibility of existing measurement technology for this paper property. A Mullen tester in the Miami University Paper Science Department was studied. Since the same two $20\,\mathrm{cm}\times 20\,\mathrm{cm}$ pieces of paper referred to in problems 23 and 24 were available, the team used them in its gauge R&R study for burst strength measurement. The burst test destroys the portion of paper tested, so repeat measurement of exactly the same paper specimen is not possible. Hence, a grid of 10 approximately equal-sized rectangles, $10\,\mathrm{cm}\times 4\,\mathrm{cm}$ (each large enough for the burst tester), was marked on each large paper piece. Each of the analysts was assigned to measure burst strength on two randomly selected rectangles from each piece. The measurement order was also randomized among the five operators for each paper piece. The data obtained are below. The ordered pairs (a, b) specify the rectangle measured and the order of measurement. (For example, the ordered pair (2,9) in the top half of the table indicates that $8.8\,\mathrm{lb}/\mathrm{in}^2$ was obtained from rectangle number 2, the 9th rectangle measured from piece 1.) An ANOVA table for this study is also provided.

Piece	Hartong	Hart	Spears	Everingham	Jobe
1	(9,2) 13.5	(6,6) 10.5	(4,8) 12.9	(2,9) 8.8	(3,10) 12.4
	(7,5) 14.8	(5,1) 11.7	(1,4) 12.0	(8,3) 13.5	(10,7) 16.0
2	(3,9) 11.3	(1,8) 14.0	(5,6) 13.0	(6,7) 12.6	(2,1) 11.0
	(8,10) 12.0	(7,5) 12.5	(9,3) 13.1	(4,2) 12.7	(10,4) 10.6

ANOVA Table for Burst Strength

Source	SS	df	MS
Piece	.5445	1	.5445
Operator	2.692	4	.6730
Piece×Operator	24.498	4	6.1245
Error	20.955	10	2.0955
Total	48.6895	19	

In the following, assume that specimen-to-specimen variation within a given piece of paper is negligible.

(a) To what set of operators can the conclusions of this study be applied?

(b) To what set of paper pieces can the conclusions of this study correctly be applied?

(c) What are the values of I, J, and m in this study?

(d) Give an ANOVA-based estimate of the repeatability standard deviation, σ.

(e) Give another estimate of the repeatability standard deviation, σ, this time based on ranges.

(f) Find an ANOVA-based estimate of $\sigma_{\text{reproducibility}}$.

(g) Find another estimate of $\sigma_{\text{reproducibility}}$, this one based on ranges

(h) Estimate the standard deviation of single burst measurements on a fixed piece of paper made by many operators, $\sigma_{\text{R\&R}}$.

26. **Paper Tensile Strength.** The final type of measurement method studied by Everingham, Hart, Hartong, Spears, and Jobe in the Paper Science Lab at Miami University was that for paper tensile strength. Since the burst tests discussed in problem 25 destroyed the $20\,\text{cm} \times 20\,\text{cm}$ pieces of $20\,\text{lb}$ bond paper referred to there, two new $20\,\text{cm} \times 20\,\text{cm}$ pieces of paper were selected from the same run of paper. Ten $15\,\text{mm} \times 20\,\text{cm}$strips were cut from each $20\,\text{cm} \times 20\,\text{cm}$ piece. Each set of ten strips was randomly allocated among the five operators (2 strips per operator for each set of ten). The order of testing was randomized for the ten strips from each piece, and the same Thwing-Albert Intellect 500 tensile tester was used by each operator to measure the load required to pull apart the strips. The data appear below in kg. (Consider, for example, the data given for piece 1, Hartong, (9,2) 4.95. A 4.95 kg load was required to tear strip number 9 from

piece 1 and the measurement was taken second in order among the ten strips measured for piece 1.) Since the testing destroyed the strip, the analysts had to assume strip-to-strip variation for a given piece to be negligible. An appropriate ANOVA is also given below (units for SS's and MS's are kg^2).

Piece	Everingham	Hart	Hartong	Spears	Jobe
1	(2,8) 4.34	(1,5) 4.34	(9,2) 4.95	(6,6) 4.03	(10,4) 4.51
	(8,10) 4.71	(4,3) 4.61	(7,7) 4.53	(3,9) 3.62	(5,1) 4.56
2	(4,7) 5.65	(6,6) 4.80	(1,1) 4.38	(2,2) 4.65	(9,5) 4.30
	(8,9) 4.51	(10,8) 4.75	(3,3) 3.89	(5,4) 5.06	(7,10) 3.87

ANOVA Table for Tensile Strength

Source	SS	df	MS
Piece	.13778	1	.1378
Operator	.69077	4	.17269
Piece×Operator	1.88967	4	.47242
Error	1.226	10	.1226
Total	3.9442	19	

(a) Make a table of load averages, $\overline{y}_{ij}$, for the 10 operator×piece combinations.

(b) Plot the load averages $\overline{y}_{ij}$ versus piece number for each of the operators (connect the two $\overline{y}_{ij}$'s for each operator).

(c) Suppose the target tensile strength for strips of $20\,\mathrm{lb}$ bond paper is $4.8\,\mathrm{kg}$. Typically, upper and lower specifications for paper properties are set 5% above and below a target. Estimate the gauge capability ratio under these conditions, using ANOVA-based calculations.

(d) If upper and lower specifications for tensile strength of $20\,\mathrm{lb}$ bond paper are equal distances above and below a target of $4.8\ \mathrm{kg}$, find the upper and lower limits such that the estimated gauge capability ratio is .01.

(e) Redo part (d) for an estimated gauge capability ratio of .1.

(f) Is it easier to make a gauge capability ratio better (smaller) by increasing its denominator or decreasing its numerator? Will your answer lead to a more consistent final product? Why or why not?

27. **Thorium Detection.** In the article "Limits for Qualitative Detection and Quantitative Determination," which appeared in *Analytical Chemistry* in 1968, L. Currie reported some experimental observations in the spectrophotometric determination of thorium using thorin. The response variability from measurements on "blank" material was observed to be essentially the same as that from any (fixed) sample of interest, and extensive analysis of blank material produced a standard

deviation of measurement ($\sigma_{\text{measurement}}$) of around .002. The absorbance response for a field sample is typically expressed as the measured response minus that of a blank, and in these terms a response for a particular sample of interest was .006.

(a) If an analyst is willing to tolerate a 5% risk of incorrectly concluding that a sample contains more thorium than is present in blank material, give the critical limit L_{c}. For the sample mentioned above, what conclusion does one reach using this critical limit? (Here, both n_{new} and n_{old} are 1.)

(b) Suppose one can tolerate only a 1% chance of incorrectly concluding there is more thorium in a sample than in a blank. Give the corresponding critical limit and say what conclusion would be reached about a sample with a response of $y_{\text{new}} = y_{\text{old}} + .006$.

(c) What risk level corresponds to a critical limit of .006?

(d) What model assumptions must be made in order to answer (a) through (c)?

(e) Suppose one can tolerate a 5% risk of incorrectly concluding a sample contains more thorium than a blank and the critical value from (a) will be employed. What is the value A (expressed in terms of an excess over the mean value for a blank) such that the chance of not detecting a mean absorbance of at least A (and corresponding thorium content above that of the blank material) is only 5%?

(f) In the vocabulary of Section 2.6 what is A in part (e)?

28. Refer to the **Thorium Detection** case in problem 27.

(a) Consider part (e) in problem 27. Another analyst can tolerate only a 1% risk for both types of possible errors. Find A for this analyst.

(b) Currie stated that the calibration factor used to translate absorbance readings to concentration values for thorium is about $k = 58.2\ \text{l}/\text{g}$. (An absorbance value divided by k gives a corresponding concentration.) Express your value for part (e) of problem 27 in $\mu\,\text{g}/\,\text{l}$. (Note that $1\mu\,\text{g}$ is $10^{-6}\,\text{g}$.)

(c) Express your answer to part (a) of this problem in $\mu\,\text{g}/\,\text{l}$.

(d) Find an increase in thorium concentration (over that in a blank) that is 10 times the standard deviation of the difference in a field sample reading and a blank reading (that is, $10\sigma_{\text{measurement}}\sqrt{1+1}$). Express your answer in absorbance units and then in $\mu\,\text{g}/\,\text{l}$.

29. **Carbon Atmospheric Blank.** Currie, et al. presented the paper "Impact of the Chemical and Isotopic Blank on the Interpretation of Environmental Radiocarbon Results" at the International Radiocarbon Conference in Glasgow, Scotland, August 1994. Part of their presentation included discussion of six carbon content measurements made on a single urban atmospheric aerosol field filter blank. These six responses (in $\mu\,\text{g}$) were (in the order produced)

$$95.6,\ 73.1,\ 56.9,\ 114.4,\ 42.3,\ \text{and } 35.6.$$

(a) Give an estimate of $\sigma_{\text{measurement}}$ (using a sample standard deviation). Use this estimate in the balance of this problem as if it were exactly $\sigma_{\text{measurement}}$.

(b) Assuming measured carbon follows a normal distribution, find a critical limit, L_{c}, for determining whether a single measurement from a field sample indicates a carbon content in excess of that in a blank (that will also be measured once). Use 2.5% for the probability of incorrectly deciding that there is more carbon in the field sample than in the blank.

(c) What additional assumption (beyond normality) must hold for your answer in (b) to be valid?

(d) Suppose now that every blank has the same and *known* amount of carbon. Using your answer to (a) as the standard deviation of a single field sample measurement, find a critical limit for deciding if the average of two measurements on a field sample indicates carbon in excess of that in any blank. Use 1% as the largest risk of incorrectly deciding that the sample contains excess carbon that can be tolerated. (The intention here is that $n_{\text{new}} = 2$ and $n_{\text{old}} = \infty$.)

(e) Find a lower limit of detection based on your estimate in (a). In making this calculation use a 5% risk of incorrectly concluding a field sample contains more carbon than a blank (of unknown content). Also use 5% for the probability that a field sample at the lower limit of detection will fail to produce a difference exceeding the critical value. (Suppose that both the field sample and the blank will be measured once.)

(f) Let the two risks in (e) be 2.5%. Find a lower limit of detection.

(g) As the two risk levels decrease, what happens to the lower limit of detection? Explain in the context of this problem.

30. Refer to the **Carbon Atmospheric Blank** case in problem 29. As in problem 29, assume that the estimate from part (a) is in fact exactly equal to $\sigma_{\text{measurement}}$.

(a) Find the increase in carbon content that is 10 times the (estimated) standard deviation of the difference between the measured carbon content for a single field sample and that from a single blank. (That is, find $10\sigma_{\text{measurement}}\sqrt{1+1}$.)

(b) Suppose many measurements from the same blank are available. Answer part (a) under these new conditions, if the difference of interest is that between a single field sample measurement and the average of the large number of measurements on a single blank. (That is, find $10\sigma_{\text{measurement}}\sqrt{1+\frac{1}{\infty}}$.)

(c) Does knowing the true carbon content of a blank affect the lower limit of detection? Why or why not? (Hint: consider a single field sample measurement and a single blank measurement. Find the lower limit of detection using, say, 5% risks. Compute lower limits of detection both when the blank

content is considered to be known and then when it is unknown. Knowing the true content of a blank means that $n_{\text{old}} = \infty$.)

31. **Dimethyl Phenanthrene (DMP4) Atmospheric Blank.** The presentation referred to in problem 29 also included a discussion of six measurements of the DMP4 content of an atmospheric field filter blank. (DMP4 is a product of softwood pyrolysis.) The six responses in nanograms (10^{-9} g) were

$$8.25, 7.30, 7.27, 6.54, 6.75 \text{ and } 7.32 \text{ .}$$

(a) Suppose the first two measurements above were taken by operator 1, the second two by operator 2, and the last two readings by operator 3, but the same blank and measuring instrument were involved in each of the three pairs of measurements. Find an appropriate range-based estimate of $\sigma_{\text{measurement}} = \sigma_{\text{R\&R}}$. In the rest of this problem use this estimate as if it were perfect. (Hint: $\sigma_{\text{R\&R}} = \sqrt{\sigma^2_{\text{repeatability}} + \sigma^2_{\text{reproducibility}}}$.)

(b) Find a critical limit, L_{c}, if the probability of a false positive is to be 10% and a randomly selected analyst is to measure a single response on both a field sample and a new blank.

(c) Continuing with the scenario in (a), find a lower limit of detection where the probability of a false positive detection is 10% and the probability of a false negative is 5% for true content at the lower limit of detection.

32. In the context of Section 2.6, consider a situation where the probabilities of both a false positive and a false negative (when the new mean is at the lower limit of detection) are set at .05.

(a) Make a 2×2 table giving formulas for critical limits in column 1 and formulas for lower limits of detection in column 2, where row 1 in the table corresponds to the case of one measurement from a field sample and one from a blank and row 2 in the table corresponds to one measurement from a field sample and a large number from a single blank.

(b) Add an additional column to the table in (a). Fill in this column with expressions for 10 times the standard deviation of $\overline{y}_{\text{field}} - \overline{y}_{\text{blank}}$.

33. **Lab Carbon Blank.** The following data were provided by L. A. Currie of the National Institute of Standards and Technology (NIST). The data are preliminary and exploratory, but real. The unit of measure is "instrument response" and is approximately equal to $1\mu\,\text{g}$ of carbon. That is, 5.18 corresponds to 5.18 instrument units of carbon or about $5.18\mu\,\text{g}$ of carbon. The responses come from blank material generated in the lab.

Test Number	1	2	3	4	5	6	7
Measured Carbon	5.18	1.91	6.66	1.12	2.79	3.91	2.87
Test Number	8	9	10	11	12	13	14
Measured Carbon	4.72	3.68	3.54	2.15	2.82	4.38	1.64

(a) Plot measured carbon content versus order of measurement.

(b) The data are ordered in time, but intervals between measurements are not equal and an appropriate plan for obtaining data was not necessarily in place. What feature of the plot in (a) might still have meaning?

(c) If one treats the measurement of lab-generated blank material as repeated measurements of a single blank, what does a trend on a plot like that in (a) suggest regarding $\sigma_{\text{repeatability}}$? (Assume the plot is made from data equally spaced in time and collected by a single individual.)

(d) Make a frequency histogram of these data with categories $1.00 - 1.99$, $2.00 - 2.99$, etc.

(e) What could be missed in a gauge R&R study if order of measurement was important (and one didn't make a plot like that in (a)) for data like these?

CHAPTER 3

Process Monitoring

This chapter discusses the important topic of process monitoring using so-called "control charts." These are devices for the routine and organized plotting of process performance measures, with the goal of identifying process changes. When these are detected, those running the process can either intervene and set things aright (if the change is detrimental) or try to make permanent the source of an unexpected process improvement.

The discussion begins with some control charting philosophy. Then the standard Shewhart control charts for both measurements/"variables data" and counts/"attributes data" are presented in consecutive sections. A fourth section discusses qualitative interpretation and practical implications of patterns sometimes seen on Shewhart charts, and some sets of rules often applied to check for such patterns. Then there is a presentation of the so-called Average Run Length concept that can be used to quantify what a particular process monitoring scheme can be expected to provide. Finally, the chapter closes with a section clarifying the relationship between "statistical process control" and "engineering control" and presenting some basic concepts of so-called PID engineering control schemes.

3.1 Generalities About Shewhart Control Charting

Section 1.2.1 introduced the notion of process "stability" as consistency over time in the pattern of process variation. Walter Shewhart, working at Bell Labs in the late 1920s and early 1930s, developed an extremely powerful and simple tool for investigating whether a process can be sensibly thought of as stable. He called it a "control

chart." Nearly 80 years after the fact, your authors would prefer (for reasons laid out in Section 3.6) that Shewhart had chosen instead the name "monitoring chart." Nevertheless, this book will use Shewhart's terminology and the "monitoring chart" terminology interchangeably.

Shewhart's Grand Conceptualization

Shewhart's fundamental conceptualization was that while some variation is inevitable in any real process, overall variation seen in process data can be decomposed as

$$\text{observed variation} = \text{baseline variation} + \text{variation that can be eliminated}\,. \tag{3.1}$$

Shewhart conceived of baseline variation as that variability in production and measurement which will remain even under the most careful process monitoring and appropriate physical intervention. It is an inherent property of a combination of system configuration and measurement methodology that cannot be reduced without basic changes in the physical process or how it is run or observed. This variation is sometimes called variation due to "system" or "common" (universal) causes. Other names for it that will be used in this book are "random" or "short-term" variation. It is the kind of variation expected under the best of circumstances, measuring item-to-consecutive-item produced on a production line. It is variation that comes from many small, unnameable, and unrecognized physical causes. When only this kind of variation is present, it is reasonable to call a process "stable" and model observations on it as independent random draws from a fixed population or universe.

The second component of overall variability portrayed in equation (3.1) is that which can potentially be eliminated by careful process monitoring and wise physical intervention (when such is warranted). This has variously been called "special cause" or "assignable" cause variation, "nonrandom" and "long-term" variation. It is the kind of systematic, persistent change that accompanies real (typically unintended) physical alteration of a process (or the measurement system used to observe it). It is change that is large enough that one can potentially track down and eliminate its root cause, leaving behind a stable process.

If one accepts Shewhart's conceptualization (3.1), the problem then becomes one of detecting the presence of the second kind of variation so that appropriate steps can be taken to eliminate it. The Shewhart control chart is a tool for making such detection.

The basic working of Shewhart's charting method is this. One periodically takes samples from the process of interest (more will be said later about the timing and nature of these samples) and computes a statistic meant to summarize process behavior at the period in question. Values of the statistic are plotted against time order of observation and compared to so-called **control limits** drawn on the chart. These separate values of the statistic that are plausible if the process is in fact stable, from those that are rare or implausible under this scenario. As long as the plotted points remain inside the control limits, one presumes that all is well (the process is stable) and does not intervene in its workings. (This is an oversimplification of how these charts are often used that will be corrected in Section 3.4. But for the time being this simplified picture will suffice.) When a point plots outside control limits, there is an indication that a physical change has probably taken place and that intervention is appropriate. Figure 3.1 shows a generic Shewhart control chart where the plotted statistic is Q, upper

Control Limits

and lower control limits are UCL_Q and LCL_Q respectively, and there is one "out of control" point.

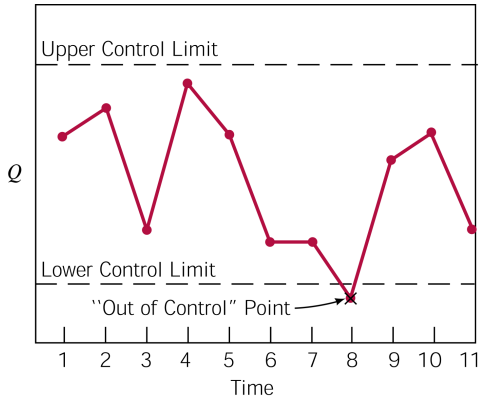


FIGURE 3.1. Generic Shewhart control chart (for a statistic Q)

There are many different kinds of Shewhart charts, corresponding to various choices of the plotted statistic, Q. Some of these chart types will be discussed in the next two sections. But before moving to discussion of specific charts, several generalities remain to be considered. First, there is the question of how one sets the control limits, UCL_Q and LCL_Q.

Shewhart's suggestion for setting control limits was essentially the following. If one can model the process output under stable conditions (i.e., if one can specify a sensible probability distribution for individual observations made on the process) then probability theory can often be invoked to produce a corresponding distribution for Q. Then small upper and lower percentage points for this distribution can provide the necessary control limits. The thinking is that only rarely will values outside these be seen under stable process conditions. Further, rather than working explicitly with probability tables or formulas for a distribution of Q, one often simply makes use of the fact that for many probability distributions most of the probability is within three standard deviations of the mean. So, if μ_Q and σ_Q are respectively a stable-process mean and standard deviation for Q, then common control limits are

Generic 3-sigma Control Limits

$$UCL_Q = \mu_Q + 3\sigma_Q \quad \text{and} \quad LCL_Q = \mu_Q - 3\sigma_Q \; . \tag{3.2}$$

Further, it is common to draw in a "center line" on a Shewhart control chart at

Generic Center Line

$$CL_Q = \mu_Q \; . \tag{3.3}$$

To make this discussion slightly more concrete, consider briefly the situation where the plotted statistic is the sample mean of n individual measurements, $Q = \overline{x}$. If the process output can be modeled as independent selections from a distribution with mean μ and standard deviation σ, the statistic $\overline{x}$ has a distribution with mean $\mu_Q = \mu_{\overline{x}} = \mu$

and standard deviation $\sigma_Q = \sigma_{\overline{x}} = \sigma/\sqrt{n}$. Then applying relationships (3.2) and (3.3) it follows that typical control limits for $\overline{x}$ are

$$UCL_{\overline{x}} = \mu + 3\,\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad LCL_{\overline{x}} = \mu - 3\,\frac{\sigma}{\sqrt{n}}\,, \tag{3.4}$$

with a center line drawn at μ.

Display (3.4) helps bring into focus another general issue regarding Shewhart control charting. As in limits (3.4), process parameters (like μ and σ) typically appear in formulas for control limits. Values for them must come from somewhere in order to apply a control chart, and there are two possibilities in this regard. Sometimes past experience with a process, engineering standards, or other considerations made prior to process monitoring specify what values should be used. This kind of situation is commonly known as a **standards given** scenario. In other circumstances, one has no information on a process outside a series of samples that are presented along with the question "Is it plausible that the process was physically stable over the period represented by these data?" In such a case, all that one can do is tentatively assume that in fact the process was stable, make provisional estimates of process parameters and plug them into formulas for control limits, and apply those limits to the data in hand as a means of criticizing the tentative assumption of stability. This kind of situation is sometimes called an **as past data** scenario, and will often be referred to in this text as a **retrospective** scenario.

Standards Given Context

Retrospective Context

The difference between what is possible in standards given and retrospective contexts can be thought of in terms of two different questions addressed in the two situations. In a standards given context, with each new sample one can face the question

"Are process parameters currently at their standard values?"

In a retrospective context, one can only wait until a number of samples have been collected (often, a minimum of 20–25 time periods is recommended) and then looking back over the data ask the question

"Are these data consistent with *any* fixed set of process parameters?"

Having introduced the notion of control limits, it is important to warn readers of a common pitfall. That is the confusion that students (and even practicing engineers) often have regarding the *much different* concepts of control limits and engineering specifications. Control limits have to do with assessing process stability. They refer to a statistic Q. They are usually derived from what a process has done in the past or is currently doing. On the other hand, engineering specifications have to do with assessing product acceptability or functionality. They almost always refer to individual measurements. They are usually derived from product performance requirements, and may have little or nothing to do with the inherent capability of a process to produce a product meeting those requirements.

Control Limits vs Engineering Specifications

Despite these real differences in meaning, people often confuse these concepts (for example, applying specifications to sample means as if they were control limits, or arguing that since a mean or individual is inside control limits for $\overline{x}$, the product being

monitored is acceptable). But it is vital that these notions be kept separate and applied in their proper contexts. (Notice that a process that is stable and producing Q's inside appropriate control limits need *not* be producing mostly acceptable product. And conversely, a process may produce product that is acceptable by current engineering standards, but nevertheless be very unstable!)

Another issue needing some general discussion here is the matter of sampling. How should one go about gathering the data to be used in control charting? This matter includes the issue sometimes referred to as **rational subgrouping** or rational sampling. When one is collecting process-monitoring data, it is important that anything one intends to call a single "sample" be collected over a short enough time span that there is little question that the process was physically stable during the data collection period. It must be clear that an "independent draws from a single population/universe" model is appropriate for describing data in a given sample. This is because the variation within such a sample essentially specifies the level of background noise against which one looks for process changes. If what one calls "samples" often contain data from genuinely different process conditions, the apparent level of background noise will be so large that it will be hard to see important process changes. In high-volume manufacturing applications of control charts, single samples (rational subgroups) typically consist of n consecutive items taken from a production line. On the other hand, in extremely low-volume operations, where one unit might take many hours to produce and there is significant opportunity for real process change between consecutive units, the only natural samples may be of size $n = 1$.

Once one has determined to group only observations close together in time into samples or subgroups, there is still the question of how often these samples should be taken. When monitoring a machine that turns out 1000 parts per hour, where samples are going to consist of $n = 5$ consecutive parts produced on the machine, does one sample once every minute, once every hour, once every day, or what? An answer to this kind of question depends upon what one expects in terms of process performance, and the consequences of process changes. If the consequences of a process change are disastrous, one is pushed toward frequent samples. The same is true if significant process upsets are a frequent occurrence. On the other hand, if a process rarely experiences changes and even when those occur only a moderate loss is incurred when it takes a while to discover them, long intervals between samples are sensible. Various operations-research type attempts have been made to provide quantitative guidelines in answer to this sampling frequency question, but these have proved largely unsatisfactory for practice. But the qualitative matters noted here clearly need to be the major considerations as looks for an appropriate sampling frequency for process monitoring.

As a final matter in this introductory discussion of Shewhart charting philosophy we should say what control charting can and cannot reasonably be expected to provide. It can signal the need for process intervention and can keep one from ill-advised and detrimental over-adjustment of a process that is behaving in a stable fashion. But in doing so, what is achieved is simply reducing variation to the minimum possible for a given system configuration (in terms of equipment, methods of operation, methods of measurement, etc.). Once that minimum has been reached, what is accomplished

is maintaining a *status quo* best possible process performance. (Remember, for example, the use of the "control" step in the six-sigma paradigm discussed on page 10.) In today's global economy, standing still is never good enough for very long. Achieving process stability provides a solid background against which to evaluate possible innovations and fundamental/order-of-magnitude improvements in production methods. But it does not itself guide their discovery. Of the tools discussed in this book, it is the methods of experimental design and analysis covered in Chapters 5 and 6 that have the most to say about aiding fundamental innovations.

Section 3.1 Exercises

1. What can control charting contribute to a process improvement effort?
2. What is the difference between "standards given" and "retrospective" control charting?
3. What is the difference between common cause and special cause variation? Which type of variation are control charts designed to detect?
4. What happens to the control limits (3.4) for an $\overline{x}$ chart as the subgroup size gets large?
5. How do you expect the behavior of a control charting scheme to change if a value smaller than 3 is used in limits (3.2)?
6. How do you expect the behavior of a control charting scheme to change if a value larger than 3 is used in limits (3.2)?
7. If the plotted statistic Q is inside appropriately constructed control limits (indicating that a process is stable), does that necessarily imply that the process is producing acceptable product? Briefly explain.
8. If the plotted statistic Q is regularly outside appropriately constructed control limits (indicating that a process is unstable), does that necessarily imply that the process is producing unacceptable product? Briefly explain.
9. The same item is being produced on two production lines. Every 15 minutes 5 items are sampled from each line and a feature of interest is measured on each item. Some statistic Q is calculated for each set of 5 measurement from each line and plotted versus time. Analyst 1 puts all 10 items together into a single group (5 from line 1 and 5 from line 2), calculates a value of the statistic Q and plots it. (This person says, "After all, isn't a larger sample size better?") Analyst 2 keeps the data from the two different lines separate and makes a different control chart for each production line.

   (a) What subgroup size is Analyst 1 using?

(b) What subgroup size is Analyst 2 using?

(c) Which analyst is making the most appropriate chart? Why? (Hint: Consider the concept of rational subgrouping.)

3.2 Shewhart Charts for Measurements/"Variables Data"

This section considers the problem of process monitoring when the data available are measurements (as opposed to counts or the kind of $0/1$ calls considered in Section 2.7). Sometimes the terminology "variables data" is used in this context. In such situations, it is common to make charts for both the process location and also for the process spread (size of the process short-term variability). So this section will consider the making of $\overline{x}$ and median ($\tilde{x}$) charts for location, and R and s charts for spread.

3.2.1 Charts for Process Location

The most common of all Shewhart control charts is that for means of samples of n measurements, the case where $Q = \overline{x}$. As was discussed in the previous section (and portrayed in display (3.4)), the fact that sampling from a distribution with mean μ and standard deviation σ produces sample averages with expected value $\mu_{\overline{x}} = \mu$ and standard deviation $\sigma_{\overline{x}} = \sigma/\sqrt{n}$ suggests **standards given** Shewhart control limits for $\overline{x}$

$$UCL_{\overline{x}} = \mu + 3\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad LCL_{\overline{x}} = \mu - 3\frac{\sigma}{\sqrt{n}}\,, \tag{3.5}$$

Standards Given $\overline{x}$ Chart Control Limits

and center line at

$$CL_{\overline{x}} = \mu\,.$$

Standards Given $\overline{x}$ Chart Center Line

Example 25 ***Monitoring the Surface Roughness of Reamed Holes.*** *Dohm, Hong, Hugget, and Knoot worked with a manufacturer on a project involving roughness measurement after the reaming of preformed holes in a metal part. Table 3.1 contains some summary statistics (the sample mean $\overline{x}$, the sample median $\tilde{x}$, the sample range R, and the sample standard deviation s) for 20 samples (taken over a period of 10 days) of $n = 5$ consecutive reamed holes.*

Suppose for the time being that standards (established on the basis of previous experience with this reaming process) for surface roughness are $\mu = 30$ and $\sigma = 4$. Then, standards given control limits for the $\overline{x}$ values in Table 3.1 are

$$\text{UCL}_{\overline{x}} = 30 + 3\frac{4}{\sqrt{5}} = 35.37$$

and

$$\text{LCL}_{\overline{x}} = 30 - 3\frac{4}{\sqrt{5}} = 24.63\ .$$

Figure 3.2 is a standards given $\overline{x}$ chart for the surface roughness measurements. Based on this chart, one would detect the fact that the reaming process is not stable at standard process parameters as early as the second sample. Several of the sample means fall outside control limits, and had the control limits been applied to the data as they were collected, the need for physical intervention would have been signaled.

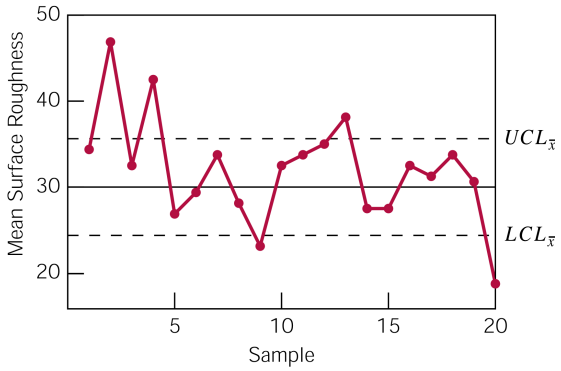


FIGURE 3.2. Standards given $\overline{x}$ chart for surface roughness

In order to make a **retrospective** $\overline{x}$ chart one must derive estimates of the process parameters μ and σ from data in hand (temporarily assuming process stability) and plug them into the formulas (3.5). There are many possible ways of doing this, each leading to slightly different retrospective control limits. Here only the most common ones will be considered and we begin with the matter of estimating μ.

Let r stand for the number of samples available in a retrospective $\overline{x}$ chart analysis. One way of estimating a supposedly common process mean for the r periods is to simply average the r sample means. Standard control charting practice is to use

TABLE 3.1. Summary Statistics for 20 Samples of 5 Surface Roughness Measurements on Reamed Holes (μ in)

Sample	$\overline{x}$	$\tilde{x}$	R	s
1	34.6	35	9	3.4
2	46.8	45	23	8.8
3	32.6	34	12	4.6
4	42.6	41	6	2.7
5	26.6	28	5	2.4
6	29.6	30	2	0.9
7	33.6	31	13	6.0
8	28.2	30	5	2.5
9	25.8	26	9	3.2
10	32.6	30	15	7.5
11	34.0	30	22	9.1
12	34.8	35	5	1.9
13	36.2	36	3	1.3
14	27.4	23	24	9.6
15	27.2	28	3	1.3
16	32.8	32	5	2.2
17	31.0	30	6	2.5
18	33.8	32	6	2.7
19	30.8	30	4	1.6
20	21.0	21	2	1.0

Average
Sample Mean

$$\overline{\overline{x}} = \frac{1}{r}\sum_{i=1}^{r}\overline{x}_i$$

as an estimator of μ in making retrospective control limits for $\overline{x}$.

An answer to the question of how to estimate σ is not quite so obvious. The estimator of σ with the best theoretical properties is obtained by pooling the r sample variances to obtain (in the constant sample size case)

$$s^2_{\text{pooled}} = \frac{1}{r}\sum_{i=1}^{r} s_i^2 \ ,$$

and then taking the square root. However, this method is not common in practice (due to historical precedent). Instead, common practice is to use estimators based on the average sample range or the average sample standard deviation.

Consider first the estimation of σ based on

Average Sample Range

$$\overline{R} = \frac{1}{r}\sum_{i=1}^{r} R_i \ .$$

As in the discussion of range-based estimation in gauge R&R on page 67, if process output is normally distributed at time period i,

$$\mathrm{E}R_i = d_2\sigma$$

and thus

$$\mathrm{E}\left(\frac{R_i}{d_2}\right) = \sigma \ .$$

(The dependence of d_2 on n is not being displayed here, since there is no chance of confusion regarding which "sample size" is under discussion.) So assuming the process is stable over all r periods, all sample sizes are n, and that a normal distribution governs the data generation process,

$$\frac{\overline{R}}{d_2}$$

is a sensible estimator of σ. Plugging this and $\overline{\overline{x}}$ into the standards given control limits for $\overline{x}$ provided in display (3.5) one obtains retrospective Shewhart control limits for $\overline{x}$,

$$UCL_{\overline{x}} = \overline{\overline{x}} + 3\frac{\overline{R}}{d_2\sqrt{n}} \qquad \text{and} \qquad LCL_{\overline{x}} = \overline{\overline{x}} - 3\frac{\overline{R}}{d_2\sqrt{n}} \ . \tag{3.6}$$

Further, one can define a constant A_2 (depending upon n) by

$$A_2 = \frac{3}{d_2\sqrt{n}} \ ,$$

and rewrite display (3.6) more compactly as

Retrospective Control Limits for $\overline{x}$ Based on the Average Range

$$UCL_{\overline{x}} = \overline{\overline{x}} + A_2\overline{R} \qquad \text{and} \qquad LCL_{\overline{x}} = \overline{\overline{x}} - A_2\overline{R} \ . \tag{3.7}$$

Values of A_2 can be found in the table of control chart constants, Table A.1.

As an alternative to estimating σ on the basis of sample ranges, next consider estimating σ based on the average sample standard deviation,

$$\overline{s} = \frac{1}{r}\sum_{i=1}^{r} s_i \ .$$

Average Sample Standard Deviation

When sampling from a normal distribution with standard deviation σ, the sample standard deviation, s, has a mean that is not quite σ. The ratio of the mean of s to σ is commonly called c_4. (c_4 depends upon the sample size and again is tabled in Table A.1, but it will not be necessary to display the dependence of c_4 on n.) Thus, if one assumes that process output is normally distributed at period i,

$$\mathrm{E}\left(\frac{s_i}{c_4}\right) = \sigma \ .$$

So assuming the process is stable over all r periods, all sample sizes are n, and that a normal distribution governs the data generation process,

$$\frac{\overline{s}}{c_4}$$

is a sensible estimator of σ. Plugging this and $\overline{\overline{x}}$ into the standards given control limits for $\overline{x}$ provided in display (3.5) one obtains retrospective Shewhart control limits for $\overline{x}$,

$$UCL_{\overline{x}} = \overline{\overline{x}} + 3\frac{\overline{s}}{c_4\sqrt{n}} \qquad \text{and} \qquad LCL_{\overline{x}} = \overline{\overline{x}} - 3\frac{\overline{s}}{c_4\sqrt{n}} \ . \tag{3.8}$$

Further, one can define another constant A_3 (depending upon n) by

$$A_3 = \frac{3}{c_4\sqrt{n}} \ ,$$

and rewrite display (3.8) more compactly as

$$UCL_{\overline{x}} = \overline{\overline{x}} + A_3\overline{s} \qquad \text{and} \qquad LCL_{\overline{x}} = \overline{\overline{x}} - A_3\overline{s} \ . \tag{3.9}$$

Retrospective Control Limits for $\overline{x}$ Based on the Average Standard Deviation

Values of A_3 can also be found in the table of control chart constants, Table A.1.

Example 26 ***(Example 25 continued.)*** *Returning to the reaming study, from Table 3.1*

$$\overline{\overline{x}} = 32.1, \overline{R} = 8.95, \text{and } \overline{s} = 3.76 \ .$$

Further, for $n = 5$ (which was the sample size used in the study) Table A.1 shows that $A_2 = .577$ and $A_3 = 1.427$. Thus, from formulas (3.7), retrospective control limits for $\overline{x}$ based on $\overline{R}$ are

$$\text{UCL}_{\overline{x}} = 32.1 + .577(8.95) = 37.26 \quad \text{and} \quad \text{LCL}_{\overline{x}} = 32.1 - .577(8.95) = 26.94\,.$$

And from formulas (3.9), retrospective control limits for $\overline{x}$ based on $\overline{s}$ are

$$\text{UCL}_{\overline{x}} = 32.1 + 1.427(3.76) = 37.47 \text{ and } \text{LCL}_{\overline{x}} = 32.1 - 1.427(3.76) = 26.73\,.$$

Figure 3.3 shows the retrospective $\overline{x}$ control chart with control limits based on $\overline{R}$. It is clear from this chart (as it would be using the limits based on $\overline{s}$) that the reaming process was not stable over the period of the study. The mean measured roughness fluctuated far more than one would expect under any stable process model.

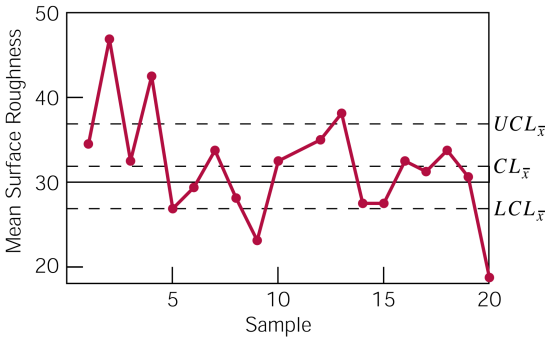


FIGURE 3.3. Retrospective $\overline{x}$ chart for surface roughness

$\overline{x}$ charts are by far the most common charts for monitoring process location, but there is an alternative worth mentioning. That is to use **sample medians** in place of sample means ($\tilde{x}$ in place of $\overline{x}$). This alternative has the advantage of requiring less in the way of computational skills from those who must compute the values to be plotted, but has the drawback of being somewhat less sensitive to changes in process location than the $\overline{x}$ chart.

The basic probability facts that lead to control limits for $\tilde{x}$ concern sampling from a normal distribution. For a sample of size n from a normal distribution with mean μ and standard deviation σ, the random variable $\tilde{x}$ has mean $\mu_{\tilde{x}} = \mu$ and standard deviation $\sigma_{\tilde{x}} = \kappa\sigma_{\overline{x}} = \kappa\sigma/\sqrt{n}$ for a constant κ (depending upon n). Table 3.2 gives a few values of κ.

Applying these facts about the probability distribution of $\tilde{x}$ under a normal process model and the generic Shewhart control limits given in display (3.2) produces standards given control limits for $\tilde{x}$

TABLE 3.2. Ratios κ Between $\sigma_{\tilde{x}}$ and $\sigma_{\overline{x}}$ When Sampling from a Normal Distribution

n	3	5	7	9	11	∞
κ	1.160	1.197	1.214	1.223	1.229	$\sqrt{\pi/2}$

Standards Given Control Limits for Medians

$$UCL_{\tilde{x}} = \mu + 3\kappa\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad LCL_{\tilde{x}} = \mu - 3\kappa\frac{\sigma}{\sqrt{n}} \; . \tag{3.10}$$

Retrospective limits can be made by replacing μ and σ with any sensible estimates.

Example 27 ***(Examples 25 and 26 continued.)*** *Returning to the reaming study, suppose once more that process standards are $\mu = 30$ and $\sigma = 4$. Then for samples of size $n = 5$ (like those used in the students' project) control limits for sample medians are*

$$\text{UCL}_{\tilde{x}} = 30 + 3(1.197)\frac{4}{\sqrt{5}} = 36.42$$

and

$$\text{LCL}_{\tilde{x}} = 30 - 3(1.197)\frac{4}{\sqrt{5}} = 23.58 \; .$$

Had these limits been applied to the data of Table 3.1 as they were collected, the need for physical intervention would have been signaled as early as the second sample.

3.2.2 Charts for Process Spread

Our exposition of control charts for measurements began with the $\overline{x}$ chart for location because it is surely the single most commonly used process monitoring tool, and because facts from elementary probability can be invoked to quickly motivate the notion of control limits for $\overline{x}$. However, in practice it is often important to deal *first* with the issue of consistency of process spread before going on to consider consistency of process location. After all, such consistency of spread (constancy of σ) is already implicitly assumed when one sets about to compute control limits for $\overline{x}$. So it is important to now consider charts intended to monitor this aspect of process behavior. The discussion here will center on charts for ranges and standard deviations, beginning with the **range chart**.

In deriving $\overline{R}/d_2$ as an estimator of σ we have employed the fact that when sampling from a normal universe with mean μ and standard deviation σ,

$$\text{E}R = \mu_R = d_2\sigma \; . \tag{3.11}$$

The same kind of mathematics that stands behind relationship (3.11) can be used to also derive a standard deviation to associate with R. (This is a measure of spread for the probability distribution of R, which is itself a measure of spread of the sample.) It turns

out that the standard deviation of R is proportional to σ. The constant of proportionality is called d_3 and is tabled in Table A.1. (Again, d_3 depends on n, but it will not be useful to display that dependence here.) That is,

$$\sigma_R = d_3\sigma \ . \tag{3.12}$$

Now the relationships (3.11) and (3.12) together with the generic formula for Shewhart control limits given in display (3.2) and center line given in display (3.3) imply that standards given control limits for R are

$$UCL_R = (d_2 + 3d_3)\sigma \qquad \text{and} \qquad LCL_R = (d_2 - 3d_3)\sigma \tag{3.13}$$

with a center line at

Standards Given R Chart Center Line

$$CL_R = d_2\sigma \ . \tag{3.14}$$

Further, if one adopts the notations $D_2 = d_2 + 3d_3$ and $D_1 = d_2 - 3d_3$ the relationships (3.13) can be written somewhat more compactly as

Standards Given R Chart Control Limits

$$UCL_R = D_2\sigma \qquad \text{and} \qquad LCL_R = D_1\sigma \ . \tag{3.15}$$

Values of the constants D_1 and D_2 may again be found in the table of control chart constants, Table A.1.

It is instructive to look at the tabled values of D_1. There are no tabled values for sample sizes $n \leq 6$. For such sample sizes the difference $d_2 - 3d_3$ turns out to be negative. Since ranges are nonnegative, a negative lower control limit would make no sense. So standard practice for $n \leq 6$ is to use no lower control limit for R.

Consider also the implications of the fact that for $n > 6$, one typically employs a positive lower control limit for R. This means that it is possible for an R chart to signal an "out of control" situation because R is *too small*. This fact sometimes causes students consternation. After all, isn't the goal to produce *small* variation? Then why signal an alarm when R is small? The answer to this conundrum lies in remembering precisely what a control chart is meant to detect, namely *process instability/change*. It is possible for unintended causes to occasionally act on a process to reduce variability. A lower control limit on an R chart simply allows one to detect such happy events. If one can detect such a change and identify its physical source, there is the possibility of making that assignable cause part of standard practice and the accompanying decrease in σ permanent. So, the practice of using positive lower control limits for R when n is sufficiently big is one that makes perfectly good practical sense.

Example 28 ***(Examples 25 through 27 continued.)*** *Consider once more the reaming example of Dohm, Hong, Hugget, and Knoot from a standards given perspective with* $\sigma = 4$*. For samples of size* $n = 5$*, Table A.1 provides the values* $d_2 = 2.326$ *and* $D_2 = 4.918$*. So using formulas (3.14) and (3.15), standards given control chart values for* R *are*

$$\mathrm{UCL}_R = 4.918(4) = 19.7 \qquad \text{and} \qquad \mathrm{CL}_R = 2.326(4) = 9.3\,.$$

Figure 3.4 is the corresponding standards given control chart for the students' ranges. There are three out-of-control points on the chart, the first coming as early as the second sample. The reaming process did not behave in a manner consistent with the $\sigma = 4$ *standard over the period of the study. Samples 2, 11, and 14 simply have too much internal variability to make consistency of* σ *at the value 4 believable. One wonders if perhaps the reamer was changed in the middle of these samples, with the effect that some holes were very rough while others were very smooth.*

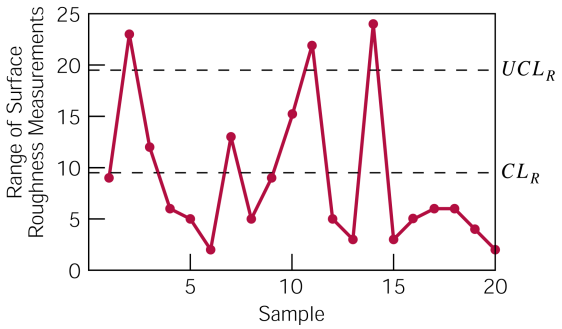


FIGURE 3.4. Standards given R chart for surface roughness

Retrospective control limits for R come about by plugging an estimate for σ derived from samples in hand into the formulas (3.14) and (3.15). A particularly natural choice for an estimator of σ in this context is $\overline{R}/d_2$. Substituting this into relationship (3.14), one gets the perfectly obvious retrospective center line for an R chart,

$$CL_R = \overline{R}\,. \tag{3.16}$$

Retrospective R Chart Center Line

Further, substituting $\overline{R}/d_2$ into equations (3.15) for σ, one gets retrospective control limits for R

$$UCL_R = D_2\left(\frac{\overline{R}}{d_2}\right) \qquad \text{and} \qquad LCL_R = D_1\left(\frac{\overline{R}}{d_2}\right)\,. \tag{3.17}$$

And adopting the notations $D_4 = D_2/d_2$ and $D_3 = D_1/d_2$, it is possible to write the relationships (3.17) more compactly as

Retrospective R Chart Control Limits

$$UCL_R = D_4\overline{R} \qquad \text{and} \qquad LCL_R = D_3\overline{R}\,. \tag{3.18}$$

As is by now to be expected, the constants D_3 and D_4 are tabled in Table A.1. And the table contains no values of D_3 for $n \leq 6$.

Example 29 ***(Examples 25 through 28 continued.)*** *Recall that the 20 samples in Table 3.1 have* $\overline{R} = 8.95$ *and note that for* $n = 5$, $D_4 = 2.114$. *So from displays (3.16) and (3.18) a retrospective control chart for the ranges (based on* $\overline{R}/d_2$ *as an estimator of* σ*) has center line at*

$$\text{CL}_R = \overline{R} = 8.95$$

and upper control limit

$$\text{UCL}_R = D_4\overline{R} = 2.114(8.95) = 18.9\,.$$

A plot of this retrospective R *chart would look very little different from Figure 3.4. The same three ranges plot outside control limits. Not only is a "*σ *constant at* 4*" view of the students' data not plausible, but neither is a "*σ *constant at some value" view. There is solid evidence of reaming process instability in the ranges of Table 3.1. The short-term process variability changes over time.*

The R chart is the most common Shewhart control chart for monitoring process spread. It requires very little from its user in the way of calculations, is based on a statistic that is very easy to understand, and is firmly entrenched in quality assurance practice dating from the days of Shewhart himself. There is, however, an alternative to the R chart that tends to detect changes in process spread more quickly, at the price of increased computational complexity. Where the quantitative sophistication of a user is high and calculations are not a problem, the s **chart** is a viable competitor for the R chart.

The fact that (when sampling from a normal distribution)

$$\text{E}s = c_4\sigma\,, \tag{3.19}$$

has already proved useful when making retrospective control limits for $\overline{x}$ based on $\overline{s}$. The same kind of mathematics that leads to relationship (3.19) can be used to find the standard deviation of s (based on a sample from a normal universe). (This is a measure of spread for the probability distribution of the random variable s, that is itself a measure of spread of the sample.) It happens that this standard deviation is a multiple of σ. The multiplier is called c_5 and it turns out that $c_5 = \sqrt{1 - c_4^2}$. That is,

$$\sigma_s = \sigma\sqrt{1 - c_4^2} = c_5\sigma\,. \tag{3.20}$$

Now relationships (3.19) and (3.20) together with the generic Shewhart control limits and center line specified in displays (3.2) and (3.3) lead immediately to standards given control limits and center line for an s chart. That is,

$$UCL_s = (c_4 + 3c_5)\sigma \qquad \text{and} \qquad LCL_s = (c_4 - 3c_5)\sigma \tag{3.21}$$

and

$$CL_s = c_4\sigma \; . \tag{3.22}$$

Standards Given s Chart Center Line

Further, if one adopts the notations $B_6 = c_4 + 3c_5$ and $B_5 = c_4 - 3c_5$, the relationships (3.21) can be written as

$$UCL_s = B_6\sigma \qquad \text{and} \qquad LCL_s = B_5\sigma \; . \tag{3.23}$$

Standards Given s Chart Control Limits

Values of the constants B_5 and B_6 may again be found in the table of control chart constants, Table A.1. For $n \leq 5$ there are no values of B_5 given in Table A.1 because for such sample sizes $c_4 - 3c_5$ is negative. For $n > 5$, B_5 is positive, allowing the s chart to provide for detection of a decrease in σ (just as is possible with an R chart and $n > 6$).

Retrospective control limits for s can be made by substituting any sensible estimate of σ into the standards given formulas (3.23) and (3.22). A particularly natural choice in this context is $\overline{s}/c_4$. Substituting this into relationship (3.23), one gets the obvious retrospective center line for an s chart

$$CL_s = \overline{s} \; . \tag{3.24}$$

Retrospective s Chart Center Line

Further, substituting $\overline{s}/c_4$ into equations (3.23) for σ produces retrospective control limits for s

$$UCL_s = B_6\left(\frac{\overline{s}}{c_4}\right) \qquad \text{and} \qquad LCL_s = B_5\left(\frac{\overline{s}}{c_4}\right) \; . \tag{3.25}$$

And adopting the notations $B_4 = B_6/c_4$ and $B_3 = B_5/c_4$, it is possible to write the relationships (3.25) more compactly as

Retrospective s Chart Control Limits

$$UCL_s = B_4\overline{s} \qquad \text{and} \qquad LCL_s = B_3\overline{s}\,. \tag{3.26}$$

As usual, the constants B_3 and B_4 are tabled in Table A.1, and the table contains no values of B_3 for $n \leq 5$.

Example 30 ***(Examples 25 through 29 continued.)*** *The 20 samples in Table 3.1 have* $\overline{s} = 3.76$. *For* $n = 5$, $B_4 = 2.089$. *So from displays (3.24) and (3.26) a retrospective control chart for the standard deviations (based on* $\overline{s}/c_4$ *as an estimator of* σ*) has center line at*

$$\text{CL}_s = \overline{s} = 3.76$$

and upper control limit

$$\text{UCL}_s = B_4\overline{s} = 2.089(3.76) = 7.85\,.$$

Figure 3.5 is a retrospective s *chart for the sample standard deviations of Table 3.1. It carries the same message as does a retrospective analysis of the sample ranges for this example. Not only is a "*σ *constant at 4" view of the students' data not plausible, neither is a "*σ *constant at some value" view. There is solid evidence of reaming process instability in the standard deviations of Table 3.1.*

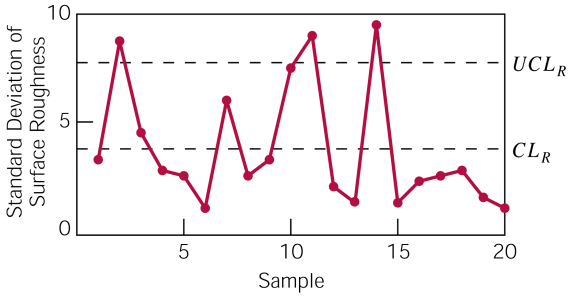


FIGURE 3.5. Retrospective s chart for surface roughness

3.2.3 What if $n = 1$?

To call n observations a "sample" or a "rational subgroup" is to implicitly guarantee that they were collected under essentially constant process conditions. The discussion in Section 3.1 has already raised the possibility (particularly in some low volume production contexts) that a natural sample or subgroup size can be $n = 1$. Sometimes it is

simply not safe to assume that even two successive process outcomes are necessarily generated under the same conditions.

There are two commonly raised questions about control charting for measurements when $n = 1$. These are

1. Exactly what should one chart (in particular, should one chart so-called "moving ranges")? and
2. How does one estimate a process standard deviation, σ?

We consider these questions before closing this section on Shewhart charting for measurements data.

Where rational subgroups are of size $n = 1$ there is really only one possible choice for a plotted statistic Q, namely x. One can "transform" the most natural measurement to some other scale (for example, by taking logarithms) but ultimately it is $Q = x$ that is available for plotting. However, people used to making $\overline{x}$ and R chart (or $\overline{x}$ and s chart) pairs in cases where $n > 1$ sometimes reason that it might be useful to supplement an x chart (or **individuals chart**) with a chart on which one plots **moving ranges**

$$MR_i = |x_i - x_{i-1}|$$

Moving Range for an tth Observation, x_i

The most commonly suggested version of this is where standards given control limits for x (the limits (3.5) for $\overline{x}$ when $n = 1$)

$$UCL_x = \mu + 3\sigma \qquad \text{and} \qquad LCL_x = \mu - 3\sigma \tag{3.27}$$

Standards Given Control Limits for Individuals, x

are used together with

$$UCL_{MR} = D_2\sigma \tag{3.28}$$

(for D_2 based on the pseudo-sample size of $n = 2$). This practice turns out to produce a very large "false alarm rate" when in fact the process is stable. And attempts to remedy this by applying control limits looser than (3.27) and (3.28) are typically not repaid with improved ability to detect process changes over what is available using only an x chart with limits (3.27). Adding a moving range chart to an individuals chart just turns out to be a bad idea that should probably disappear from control charting practice. There is, to our knowledge, only one circumstance in which adding the moving range chart to an individuals chart makes sense. That is a case where the departure from stable process behavior that one fears and needs to detect is one of *oscillation*

in consecutive individuals. There, a moving range chart is (not surprisingly) more effective than the individuals chart at "seeing" the non-standard process behavior. To be absolutely explicit, in cases where $n = 1$, *the best thing to do about control charting is typically to use only an individuals chart with corresponding control limits (3.27)*.

Chart Only x When $n = 1$

Consider then the second question above. When $n = 1$, there are no sample ranges or standard deviations to use in estimating σ. In fact, there is no really honest way to estimate a process standard deviation unless one has a sample or samples with $n \geq 2$. But some "dishonest" methods are less dishonest than others, and the best known method (the least dishonest method) is based on an average of moving ranges of successive observations. (Notice that this is *not* a matter of process monitoring based on moving ranges, but rather using moving ranges to estimate process standard deviation.)

The rationale for using moving ranges of successive observations in estimating σ is this. If process conditions can change observation to observation, observations will vary not only because $\sigma \neq 0$, but because the process mean changes. However, it is not unreasonable to expect the variability in pairs of successive observations to be less affected by mean changes than the variability of any other type of group of observations that could be made up. It is thus reasonable to use moving ranges to make an estimate of process standard deviation. While such an estimate is potentially inflated by variation in the process mean, it can be expected to be *less so* than any other estimate one might make.

The exact form of estimator of σ we'll use (based on samples of size $n = 1$) is

Moving Range-Based Estimate of σ

$$\hat{\sigma} = \frac{\overline{MR}}{d_2} \tag{3.29}$$

where d_2 is for "sample size" 2 (as there are 2 observations represented in each moving range). This is a *conservative* estimator, as it will tend to over-estimate σ when μ is not constant. But it is the best one available.

Example 31 ***A Numerical Example.*** *Consider the 8 successive observations in the table below and the corresponding 7 moving ranges.*

Sample	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
x	5	3	9	10	17	4	6	2
MR		2	6	1	7	13	2	4

The values $5, 3, 9, 10, 17, 4, 6, 2$ *certainly vary because* $\sigma \neq 0$. *They may vary beyond what can be attributed to inherent process short term variability if* μ *is not constant. That is, the* 7 *moving ranges should not be thought of as honest sample ranges, but as potentially over-representing* σ. *Nevertheless, the best available estimate of* σ

in this $n = 1$ context is from formula (3.29)

$$\begin{aligned}\hat{\sigma} &= \frac{\overline{MR}}{d_2} \\ &= \frac{(2+6+1+7+13+2+4)/7}{1.128} \\ &= 4.43\,.\end{aligned}$$

Section 3.2 Exercises

1. Some specialized containers are produced by a process that runs 8 hours per day. Nine containers are sampled hourly, each day for five days. The distance from the bottom of the container to the container's handle is of interest. The target value for this dimension is $4\,\mathrm{cm}.$ and the process standard deviation for this quality dimension is $.1\,\mathrm{cm}.$ (This is known from extensive experience with the process.)

   (a) What is a subgroup in this context? What is the subgroup size? How many subgroups make up the entire study?

   (b) Give control limits for process monitoring when subgroup averages are plotted versus time.

   (c) In addition to the chart in (b), a control chart for short term process variation is to be made. Suppose that only subgroup average and the smallest and largest values in a subgroup are available for analysis. What subgroup statistic can be used to do the charting? Give appropriate control limits and center line for the chart.

   (d) Are the limits in (b) and (c) standards given or retrospective limits? Why?

   (e) Suppose both the charts in (b) and (c) indicate that the process is stable. Is it then possible that any plotted subgroup mean is outside the limits from (b)? Is it possible that there are plotted values on the second chart outside control limits from (c)? Explain.

2. Continue in the context of problem 1, except now assume that no target value for the critical dimension or process standard deviation have previously been established..The average of the $r = 40$ subgroup averages was $3.9\,\mathrm{cm}$, the average of the subgroup ranges was $.56\,\mathrm{cm},$. and the average of the 40 subgroup standard deviations was $.48\,\mathrm{cm}.$

   (a) Find control limits and center line to assess the consistency of "hourly variation" quantified as subgroup ranges.

   (b) Find control limits and center line to assess the consistency of process aim hour-to-hour based on subgroup averages. (Estimate the "within hour" standard deviation based on subgroup ranges.)

(c) Repeat (a), except now use the subgroup standard deviations instead of ranges.

(d) Repeat (b), except now use the subgroup standard deviations to estimate the "within hour" standard deviation.

(e) Suppose that none of the charts in (a) to (d) suggests lack of process stability (so that it makes sense to talk about a single process mean and single process standard deviation). Give a valid estimate of the process average distance from container bottom to the handle. Give two valid estimates of the standard deviation of the distance from the container bottom to the handle. (Provide both the formulas you use and numerical answers.)

3. Below are sample means and standard deviations from 10 samples, each of size $n = 4$.

Sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Sum**
$\overline{x}$	7.0	7.9	7.1	7.7	5.2	5.4	6.4	6.5	5.8	6.8	65.8
$\mathbf{s}$	1.5	3.1	3.4	1.1	1.4	1.0	2.5	.7	1.4	1.1	17.2

(a) Suppose process standards $\mu = 6.0$ and $\sigma = 1.5$ are provided. Find the standards given center line and control limits for an $\overline{x}$ chart. If these limits had been applied to the values in the table as they were collected, would there have been out-of-control signals?

(b) Using the standards in (a) find the standards given center line and control limits for an s chart. If these limits had been applied to the values in the table as they were collected, would there have been out-of-control signals?

(c) Suppose that the standards in (a) were not available. Make retrospective charts and assess whether there is evidence of process instability in the values in the table.

(d) What is an estimate of σ based on the average sample standard deviation? Use this estimate and estimate the mean of a *range* for an additional sample of size $n = 6$.

4. **Transmission Housings.** Apple, Hammerand, Nelson and Seow analyzed data taken from a set of "Series 42" transmission housings. One critical dimension they examined was the diameter for a particular hole on the side cover of the housing. A total of 35 consecutively produced housings were examined and the corresponding x = hole diameter measured and recorded (in inches). Specifications for the diameter were $3.7814 \pm .002\,\text{in}$. Below are the first 10 recorded diameters. Summary statistics for all 35 housings are $\sum x = 132.319\,\text{in}$ and $\sum MR = .02472\,\text{in}$.

Housing	1	2	3	4	5	6	7	8	9	10
x	3.7804	3.7803	3.7806	3.7811	3.7812	3.7809	3.7816	3.7814	3.7809	3.7814

(a) What is a subgroup here and what is the subgroup size?

(b) The 35 consecutive hole diameters produce how many moving ranges?

(c) Compute the first two moving ranges.

(d) Make an estimate of σ. Use your estimate and the sample mean diameter to replace process parameters in the limits (3.27). Are the resulting limits for individuals standards given or retrospective limits? Why? Apply your limits to the first 10 hole diameters. Do these values provide evidence of process instability?

3.3 Shewhart Charts for Counts/"Attributes Data"

The control charts for measurements introduced in Section 3.2 are the most important of the Shewhart control charts. Where it is at all possible to make measurements, they will almost always provide more information on process behavior than will a corresponding number of qualitative observations. However, there are occasions where only attributes data can be collected. So this section presents Shewhart control charting methods for such cases. The section considers charting *counts* and corresponding *rates* of occurrence for nonconforming items (or defectives) and for nonconformities (or defects). The case of so-called np charts and p charts for "percent nonconforming" (or percent defective) contexts is treated first. Then follows a discussion of c and u charts for "nonconformities per unit" (or defects per unit) situations.

3.3.1 Charts for Fraction Nonconforming

Consider now a situation where one periodically samples n items or outcomes from a process and (making careful use of operational definitions) classifies each one as "nonconforming" or "conforming." (The old terminology for these possibilities is "defective" and "nondefective." The newer terminology is used in recognition of the fact that some kinds of failures to meet inspection criteria do not render a product functionally deficient. There is also reluctance in today's litigious society to ever admit that anything produced by an organization could possibly be "defective.")

Then let

$$X = \text{the number nonconforming in a sample of } n \text{ items or outcomes} \tag{3.30}$$

and

$$\hat{p} = \frac{X}{n} = \text{the fraction nonconforming in a sample of } n \text{ items or outcomes .} \tag{3.31}$$

Shewhart np **charts** are for the plotting of $Q = X$, and p **charts** are for the monitoring of $Q = \hat{p}$. Both are based on the same probability model for the variable X. (The fact that $\hat{p}$ is simply X divided by n implies that control limits for $\hat{p}$ should simply be those for X, divided by n.) Under stable process conditions for the creation of the n items or outcomes in a sample (under the assumption that the sample in question is a rational subgroup) it is reasonable to model the variable X with a binomial distribution for n "trials" and "success probability," p, equal to the process propensity for producing nonconforming outcomes.

Elementary properties of the binomial distribution can be invoked to conclude that

$$\mu_X = \mathrm{E}X = np \qquad \text{and} \qquad \sigma_X = \sqrt{\mathrm{Var}\, X} = \sqrt{np(1-p)}. \tag{3.32}$$

Then the mean and standard deviation in display (3.32) and the generic Shewhart control limits and center line specified in displays (3.2) and (3.3) lead to standards given control limits for both X and $\hat{p}$. That is,

Standards Given np Chart Center Line

$$CL_X = np \tag{3.33}$$

while

Standards Given np Chart Control Limits

$$UCL_X = np + 3\sqrt{np(1-p)} \qquad \text{and} \qquad LCL_X = np - 3\sqrt{np(1-p)}\,. \tag{3.34}$$

And dividing the expressions (3.33) and (3.34) through by n, one arrives at standards given values for $\hat{p}$,

Standards Given p Chart Center Line

$$CL_{\hat{p}} = p, \tag{3.35}$$

Standards Given p Chart Control Limits

$$UCL_{\hat{p}} = p + 3\sqrt{\frac{p(1-p)}{n}} \qquad \text{and} \qquad LCL_{\hat{p}} = p - 3\sqrt{\frac{p(1-p)}{n}}. \tag{3.36}$$

TABLE 3.3. Counts and Fractions of Nonconforming Pellets in Samples of Size 30

Sample	X	$\hat{p}$	Sample	X	$\hat{p}$
1	14	.47	14	9	.30
2	20	.67	15	16	.53
3	17	.57	16	16	.53
4	13	.43	17	15	.50
5	12	.40	18	11	.37
6	12	.40	19	17	.57
7	14	.47	20	8	.27
8	15	.50	21	16	.53
9	19	.63	22	13	.43
10	21	.70	23	16	.53
11	18	.60	24	15	.50
12	14	.47	25	13	.43
13	13	.43			

Example 32 ***Monitoring the Fraction Nonconforming in a Pelletizing Process.*** *Kaminiski, Rasavaghn, Smith, and Weitekamper worked with a manufacturer of hexamine pellets. Their work covered a time period of several days of production. Early efforts with the pelletizing machine (using shop standard operating procedures) produced a standard fraction nonconforming of approximately $p = .60$. On the final day of the study, after adjusting the "mix" of the powder being fed into the machine, the counts and proportions of nonconforming pellets in samples of size $n = 30$ portrayed in Table 3.3 were collected.*

From equations (3.34), standards given control limits for the numbers of nonconforming pellets in the samples represented by Table 3.3 are

$$\text{UCL}_X = 30(.6) + 3\sqrt{30(.6)(.4)} = 26.05$$

and

$$\text{LCL}_X = 30(.6) - 3\sqrt{30(.6)(.4)} = 9.95,$$

and from display (3.33) a center line at

$$\text{CL}_X = 30(.6) = 18$$

is in order. Figure 3.6 on page 144 is the standards given np control chart for the data of Table 3.3.

It is evident from Figure 3.6 that the pelletizing process was not stable at the standard value of $p = .60$ on the final day of the students' study. Notice that there are two out-of-control points on the chart (and most of the plotted points run below the center line established on the basis of the standard value of p). The message that was delivered at samples 14 and 20 (if not even before, on the basis of the plotted values running consistently below 18) was one of clear process improvement, presumably traceable to the change in powder mix.

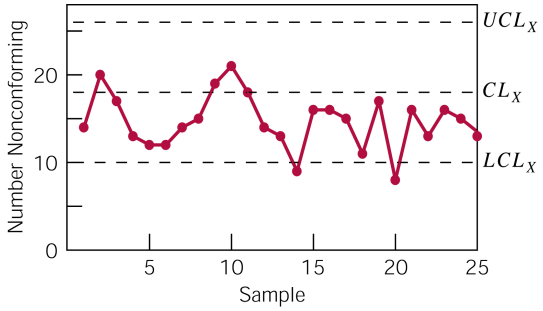


FIGURE 3.6. Standards given np chart for counts of nonconforming pellets

Example 32 nicely illustrates the fact that a positive lower control limit on an np chart or on a p chart makes perfectly good sense in terms of allowing identification of unexpectedly good process output. Remember that the objective of Shewhart charting is to detect process instability/change. On occasion, that change can be for the good.

Retrospective control limits for X or $\hat{p}$ require that one take the data in hand and produce a provisional estimate of (a supposedly constant) p for plugging into formulas (3.33) through (3.36) in place of p. If samples (of possibly different sizes) are available from r different periods, then a most natural estimator of a common p is the pooled sample fraction nonconforming

Pooled Fraction Nonconforming

$$\hat{p}_{\text{pooled}} = \frac{\sum_{i=1}^{r} n_i \hat{p}_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} X_i}{\sum_{i=1}^{r} n_i} = \frac{\text{total nonconforming}}{\text{total of the sample sizes}} . \tag{3.37}$$

Example 33 ***(Example 32 continued.)*** *Returning again to the pelletizing example, the counts of nonconforming pellets in Table 3.3 total to* 367. *There were* $30(25) = 750$ *pellets inspected, so from relationship (3.37),* $\hat{p}_{\text{pooled}} = 367/750 = .4893$. *Substituting this into equations (3.33) and (3.34) in place of* p, *one arrives at retrospective values*

$$\text{CL}_X = 30(.4893) = 14.68 ,$$

$$\text{UCL}_X = 30(.4893) + 3\sqrt{30(.4893)(.5107)} = 22.89 ,$$

and

$$\text{LCL}_X = 30(.4893) - 3\sqrt{30(.4893)(.5107)} = 6.47 .$$

Figure 3.7 is a retrospective np *chart made using these values and the data of Table 3.3. The figure shows that although it is not plausible that the pelletizing process was stable at the standard value of* p (.60) *on the final day of the students' study, it* is

plausible that the process was stable at some *value of* p*, and* .4893 *is a reasonable guess at that value.*

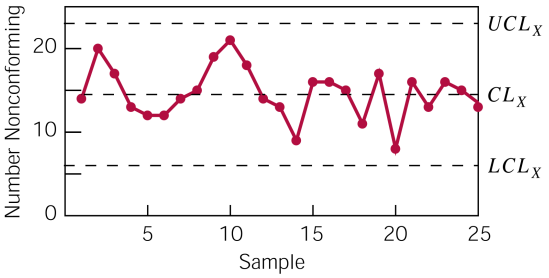


FIGURE 3.7. Retrospective np chart for counts of nonconforming pellets

A few words need to be said about cases where sample sizes vary in a fraction nonconforming context. In such situations, it makes much more sense to plot $\hat{p}$ values than it does to plot X's based on differing sample sizes. Then at least, one has a constant center line (given by expression (3.35)). Of course, the control limits represented in display (3.36) will vary with the sample size. Equations (3.36) show that the larger the sample size, the tighter will be the control limits about the central value p. This is perfectly sensible. The larger the sample, the more information about the current process propensity for producing nonconforming outcomes, and the *less* variation one should allow from the central value before declaring that there is evidence of process instability.

3.3.2 Charts for Mean Nonconformities per Unit

A second kind of situation leading to count and rate data that is fundamentally different from the fraction nonconforming scenario is the so-called "mean nonconformances/nonconformities per unit" ("or mean defects per unit") situation. In such a context, one periodically selects k inspection units from a process output and counts

$$X = \text{the total number of nonconformities on the } k \text{ units} \tag{3.38}$$

(older terminology for nonconformities is "defects" or "flaws"). In cases where k is always equal to 1, the count X itself is plotted and the resulting chart is called a c **chart.** Where k varies and/or is not equal to 1, it is common to plot instead

$$\hat{u} = \frac{X}{k} = \text{the sample mean nonconformities per unit} \tag{3.39}$$

and the resulting chart is called a u **chart.**

Control limits for c and u charts are based on the Poisson process model. If one assumes that under stable process conditions the generation of nonconformities can be

described by a Poisson process with (constant) rate parameter λ, the number of defects on one inspection unit has a Poisson distribution with mean λ. And X, the number of defects on k inspection units, is a Poisson random variable with mean $k\lambda$. Thus, under stable process conditions

$$\mu_X = \mathrm{E}X = k\lambda \qquad \text{and} \qquad \sigma_X = \sqrt{\mathrm{Var}\, X} = \sqrt{k\lambda}. \tag{3.40}$$

So using facts (3.40) and the generic Shewhart control limits and center line specified in displays (3.2) and (3.3), in the c chart situation $(k \equiv 1)$ standards given values are

Standards Given c Chart Center Line

$$CL_X = \lambda\,, \tag{3.41}$$

and

Standards Given c Chart Control Limits

$$UCL_X = \lambda + 3\sqrt{\lambda} \qquad \text{and} \qquad LCL_X = \lambda - 3\sqrt{\lambda}\,. \tag{3.42}$$

It follows from the definition of $\hat{u}$ in display (3.39) and relationships (3.40) that

$$\mu_{\hat{u}} = \mathrm{E}\hat{u} = \lambda \qquad \text{and} \qquad \sigma_{\hat{u}} = \sqrt{\mathrm{Var}\,\hat{u}} = \sqrt{\frac{\lambda}{k}}\,. \tag{3.43}$$

Then using again the generic Shewhart control limits and center line and applying the facts (3.43), standards given values for a u chart are

Standards Given u Chart Center Line

$$CL_{\hat{u}} = \lambda\,, \tag{3.44}$$

and

Standards Given u Chart Control Limits

$$UCL_{\hat{u}} = \lambda + 3\sqrt{\frac{\lambda}{k}} \qquad \text{and} \qquad LCL_{\hat{u}} = \lambda - 3\sqrt{\frac{\lambda}{k}}\,. \tag{3.45}$$

Notice that in the case $k = 1$, the u chart control limits reduce (as they should) to the c chart limits.

TABLE 3.4. Counts and Occurrence Rates of Outlet Leaks Found in 18 Daily Samples of Radiators

Day	X **(leaks)**	k **(radiators)**	$\hat{u}$ **(leaks/radiator)**
1	14	39	.36
2	4	45	.09
3	5	46	.11
4	13	48	.27
5	6	40	.15
6	2	58	.03
7	4	50	.08
8	11	50	.22
9	8	50	.16
10	10	50	.20
11	3	32	.09
12	11	50	.22
13	1	33	.03
14	3	50	.06
15	6	50	.12
16	8	50	.16
17	5	50	.10
18	2	50	.04

Retrospective control limits for X or $\hat{u}$ require that one take the data in hand and produce a provisional estimate of (a supposedly constant) λ for plugging into formulas (3.41), (3.42), (3.44), and (3.45) in place of λ. If data from r different periods are available, then a most natural estimator of a common λ is the pooled mean nonconformities per unit

$$\hat{\lambda}_{\text{pooled}} = \frac{\sum_{i=1}^{r} k_i \hat{u}_i}{\sum_{i=1}^{r} k_i} = \frac{\sum_{i=1}^{r} X_i}{\sum_{i=1}^{r} k_i} = \frac{\text{total nonconformities}}{\text{total units inspected}} \tag{3.46}$$

Pooled Mean Nonconformities Per Unit

Example 34 ***Monitoring the Number of Leaks in Assembled Radiators.*** *The article "Quality Control Proves Itself in Assembly," by Wilbur Burns (reprinted from* Industrial Quality Control*) in Volume 2, Number 1 of* Quality Engineering*, contains a classic set of data on the numbers of leaks found in samples of auto radiators at final assembly. These are reproduced in Table 3.4.*

This is a nonconformities per unit situation. Each unit (each radiator) presents the opportunity for the occurrence of any number of leaks, several units are being inspected and the total number of leaks on those units is being counted. The leaks per radiator are calculated as in display (3.39), and if one wishes to investigate the statistical evidence for process instability, a u chart is in order.

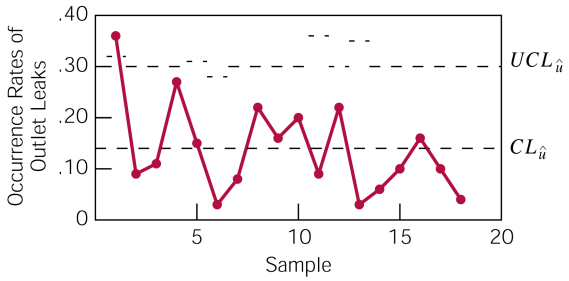


FIGURE 3.8. Retrospective u chart for rates of radiator outlet leaks

The article gives no shop standard value for λ, so consider a retrospective analysis of the data in Table 3.4. There are 116 total leaks represented in Table 3.4, and 841 radiators were tested. So from relationship (3.46)

$$\hat{\lambda}_{\text{pooled}} = \frac{116}{841} = .138,$$

and a center line for a retrospective u chart for these data can be drawn at this value. From equations (3.45) (using .138 for λ) the control limits change with k, larger k leading to tighter limits about the center line. As an example of using equations (3.45), note that for those $\hat{u}$ values based on tests of $k = 50$ radiators

$$\text{UCL}_{\hat{u}} = .138 + 3\sqrt{\frac{.138}{50}} = .296\ .$$

On the other hand, since the formula (3.45) for $LCL_{\hat{u}}$ produces a negative value for the intrinsically nonnegative $\hat{u}$, no lower control limit would be used for $\hat{u}$ based on 50 radiators. (As a matter of fact, no k in Table 3.4 is large enough to lead to the use of a lower control limit.)

Figure 3.8 is a retrospective u chart for the radiator leak data. It shows that the initial day's experience does not "fit" with the subsequent 17 days. There is evidence of process change/instability, and appearances are that things improved in the radiator assembly process after the first day.

This section opened with the disclaimer that where possible, the charts for measurements introduced in the previous section should be used in preference to the ones presented here. That advice bears repeating. The two examples in this section are reasonably convincing, but they are so in part because the relevant fraction nonconforming and mean nonconformities per unit are fairly large. Modern business pressures make standard defect rates in the "parts per million" range common. And there is really no way to effectively monitor processes that are supposed to have such performance with attributes control charts (sample sizes in the millions would be required for effective detection of even doubling of defect rates!).

Section 3.3 Exercises

1. In a packaging department of a food processor, types of packaging "imperfections" are carefully defined, and include creases, holes, printing smudges, and broken seals. 30 packages each hour are sampled and $X =$ the total number of imperfections identified on the 30 packages is recorded. On average about .05 imperfections per package have been seen in the past. Below are data from 7 hours one day in this department.

Hour	1	2	3	4	5	6	7
X	1	0	2	0	1	1	3

   (a) Are the data above variables or attributes data? Why?

   (b) What distribution (fully specify it, giving the value of any parameter(s)) can be used to model the number of imperfections observed on a single package?

   (c) What is the expected total number of imperfections observed on a set of 30 boxes? What probability distribution can be used to model this variable?

   (d) What is the standard deviation of the total number of imperfections on 30 boxes?

   (e) Find the standards given control limits and center line for a chart based on the data above, where the plotted statistic will be $X/30$. Do any values of $X/30$ plot outside of your control limits?

   (f) What is the name of the type of chart you made in (e)?

   (g) Suppose no standard is given for the rate of imperfections. Using values above, find appropriate retrospective control limits and center line for plotting $X/30$.

2. Consider a variant of problem 1 where any package with at least one imperfection (a crease, a hole, a smudge, or broken seal) is considered to be nonconforming. Reinterpret the values X in the table of problem 1 as counts of nonconforming packages in samples of size 30. Suppose that in the past .05 (5%) of packages have been nonconforming.

   (a) Does this variant of problem 1 involve variables data or attributes data? Why?

   (b) What probability distribution (fully specify it, giving the value of any parameter(s)) can be used to model the number of nonconforming packages in a sample of 30?

(c) What is the mean number of nonconforming packages in a sample of 30?

(d) What is the standard deviation of the number of nonconforming packages in a sample of 30?

(e) Find the standards given control limits and center line for monitoring the proportion of nonconforming packages in samples of size 30.

(f) Repeat (e) for monitoring the number of nonconforming packages in samples of size 30.

(g) Suppose no standard is given for the fraction of nonconforming packages. Based on the data in the table above, find appropriate retrospective control limits and center line for an np chart.

3.4 Patterns on Shewhart Charts and Special Alarm Rules

To this point all we have discussed doing with values Q plotted a Shewhart chart is to compare them to control limits one at a time. If that were the whole story, there would be little reason to actually make the plots. Simple numerical comparisons would suffice. But the plots offer the possibility of *seeing* other important things in process monitoring data besides only where points plot outside control limits. And it is standard control charting practice to examine Shewhart control charts for these other kinds of indications of process change. The purpose of this section is to discuss some types of revealing patterns that occasionally show up on control charts (providing both jargon for naming them and discussion of the kinds of physical phenomena that can stand behind them) and to provide some sets of rules that can be applied to identify them.

Under stable process conditions (leading to Q's that can be modeled as independent and identically distributed), one expects to see a sequence of plotted values that

1. are without obvious pattern or trend,
2. only on rare occasions fall outside control limits,
3. tend to cluster about the center line, about equally often above and below it, but
4. on occasion approach the control limits.

(The tacit assumption in most applications is that the stable process distribution of Q is reasonably "mound shaped" and centered at the chart's center line.) When something other than this kind of "random scatter" picture shows up on a control chart, it can be possible to get clues to what kinds of physical causes are acting on the process, that can in turn be used in process improvement efforts.

On occasion one notices **systematic variation/cycles**, regular "up then back down again" patterns on a Shewhart chart like those pictured on Figure 3.9. This suggests

that there are important variables acting on the process whose effects are periodic. Identification of the period of variation can give one strong hints where to start looking for physical causes. Examples of factors that can produce cycles on a Shewhart chart are seasonal and diurnal variables like ambient temperature. And sometimes regular rotation of fixtures or gages or shift changes in operators running equipment or making measurements can stand behind systematic variation.

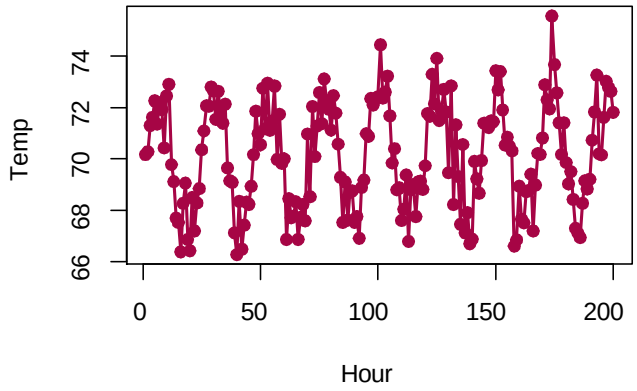


FIGURE 3.9. A plot of factory ambient temperature vs time exhibiting systematic variation or cycles

While systematic variation is variation of the "second kind" on the right side of equation (3.1), it may not always be economically feasible to eliminate it. For example, in some applications it may be preferable to live with effects of ambient temperature rather than try to control the environment in which a process operates. But recognition of its presence at least allows one to intelligently consider options regarding remedial measures, and to mentally remove that kind of variation from the baseline against which one looks for the effects of other special causes.

Instability is a word that has traditionally been applied to patterns on control charts where many points plot near or beyond control limits. This text has used (and will continue to use) the word to refer to physical changes in a process that lead to individual points plotting outside of control limits. But this second usage refers more to a pattern on the chart, and specifically to one where points outside of control limits are very frequent. Figure 3.10 on page 152 contrasts variation on a Shewhart chart that one expects to see, to a pattern of instability. Standing behind such a pattern can be more or less erratic and unexpected causes, like different lots of raw material with different physical properties mixed as process input.

Another important possible cause of many points at or beyond control limits is that of unwise operator over-adjustment of equipment. Control charting is useful both because it signals the existence of conditions that deserve physical intervention, *and* because it tells one to leave equipment untouched when it seems to be operating as

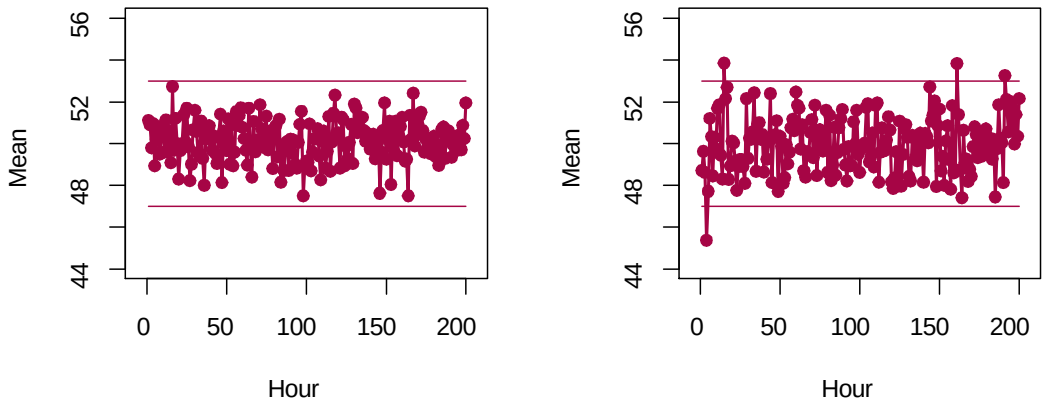


FIGURE 3.10. Two $\overline{x}$ charts, the second of which shows "instability"

consistently as possible. When that "hands-off" advice is not followed and humans tinker with physically stable processes, reacting to every small appearance of variation, the end result is not to decrease process variation, but rather to increase it. And such fiddling can turn a process that would otherwise be generating plotted values inside control limits into one that is regularly producing Q's near or beyond control limits.

Changes in level are sometimes seen on control charts, where the average plotted value seems to move decisively up or down. The change can be sudden as pictured on Figure 3.11 and traceable to some basic change at the time of the shift. The introduction of new equipment or a clear change in the quality of a raw material can produce such a sudden change in level.

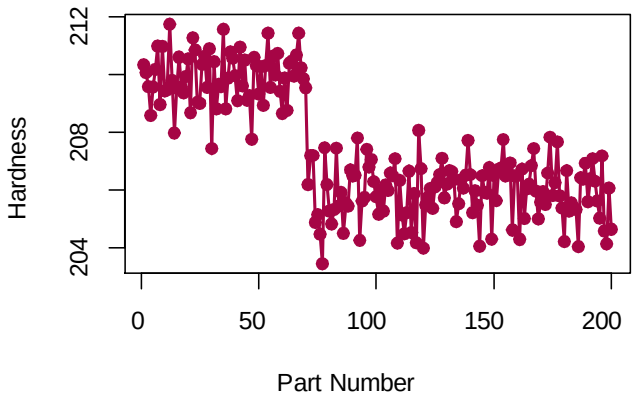


FIGURE 3.11. A sudden change in level of part hardness

A change in level can also be like that pictured in Figure 3.12, more gradual and attributable to an important cause starting to act at the beginning of the change in level, but so to speak "gathering steam" as time goes on until its full effect is felt. For example, effective worker training in machine operation and measuring techniques could well begin a gradual decrease in level on an R chart, that over time and with practice will reach its full potential for reducing observed variation.

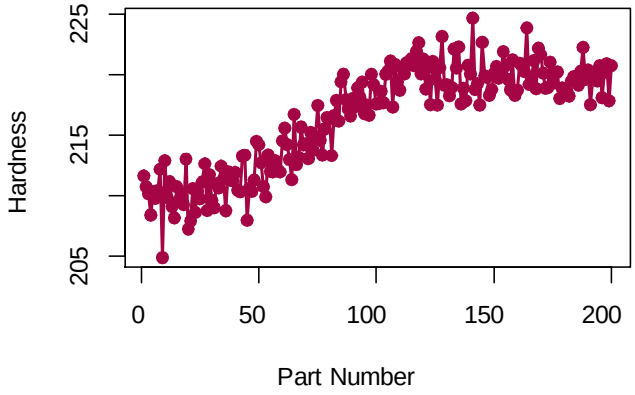


FIGURE 3.12. A gradual change in level of part hardness

Where a gradual change in level does not end with stabilization around a new mean, but would go on unabated in the absence of physical intervention, it is traditional to say that there is a **trend** on a control chart. Figure 3.13 pictures such a trend on a run chart.

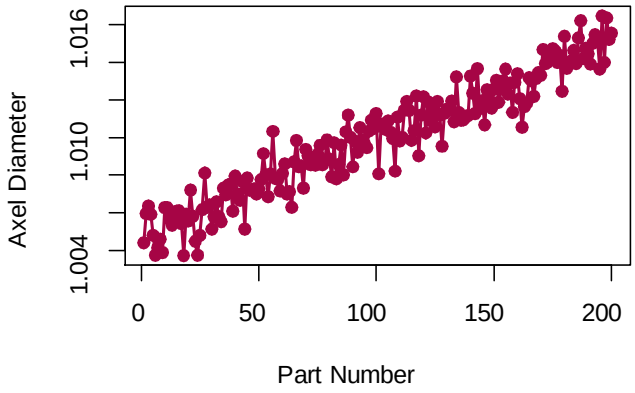


FIGURE 3.13. A run chart with an unabated trend

Many physical causes acting on manufacturing processes will produce trends if they remain unaddressed. An example is tool wear in machining processes. As a cutting tool wears, the parts being machined will tend to grow larger. If adjustments are not made and the tool is not periodically changed, machined dimensions of parts will eventually be so far from ideal as to make the parts practically unusable.

There is another phenomenon that occasionally produces strange-looking patterns on Shewhart control charts. This is something the early users of control charts called the occurrence of **mixtures.** These are the combination of two or more distinct patterns of variation (in either a plotted statistic Q, or in an underlying distribution of individual observations leading to Q) that get put together on a single control chart. In "stable" mixtures, the proportions of the component patterns remain relatively constant over time, while in "unstable" versions the proportions vary with time.

Where an underlying distribution of observations has two or more radically different components, a plotted statistic Q can be either unexpectedly variable or surprisingly consistent. Consider first the phenomenon of unexpectedly large variation in Q traceable to a mixture phenomenon. Where blunders like incomplete or omitted manufacturing operations or equipment malfunctions lead to occasional wild individual observations and correspondingly wild values of Q, the terminology **freaks** is often used. The resulting impact of mixing normal and aberrant observations can be as pictured in Figure 3.14. Where individual observations or values of Q of a given magnitude tend to occur together in time as pictured in Figure 3.15, the terminology **grouping** or **bunching** is common. Different work methods employed by different operators or changes in the calibration of a measurement instrument can be responsible for grouping or bunching. So, how mixture phenomena sometimes lead to unexpectedly large variation on a control chart is fairly obvious.

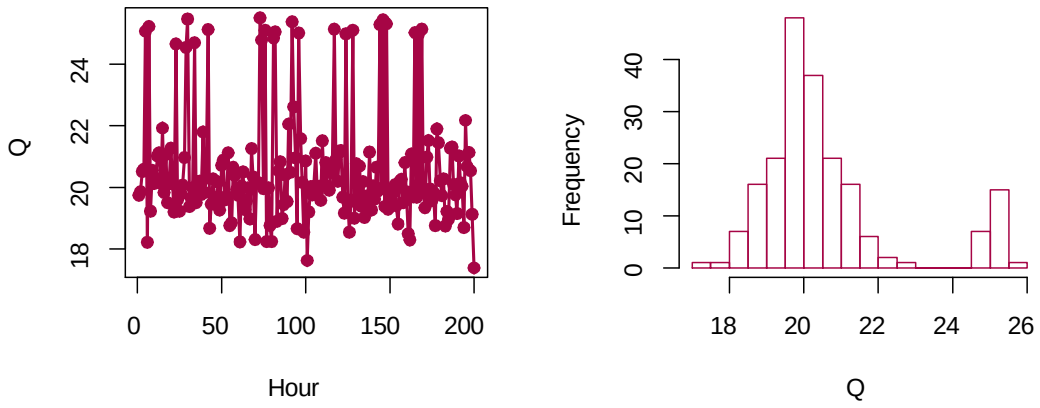


FIGURE 3.14. An example of a pattern that could be described as exhibiting "freaks" (and the corresponding histogram)

How a mixture can lead to unexpectedly small variation in a plotted statistic is more

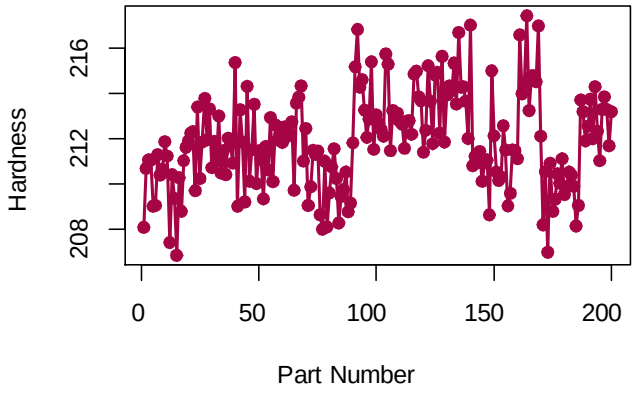


FIGURE 3.15. A run chart showing grouping or bunching

subtle, but very important. It involves a phenomenon sometimes known in quality assurance circles as **stratification.** If an underlying distribution of observations has radically different components, each with small associated variation, and these components are (wittingly or unwittingly) sampled in a systematic fashion, a series of plotted values Q with unbelievably small variation can result. One might, for example, be sampling different raw material streams or the output of different machines and unthinkingly calling the resulting values a single "sample" (in violation, by the way, of the notion of rational subgrouping). The result can be a Shewhart control chart like the one in Figure 3.16.

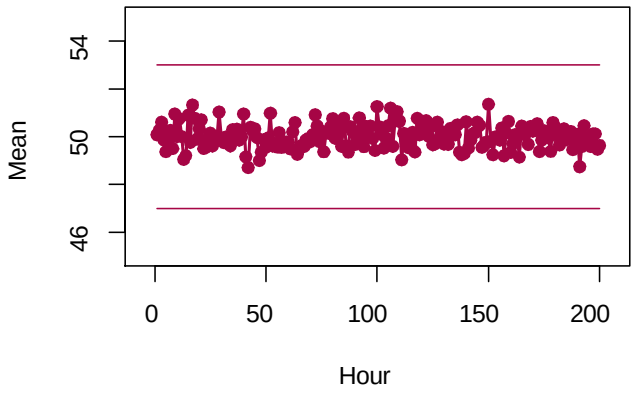


FIGURE 3.16. Unexpectedly small variation on an $\overline{x}$ chart, potentially due to stratification

To see how stratification can lead to surprisingly small variation in Q, consider the case of a p chart and a hypothetical situation where a 10-head machine has one com-

TABLE 3.5. Western Electric Alarm Rules

A single point outside 3 sigma control limits
2 out of any 3 consecutive points outside 2 sigma limits on one side of center
4 out of any 5 consecutive points outside 1 sigma limits on one side of center
8 consecutive points on one side of center

TABLE 3.6. Alarm Rules from Duncan's *Quality Control and Engineering Statistics*

A single point outside 3 sigma control limits
A run of 7 consecutive points up, down or on one side of center
2 consecutive points outside 2 sigma limits
4 consecutive points outside 1 sigma limits
"Obvious" cycles up and down

pletely bad head and 9 perfect ones. If the items from this machine are taken off the heads in sequence and placed into a production stream, "samples" of 10 consecutive items will have fractions defective that are *absolutely constant* at $\hat{p} = .10$. A p chart for the process will look unbelievably stable about a center line at $.10$. (A similar hypothetical example involving $\overline{x}$ and R charts can be invented by thinking of 9 of the 10 heads as turning out widget diameters of essentially exactly 5.000, while the 10th turns out widget diameters of essentially exactly 7.000. Ranges of "samples" of 10 consecutive parts will be unbelievably stable at 2.000 and means will be unbelievably stable at 5.200.)

So, too much consistency on a control chart is not cause for rejoicing and relaxation. When plotted points hug a center line and never approach control limits something is not as it should be. There may be a simple blunder in the computation of the control limits, or the intrinsic variation in the process may be grossly overestimated. (For example, an excessive standard value for σ produces $\overline{x}$ and R chart control limits that are too wide and plotted points that never approach them under stable conditions.) And on occasion stratification may be present. When it is and it goes unrecognized, one will never be in a position to discover and eliminate the cause(s) of the differences between the components of the underlying distribution of observations. In the 10-head machine example, someone naively happy with the "$\hat{p}$ constant at .10" phenomenon will never be in a position to discover that the one head is defective and remedy it. So, a chart

TABLE 3.7. Nelson's Alarm Rules from the *Journal of Quality Technology*

A single point outside 3 sigma control limits
9 consecutive points on one side of center
6 consecutive points increasing or decreasing
14 consecutive points alternating up and down
2 out of any 3 consecutive points outside 2 sigma limits on one side of center
4 out of any 5 consecutive points outside 1 sigma limits on one side of center
15 consecutive points inside 1 sigma limits
8 consecutive points with none inside 1 sigma limits

that looks too good to be true is as much a cause for physical investigation as is one producing points outside control limits.

Once one recognizes the possibility of looking for patterns on a Shewhart control chart, the question becomes exactly what to consider to be an occurrence of a pattern. This is important for two reasons. In the first place, there is the matter of consistency within an organization. If control charts are going to be used by more than one person, those people need a common set of ground rules for interpreting the charts that they together use. Second, without a fair amount of theoretical experience in probability and/or practical experience in using control charts, people tend to want to "see" patterns that are in actuality very easily produced by a stable process.

Since the simple "one point outside control limits" rule is blind to the interesting kinds of patterns discussed here and there is a need for some standardization of the criteria used to judge whether a pattern is present, organizations often develop sets of "special checks for unnatural patterns" for application to Shewhart control charts. These are usually based on segmenting the set of possible Q's into various zones defined in terms of multiples of σ_Q above and below the central value μ_Q. Figure 3.17 shows a generic Shewhart chart with typical zones marked on it.

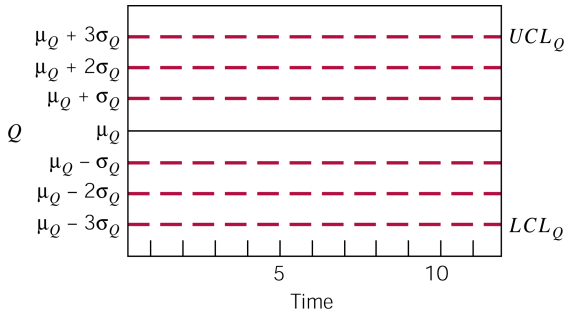


FIGURE 3.17. Generic Shewhart control chart with "zones" marked on it

By far the most famous set of special checks is the set of "Western Electric Alarm Rules" given in Table 3.5. They are discussed extensively in the *Statistical Quality Control Handbook* published originally by Western Electric and later by AT&T. Two other possible sets of rules, one taken from A.J. Duncan's excellent *Quality Control and Industrial Statistics* and the other published by Lloyd Nelson in the *Journal of Quality Technology* in 1984, are given in Tables 3.6 and Table 3.7 respectively. The reader should be able to see in these sets of rules attempts to provide operational definitions for the kinds of patterns discussed in this section. It is not at all obvious which set should be considered best, or even what are rational criteria for comparing them and the many other sets that have been suggested. But the motivation behind them should be clear.

Section 3.4 Exercises

1. When a process is stable, what do you expect to see on a control chart for a statistic Q?
2. What motivates the use of multiple rules for identifying out of control situations?
3. When "extra alarm rules" (beyond the "single point outside 3 sigma control limits" rule) are used in process monitoring, do you expect the frequency of false alarms to decrease, stay the same. or increase? (A "false alarm" occurs when the chart signals, but no physical special cause can be identified.)

3.5 The Average Run Length Concept

Realizing that alternative schemes for issuing out-of-control signals based on process-monitoring data are possible, the need arises to quantify what a given scheme can be expected to do. For example, to choose intelligently between the sets of alarm rules in Tables 3.5 through 3.7, one needs some way of predicting behavior of the alternative monitoring schemes. The most effective tool available for making this kind of prediction is the "Average Run Length" (ARL) notion. This section introduces the concept and illustrates it in some very simple situations.

Consider a context where based on values of Q plotted at periods $1, 2, 3, \ldots$ one will monitor a process until an out-of-control signal is issued. Let

$$T = \text{the period at which the process-monitoring scheme first signals .} \tag{3.47}$$

T is a random variable and is called the **run length** for the scheme. The probability distribution of T is called the **run length distribution**, and the mean or average value of this distribution is called the **Average Run Length** (ARL) for the process-monitoring scheme. That is,

$$ARL = \mathrm{E}T = \mu_T \, . \tag{3.48}$$

It is desirable that a process monitoring scheme have a large ARL when the process is stable at standard values for process parameters, and small ARLs under other conditions.

Finding formulas and numerical values for ARLs is not usually elementary. Some advanced probability and numerical analysis are often required. But there is one kind of circumstance where an explicit formula for ARLs is possible and we can illustrate the meaning and usefulness of the ARL concept in elementary terms. That is the situation where

1. the process-monitoring scheme employs only the single alarm rule "signal the first time that a point Q plots outside control limits," and
2. it is sensible to think of the process as physically stable (though perhaps not at standard values for process parameters).

Under condition 2, the values $Q_1, Q_2, Q_3, \ldots$ can be modeled as independent random variables with the same individual distribution, and the notation

$$q = P[Q_1 \text{ plots outside control limits}] \tag{3.49}$$

Probability of an Immediate Alarm

will prove useful.

In this simple case, the random variable T has a geometric distribution with probability function

$$f(t) = \begin{cases} q(1-q)^{t-1} & \text{for } t = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

It then follows from the properties of the geometric distribution and relationship (3.48) that

$$ARL = \mathrm{E}T = \frac{1}{q}. \tag{3.50}$$

ARL for a "One Point Outside Control Limits" Scheme

Example 35 ***Some ARLs for Shewhart $\overline{x}$ Charts.*** *To illustrate the meaning of relationship (3.50) consider finding ARLs for a standards given Shewhart $\overline{x}$ chart based on samples of size $n = 5$. Note that if standard values for the process mean and standard deviation are respectively μ and σ, the relevant control limits are*

$$\mathrm{UCL}_{\overline{x}} = \mu + 3\frac{\sigma}{\sqrt{5}} \qquad \text{and} \qquad \mathrm{LCL}_{\overline{x}} = \mu - 3\frac{\sigma}{\sqrt{5}}.$$

Thus, from equation (3.49)

$$q = P\left[\overline{x} < \mu - 3\frac{\sigma}{\sqrt{5}} \qquad \text{or} \qquad \overline{x} > \mu + 3\frac{\sigma}{\sqrt{5}}\right].$$

First suppose that "all is well" and the process is stable at standard values of the process parameters. Then elementary probability shows that $\mu_{\overline{x}} = \mu$ and $\sigma_{\overline{x}} = \sigma/\sqrt{5}$ and if the process output is normal, so also is the random variable $\overline{x}$. Thus

$$q = 1 - P\left[\mu - 3\frac{\sigma}{\sqrt{5}} < \overline{x} < \mu + 3\frac{\sigma}{\sqrt{5}}\right] = 1 - P\left[-3 < \frac{\overline{x} - \mu}{\sigma/\sqrt{5}} < 3\right]$$

can be evaluated using the fact that

$$Z = \frac{\overline{x} - \mu}{\sigma/\sqrt{5}}$$

is a standard normal random variable. Using a normal table with an additional significant digit beyond the one in this text it is possible to establish that

$$q = 1 - P[-3 < Z < 3] = .0027$$

to 4 digits. Therefore, from relationship (3.50) it follows that

$$\text{ARL} = \frac{1}{.0027} = 370\ .$$

The interpretation of this is that when all is OK (i.e., the process is stable and parameters are at their standard values), the $\overline{x}$ chart will issue (false alarm) signals on average only once every 370 plotted points.

In contrast to the situation where process parameters are at their standard values, consider next the possibility that the process standard deviation is at its standard value but the process mean is one standard deviation above its standard value. In these circumstances one still has $\sigma_{\overline{x}} = \sigma/\sqrt{5}$, but now $\mu_{\overline{x}} = \mu + \sigma$ (μ and σ are still the standard values of respectively the process mean and standard deviation). Then,

$$\begin{aligned}
q &= 1 - P\left[\mu - 3\frac{\sigma}{\sqrt{5}} < \overline{x} < \mu + 3\frac{\sigma}{\sqrt{5}}\right]\ , \\
&= 1 - P\left[\frac{\mu - 3\sigma/\sqrt{5} - (\mu+\sigma)}{\sigma/\sqrt{5}} < \frac{\overline{x} - (\mu+\sigma)}{\sigma/\sqrt{5}} < \frac{\mu + 3\sigma/\sqrt{5} - (\mu+\sigma)}{\sigma/\sqrt{5}}\right]\ , \\
&= 1 - P[-5.24 < Z < .76]\ , \\
&= .2236\ .
\end{aligned}$$

Figure 3.18 illustrates the calculation being done here and shows the roughly 22% chance that under these circumstances the sample mean will plot outside $\overline{x}$ chart control limits. Finally, using relationship (3.50),

$$\text{ARL} = \frac{1}{.2236} = 4.5.$$

That is, if the process mean is off target by as much as one process standard deviation, then it will take on average only 4.5 samples of size $n = 5$ to detect this kind of misadjustment.

Example 35 should agree completely with the reader's intuition about "how things should be." It says that when a process is on target, one can expect long periods between signals from an $\overline{x}$ chart. On the other hand, should the process mean shift off target by a substantial amount, there will typically be quick detection of that change.

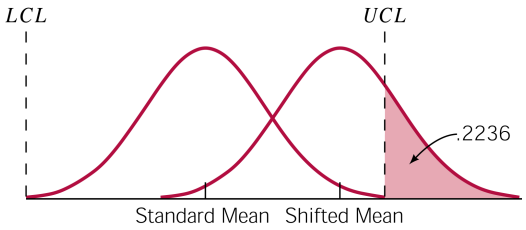


FIGURE 3.18. Two distributions for $\overline{x}$ and standard given control limits

Example 36 ***Some ARLs for Shewhart*** c ***Charts.*** *As a second example of the meaning of equation (3.50), consider finding some ARLs for two different versions of a Shewhart* c *chart when the standard rate of nonconformities is 1.5 nonconformities per unit. To begin, suppose that only one unit is inspected each period. Using relationships (3.42) with* $\lambda = 1.5$*, it follows that since* $1.5 - 3\sqrt{1.5} < 0$ *no lower control limit is used for the number of nonconformities found on an inspection unit, and*

$$\text{UCL}_X = 1.5 + 3\sqrt{1.5} = 5.2 \, .$$

So, for this situation

$$q = P[X > 5.2] = 1 - P[X \leq 5] \, .$$

Consider evaluating q *both when the nonconformity rate is at its standard value (of* $\lambda = 1.5$ *nonconformities per unit) and when it is at three times its standard value (i.e., is* 4.5 *nonconformities per unit). When the rate is standard, one uses a Poisson distribution with mean* 1.5 *for* X *and finds*

$$q = 1 - P[X \leq 5] = .005 \qquad \text{and} \qquad \text{ARL} = \frac{1}{.005} = 200 \, .$$

When the rate is three times standard, one uses a Poisson distribution with mean 4.5 *for* X *and finds*

$$q = 1 - P[X \leq 5] = .298 \qquad \text{and} \qquad \text{ARL} = \frac{1}{.298} = 3.4 \, .$$

That is, completely in accord with intuition, the mean waiting time until an alarm is much smaller when quality deteriorates than when the process defect rate is standard.

Now suppose that two units will be inspected each period. One can then either use a u *chart, or equivalently simply apply a* c *chart where the standard value of* λ *is* 3.0 *nonconformities per two units. Applying this second way of thinking and relationships (3.42) with* $\lambda = 3.0$*, it follows that since* $3.0 - 3\sqrt{3.0} < 0$ *no lower control limit is used for the number of nonconformities found on two inspection units, and*

$$\text{UCL}_X = 3.0 + 3\sqrt{3.0} = 8.2 \, .$$

So, for this situation

$$q = P[X > 8.2] = 1 - P[X \leq 8] \, .$$

TABLE 3.8. ARLs for Two c Chart Monitoring Schemes for a Standard Nonconformity Rate of 1.5 Defects per Unit

	Standard Defect Rate	**3× Standard Defect Rate**
1 Unit Inspected	200	3.4
2 Units Inspected	250	1.8

Consider again the ARLs both where the nonconformity rate is at its standard value (of $\lambda = 3.0$ nonconformities per two units) and where it is at three times its standard value (i.e., is 9.0 nonconformities per two units). When the rate is standard, one uses a Poisson distribution with mean 3.0 for X and finds

$$q = 1 - P[X \leq 8] = .004 \qquad \text{and} \qquad \text{ARL} = \frac{1}{.004} = 250.$$

When the rate is three times standard, one uses a Poisson distribution with mean 9.0 for X and finds

$$q = 1 - P[X \leq 8] = .545 \qquad \text{and} \qquad \text{ARL} = \frac{1}{.545} = 1.8.$$

Table 3.8 summarizes the calculations of this example. It shows the superiority of the monitoring scheme based on two units rather than one unit per period. The two-unit-per-period monitoring scheme has both a larger ARL when quality is standard and a smaller ARL when the nonconformity rate degrades by a factor of 3.0 than the one-unit-per-period scheme. This, of course, does not come without a price. One must do twice as much inspection for the second plan as for the first.

Examples 35 and 36 illustrate the ARL concept in very simple contexts that are covered by an elementary formula. Where the rules used to convert observed values $Q_1, Q_2, Q_3, \ldots$ into out-of-control signals or the probability model for these variables are at all complicated, explicit formulas and elementary computations are impossible. But it is not necessary to understand the nature of the numerical analysis needed to compute ARLs for more complicated cases to appreciate what an ARL tells one about a monitoring scheme.

For example, a paper by Champ and Woodall appearing in *Technometrics* in 1987 considered ARL computations for monitoring schemes using various combinations of the four Western Electric alarm rules. Example 35 showed the "all OK" ARL for an $\overline{x}$ chart scheme using only the "one point outside $3\sigma_{\overline{x}}$ control limits" rule to be about 370. When all four Western Electric rules are employed simultaneously, Champ and Woodall found that the $\overline{x}$ chart "all OK" ARL is far less than 370 (or what naive users of the rules might expect), namely approximately 92. The reduction from 370 to 92 shows the effects (in terms of increased frequency of false alarms) of allowing for signs of process change in addition to individual points outside control limits.

Section 3.5 Exercises

1. Interpret the terms "ARL" and "All OK ARL."
2. What kind of ARL does one want under a "stable at standard parameter values" process model? What kind of ARL does one hope to have under any other circumstance?
3. $n = 4$ values are sampled every hour from a process that under "All OK stable process" conditions produces observations x that are normal mean 20 and standard deviation 4. A typical Shewhart $\overline{x}$ chart is set up.
   (a) What is the All OK ARL of the monitoring scheme?
   (b) An upward shift in the process mean of at least 1 unit occurs while the process standard deviation variation does not change. At worst, how many hours on average will pass before this change produces a subgroup average outside the control limits?
4. Consider a production process where one item (the subgroup size is 1) is periodically sampled and the number of nonconformities is observed. Suppose standard nonconformity rate per item is $\lambda = 4$.
   (a) Find the All OK ARL.
   (b) Find the ARL if an increase to a rate of $\lambda = 8$ occurs.
   (c) Answer (a) and (b) if two items make up each subgroup.
5. Control charting Method A is preferred to Method B relative to an "All OK" and some "not All OK" process conditions. Which of the following is true?
   (a) $ARL_{\text{A}} > ARL_{\text{B}}$ when "All is OK" and $ARL_{\text{A}} > ARL_{\text{B}}$ when "All is not OK".
   (b) $ARL_{\text{A}} > ARL_{\text{B}}$ when "All is OK" and $ARL_{\text{A}} < ARL_{\text{B}}$ when "All is not OK".
   (c) $ARL_{\text{A}} < ARL_{\text{B}}$ when "All is OK" and $ARL_{\text{A}} > ARL_{\text{B}}$ when "All is not OK".
   (d) $ARL_{\text{A}} < ARL_{\text{B}}$ when "All is OK" and $ARL_{\text{A}} < ARL_{\text{B}}$ when "All is not OK".
6. Process standards are $\mu = 100$ and $\sigma = 7$ and observations from the process are normally distributed. A Shewhart $\overline{x}$ chart is being considered for use in monitoring the process.
   (a) The charts with $n = 5$ and $n = 10$ will have different control limits. Why?
   (b) The charts with $n = 5$ and $n = 10$ will have the same ARL if process parameters remain at standard values. Why?

3.6 Statistical Process Monitoring and Engineering Control

We have said that "Statistical Process *Control*" is really better called "Statistical Process *Monitoring*." "Engineering *Control*" is a very important subject that is largely distinct from the considerations laid out thus far in this chapter. Unfortunately, there has been a fair amount of confusion about what the two methodologies offer, how they differ, and what are their proper roles in the running of industrial processes. This section is intended to help readers better understand the relationship between them. It begins with an elementary introduction to one simple kind of engineering control, called PID control. It then proceeds to a number of general comments comparing and contrasting statistical process monitoring and engineering control.

3.6.1 Discrete Time PID Control

Engineering control has to do with guiding processes by the deliberate manipulation of appropriate process parameters. For example, in a chemical process, a temperature in a reaction vessel might be kept constant by appropriate manipulation of the position of an inlet steam valve. A very common version of engineering control in industry can be represented in terms of a feedback control diagram like that in Figure 3.19.

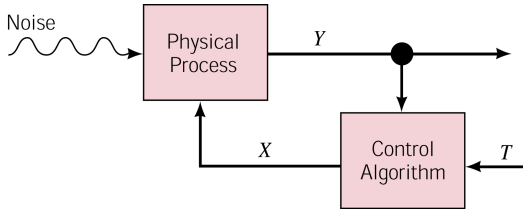


FIGURE 3.19. Schematic of an engineering feedback control system

In Figure 3.19, a process outputs a value of a variable Y, which is fed into a control algorithm along with a value of a target T for the next output, resulting in a value for some manipulated process variable X, which together with (unavoidable) noise (somehow) produces a subsequent value of Y, and so on. Depending upon what is known about the various elements in Figure 3.19, different means of choosing a control algorithm can be applied. A method that requires very little in the way of detailed knowledge about how X or the noise impact Y is that of **Proportional-Integral-Derivative (PID) control**.

The discussion here will treat the discrete time version of PID control. So consider discrete integer times $t = 1, 2, 3, \ldots$ (typically evenly spaced in real time) and as in

Figure 3.19, suppose that

$$\begin{aligned} Y(t) &= \text{the value of the controlled or output variable at time } t\text{ ,} \\ T(t) &= \text{the value of a target for } Y \text{ at time } t\text{, and} \\ X(t) &= \text{the value of a (manipulated) process variable that is} \\ &\quad \text{chosen after observing } Y(t)\text{ .} \end{aligned}$$

A control algorithm converts knowledge of $Y(1), Y(2), \ldots, Y(t)$ and $T(s)$ for all s into a choice of $X(t)$. For example, in machining $Y(t)$ could be a measured widget diameter, $T(t)$ a target diameter, and $X(t)$ a cutting tool position. A control algorithm orders a tool position in light of all past and present diameters and all targets for past, present, and future diameters.

The practice of PID control does not typically invest much effort in modeling exactly how changes in X get reflected in Y. (If the goal of a study *was* to understand that relationship, tools of regression analysis might well be helpful.) Nevertheless, in understanding the goals of engineering control, it is useful to consider two kinds of process behavior with which engineering control algorithms must sometimes deal.

For one thing, some physical processes react to changes in manipulated variables only gradually. One behavior predicted by many models of physical science is that when initially at "steady state" at time t_0, a change of ΔX in a manipulated variable introduces a change in the output at time $t > t_0$ of the form

$$\Delta Y(t) = Y(t) - Y(t_0) = G\Delta X\left(1 - \exp\left(\frac{-(t-t_0)}{\tau}\right)\right), \qquad (3.51)$$

for process-dependent constants G and τ. Figure 3.20 shows a plot of ΔY in display (3.51) as a function of time. In cases where relationship (3.51) holds, G is the limit of the ratio $\Delta Y/\Delta X$ and is called the **control gain**. τ governs how quickly the limiting change in Y is reached (τ is the time required to reach a fraction $1 - e^{-1} \approx .63$ of the limiting change in Y). It is called the **time constant** for a system obeying relationship (3.51).

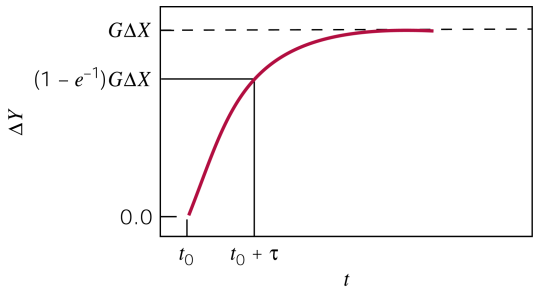


FIGURE 3.20. Change in the output Y (initially at steady state) in response to a ΔX change in the manipulated variable X

Another phenomenon that is sometimes part of the environment in which engineering control systems must operate is that of **dead time** or **delay** between when a change is made in X and when any effect of the change begins to be seen in Y. If there are δ units of dead time and thereafter a relationship similar to that in equation (3.51) holds, one might see a pattern like that shown in Figure 3.21 following an adjustment ΔX made at time t_0 on a process at steady state at that time.

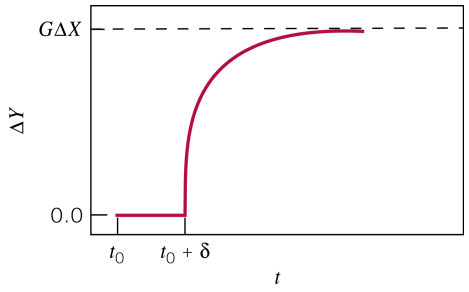


FIGURE 3.21. Change in the output Y (initially at steady state) in response to a ΔX change in the manipulated variable at time t_0 if there are δ units of dead time

Of course, not all physical systems involve the kind of gradual impact of process changes illustrated in Figure 3.20, nor do they necessarily involve dead time. (For example, real-time feedback control of machine tools will typically involve changes in tool positions that take their full effect "immediately" after being ordered.) But where these phenomena are present, they increase the difficulty of finding effective control algorithms, the dead time problem being particularly troublesome where δ is large.

To get to the point of introducing the general PID control algorithm, consider a situation where it is sensible to expect that increasing X will tend to increase Y. Define the observed "error" at time t,

Error at Time t

$$E(t) = T(t) - Y(t)\ ,$$

and the first and second differences of errors

First Difference in Errors at Time t

$$\Delta E(t) = E(t) - E(t-1)$$

and

$$\Delta^2 E(t) = \Delta(\Delta E(t)) = \Delta E(t) - \Delta E(t-1)\ .$$

Second Difference in Errors at Time t

With some additional algebra

$$\begin{aligned}\Delta^2 E(t) &= (E(t) - E(t-1)) - (E(t-1) - E(t-2)) \\ &= E(t) - 2E(t-1) + E(t-2)\ .\end{aligned}$$

Then, for constants κ_1, κ_2, and κ_3, a PID control algorithm sets

$$\Delta X(t) = \kappa_1 \Delta E(t) + \kappa_2 E(t) + \kappa_3 \Delta^2 E(t)\ . \tag{3.52}$$

PID Control Algorithm

(In cases where Y tends to increase with X, the constants κ_1, κ_2, and κ_3 are typically nonnegative.) The three terms summed on the right of equation (3.52) are respectively the **proportional, integral,** and **derivative** parts of the control algorithm.

Example 37 ***PID Control of Final Dry Weight of 20*** lb ***Bond Paper.*** *Through the kind cooperation of the Miami University Paper Science Laboratory and Mr. Doug Hart, Research Associate at the lab, one of your authors was able to help implement a PID controller on a* $13\,\mathrm{in}$ *Fourdrinier paper-making machine. This machine produces paper in a long continuous sheet beginning with vats of pulp mix. The final dry weight of paper is measured as the paper leaves the machine and can be controlled by the rate at which a Masterflex peristaltic pump delivers pulp mix to the machine. A manual knob is used to vary the pump speed and can be adjusted in "ticks." (Each 1-tick change corresponds approximately to a change of pump speed equal to .2% of its maximum capacity.) Past experience with the machine indicated that for* $20\,\mathrm{lb}$ *bond pulp mixture, a 1-tick increase in pump speed produces approximately a .3* g/m^2 *increase in paper dry weight. But unavoidable variations in the process (including the "thickness" of the mix available to the pump) produce variation in the paper dry weight and need to be compensated for by varying the pump speed.*

Since there is over a 4 min *lag between when a pump speed change is made and when paper affected by the speed change reaches the scanner that measures dry weight at the end of the machine, measurements and corresponding adjustments to pump speed were made only once every 5* min*. (This choice eliminates the effect of dead time on the control algorithm, which would be a concern if measurements and adjustments were made closer together.) Some experimentation with the machine led to the conclusion that a sensible PID control algorithm for the machine (using the 5-minute intervals and measuring the control variable changes in terms of ticks) has*

$$\kappa_1 = .83, \quad \kappa_2 = 1.66, \qquad \text{and} \qquad \kappa_3 = .83$$

TABLE 3.9. PID Control Calculations for the Control of Paper Dry Weight ($T, Y, E, \Delta E$ and $\Delta^2 E$ in g/m^2 and ΔX in ticks)

Period, t	$T(t)$	$Y(t)$	$E(t)$	$\Delta E(t)$	$\Delta^2 E(t)$	$\Delta X(t) = .83\Delta E(t) + 1.66E(t) + .83\Delta^2 E(t)$
1	70.0	65.0	5.0			
2	70.0	67.0	3.0	−2.0		
3	70.0	68.6	1.4	−1.6	.4	1.328
4	70.0	68.0	2.0	.6	2.2	5.644
5	70.0	67.8	2.2	.2	−.4	3.486
6	70.0	69.2	.8	−1.4	−1.6	−1.162
7	70.0	70.6	−.6	−1.4	0	−2.158
8	70.0	69.5	.5	1.1	2.5	3.818
9	70.0	70.3	−.3	−.8	−1.9	−2.739
10	70.0	70.7	−.7	−.4	.4	−1.162
11	70.0	70.1	−.1	.6	1.0	1.162

in formula (3.52). Table 3.9 shows an actual series of dry weight measurements and PID controller calculations made using these constants. (Since it was impossible to move the pump speed knob in fractions of a tick, the actual adjustments applied were those in the table rounded off to the nearest tick.) The production run was begun with the knob (X) *in the standard or default position for the production of 20* lb *bond paper.*

For example, for $t = 3$*,*

$$E(3) = T(3) - Y(3) = 70.0 - 68.6 = 1.4,$$

$$\Delta E(3) = E(3) - E(2) = 1.4 - 3.0 = -1.6,$$

$$\Delta^2 E(3) = \Delta E(3) - \Delta E(2) = -1.6 - (-2.0) = .4,$$

and so the indicated adjustment (increment) on the pump speed knob is $\Delta X(3) = .83\Delta E(3) + 1.66E(3) + .83\Delta^2 E(3) = .83(-1.6) + 1.66(1.4) + .83(.4) = 1.328$ *ticks. (As actually implemented, this led to a 1-tick increase in the knob position after measurement 3.)*

It is useful to separately consider the proportional, integral, and derivative parts of algorithm (3.52), beginning with the integral part. With $\kappa_2 > 0$, this part of the algorithm increases X when E is positive and thus $T > Y$. It is this part of a PID control algorithm that reacts to (attempts to cancel) **deviations from target**. Its function is to try to move Y in the direction of T.

To grasp why $\kappa_2 E(t)$ might be called the "integral" part of the control algorithm, consider a case where both $\kappa_1 = 0$ and $\kappa_3 = 0$ so that one has an "integral only" controller. In this case (supposing that $Y(t)$'s and $T(t)$'s with $t < 1$ are available so that one can begin using relationship (3.52) at time $t = 1$), note that

$$\sum_{s=1}^{t} \Delta X(s) = \kappa_2 \sum_{s=1}^{t} E(s). \tag{3.53}$$

But the sum on the left of equation (3.53) telescopes to $X(t) - X(0)$ so that one has

$$X(t) = X(0) + \kappa_2 \sum_{s=1}^{t} E(s).$$

That is, the value of the manipulated variable is $X(0)$ plus a sum or "integral" of the error.

"Integral only" control (especially in the presence of a large time constant and/or large dead time) often tends to overshoot target values and set up oscillations in the variable Y. The proportional and derivative parts of a PID algorithm are meant to reduce overshoot and damp oscillations. Consider next the proportional term from equation (3.52), namely $\kappa_1 \Delta E(t)$.

The proportional part of a PID control algorithm reacts to **changes in the error**. In graphical terms, it reacts to a nonzero slope on a plot of $E(t)$ versus t. Where $\kappa_1 > 0$, this part of the algorithm increases X if the error increases and decreases X if E decreases. In some sense, this part of the algorithm works to hold the error constant (whether at 0 or otherwise).

When κ_1 and κ_2 have the same sign, the proportional part of a PID control algorithm augments the integral part when E is moving away from 0 and "brakes" or cancels part of the integral part when E is moving toward 0. Figure 3.22 pictures two plots of $Y(t)$ versus t for cases where the target T is constant. In the first plot, Y is approaching T from below. $E(t) > 0$ while $\Delta E(t) < 0$. This is a case where the proportional part of the algorithm brakes the integral part. In the second plot, Y is above T and diverging from it. There, both $E(t) < 0$ and $\Delta E(t) < 0$, and the proportional part of the algorithm augments the integral part. The braking behavior of the proportional part of a PID algorithm helps to resist the kind of oscillation/overshoot problem produced by "integral only" control.

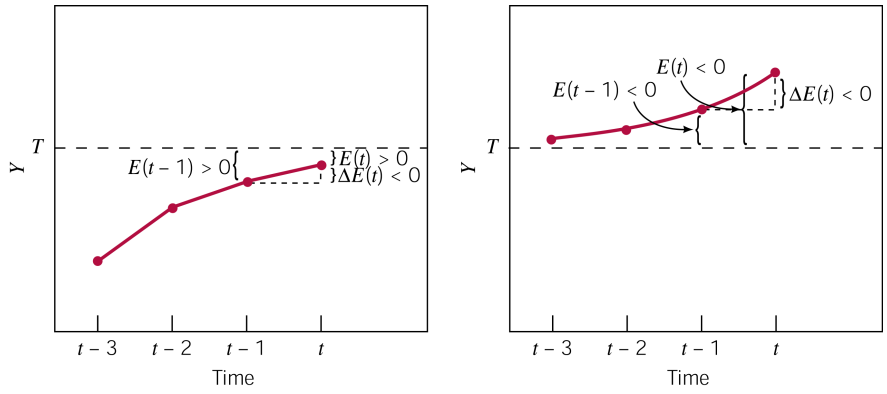


FIGURE 3.22. Two plots of Y against time

To see why $\kappa_1 \Delta E(t)$ might be called the "proportional" part of the control algorithm, consider a case where both $\kappa_2 = 0$ and $\kappa_3 = 0$ so that one has a "proportional

only" controller. In this case (supposing that $Y(t)$'s and $T(t)$'s with $t < 1$ are available so that one can begin using relationship (3.52) at time $t = 1$),

$$\sum_{s=1}^{t} \Delta X(s) = \kappa_1 \sum_{s=1}^{t} \Delta E(s). \tag{3.54}$$

But the sums on both sides of equation (3.54) telescope, so that one has

$$X(t) = X(0) - \kappa_1 E(0) + \kappa_1 E(t).$$

That is, the value of the manipulated variable is $X(0) - \kappa_1 E(0)$ plus a term "proportional" to the error.

Finally, consider the derivative part of the algorithm (3.52), namely $\kappa_3 \Delta^2 E(t)$. This part of the algorithm reacts to curvature or **changes in slope on a plot of** $E(t)$ **versus** t. That is, it reacts to changes in $\Delta E(t)$. If a plot of errors versus t is linear ($\Delta E(t)$ is constant), this part of the algorithm does nothing to change X. If $\kappa_3 > 0$ and a plot of errors versus t is concave up, the derivative part of algorithm (3.52) will increase X (and thus Y, decreasing E), while if the plot is concave down it will decrease X. For constant target T, this will tend to "straighten out" a plot of $E(t)$ or $Y(t)$ versus t (presumably then allowing the proportional part of the algorithm to reduce the slope to 0 and the integral part to put the process on target). Once again, since "integral only" control often produces unwanted oscillations of Y about a target, and it is impossible to oscillate without local curvature in a plot of E or Y versus t, the derivative part of the algorithm can be considered as corrective to a deficiency in the naive "integral only" idea.

The rationale for calling $\kappa_3 \Delta^2 E(t)$ the "derivative" part of the PID algorithm (3.52) is similar to the arguments made about the other two parts. Namely, if κ_1 and κ_2 are both 0 (so that one has "derivative only" control),

$$\sum_{s=1}^{t} \Delta X(s) = \kappa_3 \sum_{s=1}^{t} \Delta^2 E(s). \tag{3.55}$$

Telescoping both sides of equation (3.55) one then has

$$X(t) = X(0) - \kappa_3 \Delta E(0) + \kappa_3 \Delta E(t),$$

and the value of the manipulated variable is $X(0) - \kappa_3 \Delta E(0)$ plus a term proportional to the change in (or "derivative" of) the error.

The primary practical problem associated with the use of PID controllers is the matter of choosing the constants κ_1, κ_2, and κ_3, sometimes called respectively the **proportional, integral, and derivative gains** for the control algorithm. In simple situations where engineers have good mathematical models for the physical system involved, those can sometimes provide at least starting values for searches to find good values of these constants. Where such models are lacking, various rules of thumb aid searches for workable values of κ_1, κ_2, and κ_3. For instance, one such rule is to initially set κ_1 and κ_3 to zero, increase κ_2 till oscillations occur, then halve that value of κ_2 and begin

searching over κ_1 and κ_3. And it is pretty clear that in systems where a relationship like (3.51) holds, the gains κ_1, κ_2, and κ_3 should be inversely proportional to G. Further, conventional wisdom also says that in systems where there is dead time $\delta > 0$, the control gains should decrease (exponentially?) in δ. (One should not be changing a manipulated variable wildly if there's to be a long delay before one gets to measure the impact of those changes and to begin to correct any unfortunate effects one sees.)

Ultimately, the matter of finding good values for the gains κ_1, κ_2, and κ_3 is typically a problem of empirical optimization. Section 6.2 of this book discusses some experimental strategies in process optimization. These can be applied to the problem of finding good constants κ_1, κ_2, and κ_3 in the following way. For given choices of the constants, one may run the process using the PID controller (3.52) for some number of periods, say m. Then a sensible figure of merit for that particular set of constants is the random variable

$$S = \frac{1}{m}\sum_{t=1}^{m}(E(t))^2 \ ,$$

the average squared error. The empirical optimization strategies of Section 6.2 may then be applied in an attempt to find a set of values for κ_1, κ_2, and κ_3 with minimum associated mean for S, μ_s. Chapter problems 38 through 44 describe how the average squared error idea was used to arrive at the control algorithm of Example 37.

3.6.2 Comparisons and Contrasts

The PID ideas just discussed are not the only ones used to produce engineering control algorithms. For example, where good models are available for both uncontrolled process behavior and for the impact of control actions on process outputs, mathematically optimal control algorithms (that need not be of the PID type) can sometimes be derived. And the introduction just given completely ignores real issues like the multivariate nature of most industrial applications. (The Y and X just considered are one-dimensional, while real process outputs and possible manipulated variables are often multidimensional.) But the foregoing brief discussion is intended only to give the reader enough of an idea of how engineering control operates to allow the following comments on the proper roles of engineering control and statistical process monitoring to make sense.

The relative merits of the two methodologies when applied in production contexts have been at times been hotly debated by their proponents. On some occasions, zealots on one side or the other of the debate have essentially claimed that their methods are universally applicable and those of the other side are either without merit or are simply a weak version of their own. The truth is that the methods of statistical process monitoring and engineering control are not competitors. They are in fact, completely complementary, each having its own purposes and appropriate areas of application. When applied to the running of industrial processes, both are aimed at the reduction of unwanted variability. In many applications, they can and should be used *together* in an effort to reduce process variation and improve quality, engineering control helping to create stable conditions that are monitored using statistical process monitoring

methods.

In cases where a process is already physically stable about a target value, statistical process monitoring tools should only infrequently (and wrongly) signal the need for intervention, and engineering control *is of no help in reducing variation*. That is, in the classical stable process situation, tweaking process parameters can only make variation worse, not better. On the other hand, if successive observations on a process look as if they are dependent, or if they have means (either constant or moving) different from a target, engineering control may be able to improve process performance (uniformity of output) essentially by canceling predictable misadjustments of the process. Statistical process monitoring will then protect one from unexpected process changes.

Table 3.10 puts side by side a number of pairs of statements that should help the reader keep clear the basic differences between engineering control and statistical process monitoring as they are applied to industrial processes. The late Dr. Bill Tucker was fond of saying "You can't steer a car with statistical process control and you can't fix a car with engineering control." His apt and easily remembered analogy brings into focus the differences in intent of the two methodologies.

Section 3.6 Exercises

1. The target value for a process output variable, Y, is 4 units, and a controllable process parameter X is thought to impact Y in a direct fashion. In 3 successive periods $Y(1) = 2, Y(2) = 1$, and $Y(3) = 0$. You may finish filling in a table like that below to help you answer this question.

Period, t	$T(t)$	$Y(t)$	$E(t)$	$\Delta E(t)$	$\Delta^2 E(t)$	$\Delta X(t)$
1		2				
2		1				
3		0				

   (a) What are your values of $T(t)$ here? What is the practical meaning of this variable?

   (b) What values do you get for $E(t), \Delta E(t)$ and $\Delta^2 E(t)$ here? Describe what these are measuring.

   (c) Use control gains $\kappa_1 = .8$, $\kappa_2 = 1.6$, and $\kappa_3 = 1.9$ and compute a PID control action $\Delta X(3)$ to be taken after observing $Y(3)$.

   (d) How will this table and future values of Y be used to extend the PID control of part (c) beyond period 3?

2. In the context of problem 1, suppose that no one is really sure whether Y is affected by changes in X, or if it is, whether the relationship is "direct" or "inverse."

   (a) Speculate on what might happen if the PID controller of part (c) above is implemented where Y is completely unrelated to X. What might happen if in fact Y is inversely related to X?

TABLE 3.10. Contrasts Between Engineering Control and Statistical Process Control for Industrial Processes

Engineering Control	Statistical Process Control
• In a typical application, there is a sensor on a process and an electromechanical adjustment mechanism that responds to orders (for change of some process parameter) sent by a computer "brain" based on signals from the sensor.	• This is either manual or automatic plotting of process performance statistics to warn of process changes.
• This is an adjustment/compensation methodology. Formulas prescribe explicit reactions to deviations from target.	• This is a detection methodology. Corrective measures for process changes that are detected are not specified.
• This is a methodology for ongoing small process adjustments.	• There is a tacit assumption here that wise intervention following detection of a process change will set things perfectly aright (for an extended period).
• There is an explicit expectation of process instability/drift in this methodology.	• There is a tacit assumption here of process stability over long periods.
• This is typically computer (or at least mechanically) controlled.	• There is typically a human agent involved in monitoring and interventions.
• The ultimate effect is to keep a process optimally adjusted.	• The ultimate effect is to warn of the presence of sources of special cause variation, to help identify them, and to lead to their permanent removal.
• This is often "tactical" and applied to process parameters.	• This is often "strategic" and applied to final quality variables.
• In its "optimal stochastic control" version, this is what one does within a particular probability model (for process behavior) to best exploit the probabilistic predictability of a process.	• This is what one does to monitor for "the unexpected" (departures from a stable process model of expected behavior).

(b) How would you propose to figure out what, if any, PID control based on X might be fruitful?

3. In what sense are control charts tools for "controlling" a process? In what meaning of the word "control" are they *not* tools for controlling a process?

3.7 Chapter Summary

Shewhart control charts are an engineer's most widely applicable and easily understood process-monitoring tools. The first four sections of this chapter have introduced these charts for both variables data and attributes data, considered their use in both standards given and retrospective contexts, and discussed their qualitative interpretation and supplementation with sets of "extra alarm rules." Table 3.11 summarizes many of the standard formulas used in the making of elementary Shewhart charts.

TABLE 3.11. Formulas for Shewhart Control Charting

				Standards Given		Retrospective	
Chart	Q	μ_Q	σ_Q	UCL_Q	LCL_Q	UCL_Q	LCL_Q
$\overline{x}$	$\overline{x}$	μ	$\sigma/\sqrt{n}$	$\mu+3\sigma/\sqrt{n}$	$\mu-3\sigma/\sqrt{n}$	$\overline{\overline{x}}+A_2\overline{R}$ $\overline{\overline{x}}+A_3\overline{s}$	$\overline{\overline{x}}-A_2\overline{R}$ $\overline{\overline{x}}-A_3\overline{s}$
$Median$	$\tilde{x}$	μ	$\kappa\sigma/\sqrt{n}$	$\mu+3\kappa\sigma/\sqrt{n}$	$\mu-3\kappa\sigma/\sqrt{n}$		
R	R	$d_2\sigma$	$d_3\sigma$	$D_2\sigma$	$D_1\sigma$	$D_4\overline{R}$	$D_3\overline{R}$
s	s	$c_4\sigma$	$c_5\sigma$	$B_6\sigma$	$B_5\sigma$	$B_4\overline{s}$	$B_3\overline{s}$
np	X	np	$\sqrt{np(1-p)}$	$np+3\sqrt{np(1-p)}$	$np-3\sqrt{np(1-p)}$	(use $\hat{p}_{\text{pooled}}$ for p)	
p	$\hat{p}$	p	$\sqrt{\frac{p(1-p)}{n}}$	$p+3\sqrt{\frac{p(1-p)}{n}}$	$p-3\sqrt{\frac{p(1-p)}{n}}$	(use $\hat{p}_{\text{pooled}}$ for p)	
c	X	λ	$\sqrt{\lambda}$	$\lambda+3\sqrt{\lambda}$	$\lambda-3\sqrt{\lambda}$	(use $\hat{\lambda}_{\text{pooled}}$ for λ)	
u	$\hat{u}$	λ	$\sqrt{\frac{\lambda}{k}}$	$\lambda+3\sqrt{\frac{\lambda}{k}}$	$\lambda-3\sqrt{\frac{\lambda}{k}}$	(use $\hat{\lambda}_{\text{pooled}}$ for λ)	

The final two sections of the chapter have provided context and perspective for the study of Shewhart charts and other process-monitoring tools. Section 3.5 introduced the ARL concept as a means of quantifying the likely performance of a monitoring scheme. Section 3.6 contrasted methods and goals of "engineering control" with those of process monitoring when they are both applied in production.

3.8 Chapter 3 Exercises

1. What is the purpose of control charting? What is suggested by out-of-control signals?

2. What makes step 3 in the quality assurance cycle presented in Chapter 1 difficult in service contexts? Explain.

3. Why is it essential to have a clear understanding of what constitutes a nonconformance if a Shewhart c or u chart is to be made?

4. Is control charting most directly concerned with "quality of design" or with "quality of conformance" ?

5. Distinguish between "control limits" and "specification limits" for variables data.

6. Explain the difference between "control limits" and "specification limits" in an attributes data setting.

7. Explain the ARL concept in terms that a person with no statistical training could understand.

8. When designing a control chart, what kinds of ARL values are desirable for an on-target process? For an off-target process? Explain why your answers are correct from an economic point of view.

9. State why statistical methodology is an unavoidable part of quality assurance practice. (Review Chapter 1.)

10. Sometimes the plotted statistics appearing on a Shewhart control chart hug (or have little scatter around) a center line. Explain why this is not necessarily a good sign.

11. Uninformed engineers sometimes draw in lines on Shewhart $\overline{x}$ charts at engineering specifications for individual measurements. Why is that bad practice?

12. It is common to hear people imply that the job of control charts is to warn of degradation in product quality. Do you agree with that? Why or why not?

13. What is the purpose of sets of "extra alarm rules" like the Western Electric rules presented in Section 3.4?

14. What (relevant to quality improvement efforts) does a multimodal shape of a histogram for a part dimension suggest? (Review Chapter 2.)

15. In colloquial terms, the language "control" chart perhaps suggests a plot associated with continuous regulatory efforts. Is understanding correct? Why or why not? Suggest a better term than "control chart."

16. **Journal Diameters**. Below are some summary statistics (means and standard deviations) for journal diameters of tractor axles as the axles come off an automatic grinding machine. The statistics are based on subgroups of size $n = 4$ pieces taken once per hour. The values listed are in millimeters. Specifications on the journal diameter are from $44.975\,\mathrm{mm}$ to $44.990\,\mathrm{mm}$.

Subgroup	$\overline{x}$	s	Subgroup	$\overline{x}$	s
1	44.9875	.0029	**11**	44.9815	.0017
2	44.9813	.0025	**12**	44.9815	.0017
3	44.9808	.0030	**13**	44.9810	.0024
4	44.9750	.0000	**14**	44.9778	.0021
5	44.9783	.0039	**15**	44.9748	.0024
6	44.9795	.0033	**16**	44.9725	.0029
7	44.9828	.0021	**17**	44.9778	.0021
8	44.9820	.0024	**18**	44.9790	.0034
9	44.9770	.0024	**19**	44.9785	.0010
10	44.9795	.0010	**20**	44.9795	.0010

Note that $\sum \overline{x} = 899.5876$ and $\sum s = .0442$.

(a) Are the above attributes data or variables data? Why?

(b) Make a retrospective s chart for these values.

(c) Make a retrospective $\overline{x}$ chart for these values.

(d) What do these charts indicate (in retrospect) about the stability of the grinding process?

(e) Based on your conclusion in (d), can the fraction of journal diameters that currently meet specifications be reliably estimated? Why or why not?

(f) Independent of your conclusion in (d), if one judged the process to be stable based on the 20 subgroups summarized above, what could be used as an estimate of the fraction of journal diameters that currently meet specifications? (Give a number based on a normal distribution assumption for diameter measurements.)

(g) Suppose that henceforth (into the future) this process is to be monitored using subgroups of size $n = 5$. Give control limits for a (standards given) median chart based on the mid-specification (giving the center line) and your estimated process standard deviation from (b).

(h) Give control limits for future monitoring of sample ranges (use your estimated process standard deviation from (b) as a future standard value and assume $n = 5$).

17. Refer to the **Journal Diameter** case introduced in problem 16. Sometimes subgroup size is not constant. When using standard deviations from subgroups of varying sizes $n_1, n_2, ..., n_r$ to estimate σ, there are several possibilities. Of commonly used ones, the one with the best theoretical properties is

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_r-1)s_r^2}{(n_1-1) + (n_2-1) + \cdots + (n_r-1)}}.$$

Another possibility is

$$\hat{\sigma} = \frac{\dfrac{(n_1-1)s_1}{c_4(n_1)} + \dfrac{(n_2-1)s_2}{c_4(n_2)} + \cdots + \dfrac{(n_r-1)s_r}{c_4(n_r)}}{n_1+n_2+\cdots+n_r-r}.$$

(Since sample sizes vary in this development, we are displaying the dependence of c_4 on sample size here.) The most appropriate estimator of a common mean, μ, when sample sizes vary is

$$\overline{x}_{\text{pooled}} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + \cdots + n_r\overline{x}_r}{n_1+n_2+\cdots+n_r}.$$

Consider the subgroup means and standard deviations given in problem 16. Suppose subgroups were of size $n = 4$ except for the ones indicated in the following table.

Subgroup	Sample Size
1	2
8	2
10	8
15	5
18	3
19	3
20	9

(a) Find values for s_{pooled}, $\hat{\sigma}$, and $\overline{x}_{\text{pooled}}$.

(b) Give two estimates of 1) the standard deviation of a subgroup mean when $n = 2$ and 2) the standard deviation of a subgroup standard deviation when $n = 2$. (Hint: $\text{Var}\,\overline{x}_i = \sigma^2/n_i$ and $\text{Var}\,s_i = \sigma^2(1 - c_4^2(n_i))$.)

(c) With the new subgroup sizes, consider two retrospective control charts, one chart appropriate for assessing the constancy of variability of axle journal diameters and the other for monitoring average axle journal diameter. Would the control limits be constant across time for the two charts? (There is no need to actually make them here.) Why or why not? (See (a) and (b).)

(d) Do the center lines for the two charts in (c) change depending on subgroup size? (Again, there is no need to make the charts.) Why or why not?

18. **Rolled Paper.** Shervheim and Snider did a project with a company that cuts rolled paper into sheets. The students periodically sampled $n = 5$ consecutive sheets as they were cut and recorded their actual lengths, y. Data from 20 subgroups are summarized below. (Measurements corresponding to the values in the table were in 64ths of an inch above nominal, i.e., $x = y - nominal$.)

(a) Make a retrospective s chart.

(b) Make a retrospective $\overline{x}$ chart.

(c) What do these charts indicate (in retrospect) about the stability of the cutting process?

(d) Give an estimate of the process standard deviation based on $\overline{s}$.

(e) If one judges the process to be stable and sheet length to be normally distributed, estimate the fraction of sheets below nominal in length. (Hint: Find $P(x < 0)$ by transforming to a standard normal random variable Z.)

Subgroup	$\overline{x}$	s
1	12.2	.84
2	11.2	1.64
3	10.6	2.07
4	12.2	2.49
5	11.2	.84
6	12.6	1.82
7	12.2	2.95
8	13.6	1.67
9	12.2	1.30
10	10.4	1.52
11	10.4	1.95
12	10.6	1.67
13	10.4	1.67
14	12.0	2.91
15	11.2	.84
16	10.6	1.82
17	10.4	1.14
18	9.8	2.17
19	9.6	2.07
20	10.6	1.95
	224.0	35.33

(f) Each $.25\,\text{in}$ that the cutting process mean is above nominal represents a \$100,000/year loss to the company from product "given away." On the other hand, the company wants to be sure that essentially no sheets are produced with below-nominal lengths (so they want $\mu_x > 3\sigma$). With this in mind, what adjustment in mean length do you suggest, and what yearly savings or additional cost do you project if this adjustment is made?

(g) Suppose that the adjustment you recommend in (f) is made and henceforth the cutting process is to be monitored based on samples of size $n = 3$. What are standards given control limits for future monitoring of $\overline{x}$ and s?

(h) Suppose that while using your $\overline{x}$ chart from (g) the process mean suddenly drops to the point where 1% of the sheets produced are below nominal in length. On average, how many samples will be required to detect this? (Hint: find the "new μ_x" that will make $P(x < 0) = .01$, then using it find $P(\overline{x} < LCL) + P(\overline{x} > UCL)$.) How does this compare in terms of

quickness of detection to a scheme (essentially a p chart) that signals the first time a sample of $n = 3$ contains at least one sheet with below-nominal length?

19. Refer to the **Rolled Paper** case in problem 18. Again use the means and standard deviations given there, but suppose that the number of sheets per subgroup was not constant. Instead, suppose subgroups contained 5 sheets except for the ones indicated in the following table.

Subgroup	Subgroup Size
3	7
6	7
10	2
14	4
17	3
19	2
20	6

(a) Compute $\overline{x}_{\text{pooled}}$ and two different estimates of σ. (See problem 17.)

(b) For a subgroup size of $n = 7$, give two estimates of 1) the standard deviation of a subgroup mean and 2) the standard deviation of a subgroup standard deviation. (Hint: $\text{Var}\,\overline{x}_i = \sigma^2/n_i$ and $\text{Var}\,s_i = \sigma^2(1 - c_4^2(n_i))$.)

(c) With the variable subgroup sizes, consider two retrospective control charts, one s chart and one $\overline{x}$ chart. Would the control limits be constant across time for either chart? (There is need to make the charts.) Why or why not? (See (a) and (b).)

(d) Do the center lines for the two charts in (c) remain across subgroup sizes? (Again, there is no need to make the charts.) Why or why not?

20. **U-bolt Threads.** A manufacturer of U-bolts for the auto industry measures and records thread lengths on bolts that it produces. Eighteen subgroups, each of $n = 5$ consecutive bolts, were obtained and actual thread lengths y were measured. These can be expressed as deviations from nominal by transforming $x = y - nominal$. Some summary statistics are indicated below (the units are $.001\,\text{in}$ above nominal).

(a) Estimate the supposedly common subgroup standard deviation, σ, using 1) the subgroup ranges (R_i) and 2) the subgroup standard deviations (s_i).

(b) Find control limits for the subgroup ranges. (Use the estimate of σ based on the s_i.)

(c) Find control limits for the subgroup standard deviations. (Use the estimate of σ based on the s_i.)

(d) Plot the ranges and standard deviations on Shewhart charts using the retrospective limits from (b) and (c). Is it plausible that variability of thread

length was constant from sampling period to sampling period? Why or why not?

(e) Find retrospective control limits for the subgroup means. (Use your estimate of σ based on the s_i.) Plot the means on a Shewhart chart with these limits.

Subgroup	Thread Length	$\tilde{x}$	s	$\overline{x}$	R
1	$11, 14, 14, 10, 8$	11	2.61	11.4	6
2	$14, 10, 11, 10, 11$	11	1.64	11.2	4
3	$8, 13, 14, 13, 10$	13	2.51	11.6	6
4	$11, 8, 13, 11, 13$	11	2.05	11.2	5
5	$13, 10, 11, 11, 11$	11	1.10	11.2	3
6	$11, 10, 10, 11, 13$	11	1.22	11.0	3
7	$8, 6, 11, 11, 11$	11	2.30	9.4	5
8	$10, 11, 10, 14, 10$	10	1.73	11.0	4
9	$11, 8, 11, 8, 10$	10	1.52	9.6	3
10	$6, 6, 11, 13, 11$	11	3.21	9.4	7
11	$11, 14, 13, 8, 11$	11	2.30	11.4	6
12	$8, 11, 10, 11, 14$	11	2.17	10.8	6
13	$11, 11, 13, 8, 13$	11	2.05	11.2	5
14	$11, 8, 11, 11, 11$	11	1.34	10.4	3
15	$11, 11, 13, 11, 11$	11	.89	11.4	2
16	$14, 13, 13, 13, 14$	13	.55	13.4	1
17	$14, 13, 14, 13, 11$	13	1.22	13.0	3
18	$13, 11, 11, 11, 13$	11	1.10	11.8	2
		202	31.51	200.4	74

(f) Setting the center line at $\overline{\overline{x}}$, find upper and lower control limits for the subgroup medians. (Use your estimate of σ based on the s_i.) Plot the medians on a Shewhart chart with these limits.

(g) What do the charts in (e) and (f) suggest about the threading process?

(h) A U-bolt customer requires that essentially all U-bolt thread lengths are within $.011\,\text{in}$ of nominal. Assuming bolt manufacturing continues as represented by the values in the table, will the customer be satisfied with current production? Why or why not? Give a quantitative defense of your answer assuming normality of thread length. (Hint: Find $P(-11 < x < 11)$.)

21. Refer to the **U-bolt Threads** case in problem 20. Problem 17 presented ways of estimating σ when r subgroups are of varying size n_i. The formulas there are based on subgroup sample standard deviations s_i. Another expression sometimes used to estimate the process standard deviation is based on ranges, namely

$$\frac{\dfrac{(n_1-1)R_1}{d_2(n_1)} + \dfrac{(n_2-1)R_2}{d_2(n_2)} + \cdots + \dfrac{(n_r-1)R_r}{d_2(n_r)}}{n_1 + n_2 + \cdots + n_r - r}.$$

Consider the subgroup means and ranges given in problem 20 and suppose that subgroups consisted of $n = 5$ bolts except for the subgroups indicated in the following table:

Subgroup	Subgroup Size
2	8
5	4
6	6
7	2
11	3
14	7
15	2
18	2

(a) Give $\overline{x}_{\text{pooled}}$ and three estimates of σ. Base two of the estimates of σ on the subgroup standard deviations and the other on the ranges.

(b) Find three estimates of the standard deviation of a subgroup mean when $n = 8$. Base two of the estimates on subgroup standard deviations and one on the ranges. (Hint: $\text{Var}\,\overline{x}_i = \sigma^2/n_i$.)

(c) Find three estimates of the standard deviation of each subgroup sample standard deviation when $n = 8$. Base two of the estimates on subgroup standard deviations and one on the ranges. (Hint: $\text{Var}\,s_i = \sigma^2(1-c_4^2(n_i))$.)

(d) Find an estimate of the standard deviation of each subgroup range when $n = 8$. Base the estimate on the subgroup ranges. (Hint: $\text{Var}\,R_i = d_3^2(n_i)\sigma^2$.)

(e) Consider retrospective $\overline{x}$ and R charts using the new configuration of subgroup sizes. (There is no need to make the charts here.) Would control limits for either chart be constant across time? Why or why not?

(f) Are the center lines for the charts referred to in (e) constant across subgroups? Why or why not?

22. **Turning**. Allan, Robbins, and Wycoff worked with a machine shop that employs a CNC (computer numerically controlled) lathe in the machining of a part for a heavy equipment manufacturer. Some summary statistics for a particular part diameter (x) obtained from 25 subgroups of $n = 4$ parts turned on the lathe are given below. The units are inches.

(a) Find retrospective control limits for the values (both means and ranges). What do the $\overline{x}$ and R values indicate about the stability of the turning process?

(b) Suppose that one wishes to apply the four Western Electric alarm rules to the $\overline{x}$ values. Specify the different zones to be used for the mean diameters. Are any of the rules violated in the first 10 samples? (If you find any violations, say which rule is violated for the first time where.)

Subgroup	$\overline{x}$	R
1	1.18093	.0001
2	1.18085	.0002
3	1.18095	.0002
4	1.18063	.0008
5	1.18053	.0007
6	1.18053	.0005
7	1.18058	.0005
8	1.18195	.0001
9	1.18100	.0003
10	1.18095	.0001
11	1.18095	.0006
12	1.18098	.0001
13	1.18123	.0009
14	1.18128	.0002
15	1.18145	.0007
16	1.18080	.0003
17	1.18100	.0000
18	1.18103	.0001
19	1.18088	.0003
20	1.18100	.0000
21	1.18108	.0002
22	1.18120	.0004
23	1.18088	.0002
24	1.18055	.0022
25	1.18100	.0004
	29.52421	.0101

(c) Give an estimate of the process short-term standard deviation derived from the ranges (use all 25 subgroups) and the assumption that σ is constant over the study period.

(d) Engineering specifications on the diameter in question were in fact $1.1809 \pm .005\,\text{in.}$ Suppose that over short production runs, diameters can be described as normally distributed and that your estimate of σ from (c) is an appropriate description of the variation seen in short runs. Give an estimate of the best possible fraction of diameters meeting specifications available using this particular lathe.

(e) Make further use of your estimate of σ from (c), and set up control limits that could be used in the future monitoring of the process standard deviation via Shewhart charting of s based on samples of size $n = 5$.

(f) Again use your estimate of σ from (c) and consider future monitoring of $\overline{x}$ based on samples of size $n = 4$ using "3 sigma" limits and a center line at the target diameter, 1.1809. Assuming diameters are normally distributed,

on average how many subgroups would be required to detect a change in mean diameter, μ, from 1.1809 to 1.1810?

23. Refer to the **Turning** case in problem 22. Problem 21 presented a method for estimating σ based on ranges of subgroups of varying size. Use that method in this problem. Use again the subgroup means and ranges given in problem 22 and suppose all subgroups were of size $n = 4$ parts except for the ones indicated in the following table.

Subgroup	Subgroup Size
1	2
4	3
5	6
9	2
11	7
13	5
15	3
16	2
17	8
18	2

(a) Give $\overline{x}_{\text{pooled}}$ and an estimate of σ.

(b) Find an estimate of the standard deviation for a subgroup mean, $\overline{x}_i$, when $n = 7$.

(c) Find an estimate of the standard deviation for a subgroup range, R_i, when $n = 7$.

(d) Consider retrospective $\overline{x}$ and R charts using the new configuration of subgroup sizes. (There is no need to make the chart here.) Are the control limits for the two charts constant across subgroups? Why or why not?

(e) Are the center lines for the charts considered in (d) constant across subgroups? Why or why not?

24. **Package Sorting**. Budworth, Heimbuch, and Kennedy analyzed a company's package sorting system. As packages arrive at the sorting system, they are placed onto trays and the bar codes affixed to the packages are scanned (in an operation much like the scanning process at a grocery store checkout). Bar code identification numbers begin with the zip code of the package destination. This permits packages to be sorted into 40 bins, each of which represents a different bulk mail center (BMC) or auxiliary service facility (ASF). All packages in a given bin are shipped by truck to the same mail center. The bulk transportation of these packages is much cheaper than if they were mailed directly by the nearest U.S. Post Office. The large number of BMC packages handled daily by the company produces tremendous cost savings.

Initially, the team tackled the so-called "no chute open" problem. When one of the BMC bins is full and packages destined for that bin cannot be dropped into

it. They end up in a "no chute open" bin. This eventuality produced many inefficiencies and even shutdowns of the entire system. In fact, the system was shut down about $10\,\mathrm{min}$/day on average because of this problem. This lost time cost the company the ability to process about 400 packages/day, and accumulated over a year, this represents a serious loss. The team decided to document the number of packages per shift dumped in the "no chute open" bin. The data they collected are below.

Date	Shift	Number in "No Chute Open" Bin
10/16	1	1510
10/17	3	622
10/18	1	2132
10/18	2	1549
10/19	1	1203
10/19	2	2752
10/19	3	1531
10/20	1	1314
10/20	2	2061
10/20	3	981
10/21	1	1636
10/21	2	2559
10/21	3	1212
10/22	1	2016
10/22	2	2765
10/22	3	574

(a) Is this an attributes data problem or a variables data problem? Why?

(b) What constitutes a "subgroup" in the context of this problem?

(c) What probability model is a possible description of the number of packages routed to the "no chute open" bin during a given shift?

(d) Assuming the sorting process is stable, estimate the average number of packages routed to the "no chute open" bin during a particular shift. Estimate the standard deviation of the number of packages routed to the "no chute open" bin. These estimates should be consistent with your answer to (c).

(e) Was the number of packages in the "no chute open" bin apparently constant except for random fluctuation? Why or why not? Defend your answer using a control chart.

25. Refer to the **Package Sorting** case in problem 24. Budworth, Heimbuch, and Kennedy were told that the sorting system was set up to let a package circle on the conveyor belt for 10 cycles (once each cycle the package would fall into the correct chute if that chute was not occupied). If after 10 cycles the correct chute

was always occupied, a package would be consigned to the inefficient "no chute open" bin. Upon observing the system in operation, the team immediately recognized packages dropping into the "no chute open" bin after only 2 or 3 cycles. Management was notified and the programming of the system was corrected. The team gathered the data below after the correction.

Date	Shift	Number in "No Chute Open "Bin
10/23	1	124
10/24	3	550
10/25	1	0
10/25	2	68
10/25	3	543
10/26	1	383
10/26	2	82
10/26	3	118

(a) Extend the control limits from your chart in part (e) of problem 24. Plot the data above on the same chart. Does it appear the system change was effective? Why or why not?

(b) Make a chart to assess stability of the number of packages in the "no chute open" bin using only the data above. Does it appear the system was stable? Why or why not?

(c) Has the team solved the problem of a large number of packages in the "no chute open" bin? Defend your answer.

26. Refer to the **Package Sorting** case of problems 24 and 25. Budworth, Heimbuch, and Kennedy also investigated the performance of the package scanning equipment. Just as items at a cashier's scanner often are not read on the first scan, so too were bar codes on packages not necessarily read on the first or second scan. Label damage and incorrect orientation, erroneous codes, and some simply unexplained failures all produced "no read" packages. If a package was not read on the first pass, it continued on the carousel until reaching a second scanner at the end of the carousel. Failure to read at this second scanner resulted in the package being dropped into a "no read" bin and scanned manually with a substantial loss in efficiency. The team took data over 30 consecutive one-minute periods on the variables

n = the number of packages entering the system during the $1\ \mathrm{min}$ period ,
X_1 = the number of those packages failing the first scan, and
X_2 = the number of those packages failing both scans .

The values they recorded follow.

Minute	n	X_1	X_2	Minute	n	X_1	X_2
1	54	10	2	**16**	66	17	0
2	10	3	2	**17**	56	11	3
3	55	22	3	**18**	26	6	1
4	60	18	5	**19**	30	6	0
5	60	12	1	**20**	69	14	1
6	60	14	1	**21**	58	23	5
7	37	14	0	**22**	51	18	5
8	42	17	1	**23**	32	15	1
9	38	20	10	**24**	44	23	4
10	33	6	2	**25**	39	13	2
11	24	6	3	**26**	26	3	1
12	26	7	5	**27**	41	17	1
13	36	12	0	**28**	51	25	5
14	32	10	3	**29**	46	18	1
15	83	25	2	**30**	59	23	6

(a) What constitutes a "subgroup" in this problem?

(b) Is this an attributes data or a variables data problem? Why?

(c) Make a retrospective control chart to assess consistency of the proportion of packages failing both scans and comment on what it indicates.

(d) Make a retrospective control chart to assess consistency of the proportion of packages that are not read on the first scan and comment on what it indicates.

(e) Make a retrospective control chart to assess consistency of the proportion of all packages in a given minute that are not read on the first scan and are read on the second scan. Comment on what it indicates.

(f) Calculate the proportions of those packages failing the first scan that also fail the second scan.

(g) Make a retrospective control chart to assess consistency of the proportions in (f). Comment on what it indicates.

27. **Jet Engine Visual Inspection**. The data below are representative of counts of nonconformances observed at final assembly at an aircraft engine company. Suppose that one final assembly is inspected per day.

(a) Is this a variables data problem or is it an attributes data problem? Explain.

(b) In the context of the problem, what is a "subgroup"?

(c) What probability distribution is a likely model for counts of nonconformances on these engines? Defend your answer.

(d) Find an estimated mean number of visually identified nonconformances and the corresponding estimated standard deviation.

Day	Number of Nonconformances	Day	Number of Nonconformances	Day	Number of Nonconformances
7/5	15	7/15	18	7/29	16
7/6	19	7/16	4	8/1	30
7/7	12	7/19	16	8/2	34
7/8	24	7/20	24	8/3	30
7/9	18	7/21	16	8/4	40
7/10	10	7/22	12	8/5	30
7/11	16	7/25	0	8/6	36
7/12	26	7/26	16	8/8	32
7/13	16	7/27	26	8/9	42
7/14	12	7/28	12	8/10	34

(e) Find appropriate upper and lower control limits and center line to apply to the counts. Make the corresponding control chart for these data. Does it appear that the process was stable over the period of the study? Why or why not? Identify any out-of-control points. Apply Nelson's rules.

(f) Suppose two inspectors were involved in the data collection. Briefly discuss what must be true (in terms of data collection protocol) to assure that the chart and analysis in (e) are credible.

28. Refer to the **Jet Engine Visual Inspection** case in problem 27.

(a) When possible causes for out-of-control points on a control chart are addressed and physically eliminated, it is common practice to discard the data associated with those out-of-control points and recalculate control limits. Apply this thinking to part (e) of problem 27, assuming causes of the out-of-control points have been addressed (you should "throw out" July 16, 25 and August 1 through 10—a total of 11 out of 30 points).

(b) Suppose the following data are obtained in visual inspection of final engine assemblies over the next three days.

Day	Assemblies Inspected	Number of Nonconformances
1	.5	8
2	2.0	31
3	1.5	26

(Partial inspection of final engine assemblies could possibly occur because of unforeseen labor problems. More than one engine assembly might be inspected on days 2 and 3 to, in some sense, make up for the partial inspection on day 1.) Using the information from (a) above, find control limits for nonconformance rates on these three days (don't use the number of nonconformances during these 3 new days to find the limits). Also give the center line and three plotted values (nonconformances per engine assembly inspected).

(c) Do your values from part (b) suggest process instability? Explain.

(d) Your center line should be constant across the three days represented in (b). Why is this?

29. The number of standard units inspected may vary from period to period. Let

$$\begin{aligned} X_i &= \text{the number of nonconformances observed at period } i, \\ k_i &= \text{the number of standard units inspected at period } i, \text{ and} \\ \hat{u}_i &= X_i/k_i \ . \end{aligned}$$

The following values were obtained over 9 periods.

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
k_i	1	2	1	3	2	1	1	3	1
$\hat{u}_i$	0	3.00	0	1.33	4.00	0	0	.67	1.00

(a) From these values, what conclusions can you make about stability of the process being monitored? Make the appropriate control chart.

(b) Suppose that in the future k_i will be held constant at 1 and that 2.4 nonconformances per inspection unit will be considered to be "standard quality." Find the probability of an out-of-control signal on a 3-sigma Shewhart control chart, if the true nonconformance rate is at the standard quality level ($\lambda = 2.4$). Find the probability of an out-of-control signal if the true nonconformance rate changes to $\lambda = \ 4.8$. (Remember that the Poisson(μ) probability function is $P(X = x) = (\exp(-\mu)\,\mu^x)\,/x!$.)

(c) Suppose that in the future k_i will be held constant at 2. Find the probability of an out-of-control signal if the true nonconformance rate is at the standard quality level ($\lambda = 2.4$). Find the probability of an out-of-control signal if the true nonconformance rate changes to $\lambda = 4.8$

(d) Compare your answers to (b) and (c). Which subgroup size ($k = 1$ or $k = 2$) is more appealing? Why?

30. **Electrical Switches.** The following scenario is taken from an aircraft engine company's training material. One hundred electrical switches are sampled from each of 25 consecutive lots. Each sampled switch is tested and the sample numbers failing are recorded below.

(a) Find the sample fractions of switches failing the test.

(b) What is a plausible probability model for describing the count of switches in a particular sample failing the test? Explain.

(c) Plot the number failing versus the sample period. Plot an appropriate center line and control limits on the same graph.

(d) What does your plot in (c) monitor here?

(e) Interpret your plot in (c). Identify any out-of-control points.

Sample	Number Failing	Sample	Number Failing
1	11	**14**	18
2	9	**15**	7
3	15	**16**	10
4	11	**17**	8
5	22	**18**	11
6	14	**19**	14
7	7	**20**	21
8	10	**21**	16
9	6	**22**	4
10	2	**23**	11
11	11	**24**	8
12	6	**25**	9
13	9		

(f) What is the usual name of the chart you prepared in part (c)?

(g) Suppose causes for out-of-control points identified in (e) are identified and physically removed. It would then make sense to delete the out-of-control points and recalculate limits. Do this recalculation and redo (c). You should have identified and eliminated 2 out-of-control points.

(h) Suppose the number of switches sampled and the number failing for the next three consecutive lots are as follows.

Number Sampled	Number Failing
75	8
144	12
90	11

Using your estimated fraction failing from (g) as a standard for judging these samples, find control limits and center lines appropriate for the three new "number failing" data points. Are the three sets of control limits and center lines the same? Why is this to be expected?

31. A data set in the book *Elementary Statistical Quality Control* by Burr indicates that in the magnaflux inspection for cracks in a type of malleable casting, about $p \approx .11$ of the castings will have detectable cracks. Consider the examination of 12 such castings. Let X be the number of castings from the set of 12 identified as being cracked.

(a) Find $P[X = 5]$.

(b) Find $P[X > 5]$.

(c) Find $\mathrm{E}X$.

(d) Find Var X.

(e) Ten sets of 12 castings are to be inspected. What is the probability that at least one set of 12 will have one or more cracked castings?

32. **Plastic Packaging**. This is a plastic packaging case investigated by Hsiao, Linse and McKay. Plastic bags were supposed to hold three bagels each. An ideal bag is $6.75\,\text{in}$ wide, has a $1.5\,\text{in}$ lip, and has a total length of $12.5\,\text{in}$ (including the lip). The ideal hole positions are on the lip. The hole position on selected bags was measured as the distance from the bottom of the bag to the hole. Five bags were obtained at six times on each of three days. Hole position, bag width, bag length, and lip width were measured and recorded for each bag. The data for hole position (in inches) are below.

Day	Time	Hole Position
1	10:10 am	1.87500, 1.84375, 1.87500, 1.84375, 1.84375
1	10:25 am	1.90625, 1.90625, 1.90625, 1.87500, 1.90625
1	10:55 am	1.87500, 1.93750, 1.93750, 1.93750, 1.96875
1	11:12 am	2.09375, 2.12500, 2.21875, 2.15625, 2.12500
1	11:35 am	2.00000, 2.00000, 2.00000, 2.00000, 2.03125
1	11:41 am	1.87500, 1.90625, 1.90625, 1.87500, 1.93750
2	8:15 am	1.62500, 1.62500, 1.59375, 1.65625, 1.59375
2	8:54 am	1.62500, 1.62500, 1.59375, 1.68750, 1.65625
2	9:21 am	1.62500, 1.59375, 1.62500, 1.59375, 1.62500
2	9:27 am	1.62500, 1.59375, 1.62500, 1.65625, 1.65625
2	9:51 am	1.56250, 1.59375, 1.56250, 1.56250, 1.56250
2	9:58 am	1.56250, 1.56250, 1.56250, 1.53125, 1.56250
3	10:18 am	1.50000, 1.56250, 1.53125, 1.53125, 1.50000
3	10:33 am	1.53125, 1.53125, 1.53125, 1.53125, 1.50000
3	10:45 am	1.50000, 1.53125, 1.50000, 1.53125, 1.46875
3	11:16 am	1.50000, 1.50000, 1.50000, 1.53125, 1.50000
3	11:24 am	1.53125, 1.53125, 1.50000, 1.50000, 1.50000
3	11:39 am	1.50000, 1.50000, 1.53125, 1.53125, 1.53125

(a) What is a natural subgroup in this situation?

(b) How many items are in each subgroup described in (a)? How many subgroups are there here in total?

(c) Calculate the subgroup means and subgroup ranges.

(d) Make a retrospective control chart for mean hole position. Give the center line, control limits, and zone limits.

(e) Make a retrospective control chart for variability in position using your values from (c). Give the control limits and zone limits.

(f) What is the usual name of the chart in (d)? What is the usual name of the chart in (e)?

(g) Is it important which of the charts developed in (d) and (e) is analyzed first? Why or why not?

(h) Find the estimated standard deviation of hole position based on the ranges.

33. Refer to the **Plastic Packaging** case in problem 32.

(a) Calculate the 18 subgroup means and 18 subgroup ranges.

(b) For each day separately, make retrospective control charts for mean hole position. Give center lines, control limits, and zone limits. What do these charts suggest about process performance?

(c) For each day separately, make retrospective control charts for variability of hole position.

(d) Based on your answer to (c), is variability of hole location constant within any one of the days? Why or why not?

(e) According to your charts in (c), is there a day in which a single standard deviation of hole position is plausible? Why or why not?

(f) Suppose your answer in (e) is "yes" for each day. Find estimated σ's for the three different days treated separately. (Base your estimates on sample ranges.)

(g) Comment on how your estimates in (f) compare to the estimate in part (h) of problem 32.

34. Refer to the **Plastic Packaging** case in problems 32 and 33. The ideal lip width is $1.5\,\text{in}$. The lip width data below (in inches) were taken on the same bags represented in problem 32.

(a) Is this a variables data or an attributes data scenario? Why?

(b) Find the subgroup means, ranges, and standard deviations.

(c) Make retrospective control charts for lip width variability and lip width mean based on the sample ranges.

(d) In view of the appearance of your chart for variability of lip width, does it make sense to seriously examine the chart for mean lip width? Why or why not?

(e) Instead of making the completely retrospective charts asked for in (c), is it possible to incorporate some "standards" information and make a different chart for mean lip width? Explain.

(f) Instead of treating all 18 samples at once as in part (c), for each day *separately*, make retrospective R and $\overline{x}$ charts. What are your conclusions regarding process stability for each day *separately*?

Day	Time	Lip Width
1	10:10 am	1.75000, 1.62500, 1.62500, 1.65625, 1.62500
1	10:25 am	1.62500, 1.62500, 1.62500, 1.65625, 1.65625
1	10:55 am	1.53125, 1.53125, 1.50000, 1.50000, 1.50000
1	11:12 am	1.40625, 1.43750, 1.43750, 1.46875, 1.46875
1	11:35 am	1.46875, 1.46875, 1.46875, 1.46875, 1.40625
1	11:41 am	1.43750, 1.43750, 1.46875, 1.50000, 1.46875
2	8:15 am	1.37500, 1.40625, 1.37500, 1.40625, 1.37500
2	8:54 am	1.37500, 1.43750, 1.43750, 1.40625, 1.40625
2	9:21 am	1.40625, 1.37500, 1.43750, 1.40625, 1.40625
2	9:27 am	1.50000, 1.46875, 1.43750, 1.46875, 1.43750
2	9:51 am	1.43750, 1.43750, 1.43750, 1.43750, 1.43750
2	9:58 am	1.53125, 1.46875, 1.53125, 1.50000, 1.53125
3	10:18 am	1.53125, 1.56250, 1.50000, 1.50000, 1.53125
3	10:33 am	1.50000, 1.53125, 1.53125, 1.50000, 1.50000
3	10:45 am	1.34375, 1.34375, 1.34375, 1.37500, 1.37500
3	11:16 am	1.46875, 1.46875, 1.46875, 1.43750, 1.43750
3	11:24 am	1.37500, 1.40625, 1.40625, 1.40625, 1.40625
3	11:39 am	1.43750, 1.43750, 1.40625, 1.37500, 1.43750

(g) Find three daily estimated lip width standard deviations. How do these estimates compare to that calculated when the complete set of data is used? (See (c) above.)

(h) Would there be advantages to using subgroup standard deviations instead of subgroup ranges in parts (c) and (f) above? Explain.

35. Refer to the **Plastic Packaging** case in problem 32.

(a) Make a control chart for the standard deviation of hole position. Is short term variation stable?

(b) Make a control chart for mean hole position based on subgroup standard deviations. Is process aim stable?

(c) For each day *separately*, make charts for the standard deviation of hole position. Is short term variation stable for each day?

(d) For each day *separately*, make charts for mean hole position (use an estimate of σ based on the subgroup standard deviations). Is process aim stable for each day?

(e) For each of (a), (b), (c), and (d), was it helpful to use the subgroup standard deviations instead of the subgroup ranges as in problem 32? Why or why not?

36. Refer to the **Hose Skiving** case of problem 11 in the Chapter 1 exercises. The plant works two shifts/day. Five hoses were sampled every two hours from each

of three production lines and skive length, y measured. Specifications for skive length are $target \pm .032$ inches. The values ($x = y - target$) in the accompanying tables are in units of $.001$ in above target.

(a) Explain (possibly using the notions of "rational subgrouping" and "stratification") why it would not make good sense to combine data taken at a particular time period from the three different production lines to make a single "sample." (Particularly in cases where it is only possible to select a single item from each line at a given time period, the urge to make such a "sample" is strong, and this kind of error is a common one.)

(b) Compute the 48 sample means and ranges for the data given here Then separately for lines 1, 2, and 3 make an $\overline{x}$ chart and an R chart. Comment on what they indicate about the stability of the skiving process on the three lines over the two days of this study.

(c) One *could* think about plotting all 48 sample means on a single chart, for example plotting means from lines 1, 2, and 3 in that order at a given time period. Discuss why that is not a terribly helpful way of summarizing the data. (Will it be easier or harder to see trends for a given production line on this kind of plot or on the separate charts of part (b)?)

		Line 1	Line 2	Line 3
Day	**Time**	**Skive Length**	**Skive Length**	**Skive Length**
1	8:00 am	$3, 2, 4, -5, 2$	$-17, 3, 2, 10, 4$	$-3, -5, 7, 10, 3$
1	10:00 am	$5, -4, -3, 0, -2$	$13, 3, -2, 12, 15$	$3, 5, 5, 8, 1$
1	12:00 pm	$-5, 5, 5, -3, 2$	$14, 6, 10, 5, 1$	$3, 6, 6, 5, 5$
1	2:00 pm	$-2, 5, 4, -3, 2$	$7, 2, 10, 16, 13$	$5, -2, 5, 4, 6$
1	4:00 pm	$-10, 2, 1, 2, 1$	$-15, -12, -2, -4, 0$	$2, 5, 4, 1, 1$
1	6:00 pm	$-5, -6, -3, -3, -7$	$-4, -6, -4, -4, 4$	$2, 1, 0, 1, 1$
1	8:00 pm	$-5, 0, -3, -3, -8$	$2, -5, -5, -3, -4$	$1, 3, 5, -6, -10$
1	10:00 pm	$-5, -10, 10, -9, -3$	$0, -1, -2, -1, 0$	$-7, -5, 4, 2, -9$
2	8:00 am	$2, 4, 1, 0, -5$	$15, 2, 16, 10, 14$	$18, 15, 5, 3, 4$
2	10:00 am	$-3, 3, -4, 5, 3$	$12, 4, -10, 10, -3$	$3, 2, -2, -5, 2$
2	12:00 pm	$-5, -7, 6, 8, -10$	$1, -7, 4, -5, -9$	$4, 2, 2, 1, 3$
2	2:00 pm	$3, -4, 4, 6, -3$	$-6, 8, -5, 18, 20$	$6, 5, 4, 2, 5$
2	4:00 pm	$-10, -7, -3, -1, -3$	$-2, -4, -5, -1, -3$	$2, 0, 1, -3, 5$
2	6:00 pm	$0, -1, -6, -2, 0$	$-2, -2, -2, -4, -2$	$2, -5, -7, -3, -5$
2	8:00 pm	$2, 4, -2, -3, 5$	$0, 2, -1, -1, -2$	$-6, -3, -10, -4, -7$
2	10:00 pm	$1, 0, -1, 7, -5$	$-1, -2, 0, -1, -1$	$0, -4, -7, -10, -2$

37. Consider the following hypothetical data from a process where $T(t)$ is the target value, $Y(t)$ is the realized value of the characteristic of interest, $E(t) = T(t) - Y(t)$, $\Delta E(t) = E(t) - E(t-1)$, and $\Delta^2 E(t) = \Delta E(t) - \Delta E(t-1)$. A PID controller $\Delta X(t) = \kappa_1 \Delta E(t) + \kappa_2 E(t) + \kappa_3 \Delta^2 E(t)$ has been used.

Period, t	$T(t)$	$Y(t)$	$E(t)$	$\Delta E(t)$	$\Delta^2 E(t)$	$\Delta X(t)$
1	4	2	2			
2	4	1	3	1		
3	4	0	4	1	0	18
4	4	2	2	−2	−3	1
5	4	2	2	0	2	
6	4	3	1	−1	−1	
7	5	3	2	1	2	
8	5	4	1	−1	−2	
9	5	5	0	−1	0	
10	5	6	−1	−1	0	
11	5	6	−1	0	1	
12	5	6	−1	0	0	

(a) What does $\Delta X(t)$ represent? Was the adjustment $\Delta X(3)$ made before or after observing $Y(3) = 0$?

(b) Suppose the integral gain in the control algorithm is 4. What are the proportional and derivative gains?

(c) Using your answer in (b), find the complete set of $\Delta X(t)$'s.

(d) Find the average squared error for the last 9 periods, the last 8 periods, . . . , and the last 3 periods.

(e) Make a plot of your values from (d) as follows. Label the horizontal axis with t. For $t = 4$ plot the average squared error for periods 4 through 12, for $t = 5$ plot the average squared error for periods 5 through 12, . . . , for $t = 10$ plot the average squared error for periods 10 through 12. Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance? Explain.

38. **Paper Dry Weight**. Before progressing to the collection of the data in Table 3.9, several different PID algorithms were tried. Miami University Paper Science Lab Research Associate Doug Hart set up the paper-making machine with 1% de-inked pulp stock to produce $20\,\mathrm{lb}$ bond paper. No filler was used. Then Jobe and Hart began an investigation into how to best control the dry weight variable. Twelve periods of data were obtained to benchmark the process behavior without any pump speed adjustments. (It is well known that the pump speed does affect final dry weight.) A standard setting of 4.5 (45% of maximum speed that was known to produce paper with a dry weight in the vicinity of the target of $70\,\mathrm{g/m^2}$) was used. Paper dry weight measurements were made at roughly $5\,\mathrm{min}$ intervals, and these are presented below as $Y(t)$. Units are $\mathrm{g/m^2}$.

Time	**Period,** t	$T(t)$	$Y(t)$	$E(t)$
8:45	1	70	75.3	−5.3
8:50	2	70	75.8	−5.8
8:55	3	70	73.1	−3.1
9:00	4	70	72.4	−2.4
9:05	5	70	73.5	−3.5
9:10	6	70	72.8	−2.8
9:15	7	70	72.6	−2.6
9:20	8	70	71.7	−1.7
9:25	9	70	69.8	.2
9:30	10	70	66.9	3.1
9:45*	11	70	70.9	−.9
9:50	12	70	71.7	−1.7

(a) Plot the measured values $Y(t)$ versus t.

(b) Plot the errors $E(t)$ versus t.

(c) Find the average squared error for periods 1 through 12, for periods 2 through 12,. . . , for periods 10 through 12.

(d) Make a plot of your values from (c) for $t = 1, 2, \ldots, 10$. (At time t plot the average squared error for periods t through 12.)

39. Refer to the **Paper Dry Weight** case of problem 38. Hart informed Jobe that for every 5-tick increase on the speed pump dial, paper dry weight increases about $1.5\,\mathrm{g/m^2}$. This means that in rough terms, to increase a dry weight by $1\,\mathrm{g/m^2}$, an increase of pump speed setting of about 3.33 ticks is needed.

(a) If one were to consider an "integral only" version (a $\kappa_1 = \kappa_3 = 0$ version) of the control equation (3.52) for use with the paper-making machine, why might $\kappa_2 = 3.33$ be a natural first choice? (X is in ticks, while T and Y are in $\mathrm{g/m^2}$.)

(b) The "integral only" controller of part (a) was used for 7 time periods and paper dry weight data collected. This is summarized in the table below. Fill in the $\Delta X(t)$ and $E(t)$ columns in that table for $t = 1, 2, \ldots, 8$. (The machine was running without adjustment with X set at 4.5 until 9:55. The measurements were taken far enough apart in time that the entire effect of a pump speed change ordered on the basis of data through a given period was felt at the next measuring period.)

(c) Plot $Y(t)$ versus t.

(d) Plot $E(t)$ versus t.

(e) Find the average squared error for periods 2 through 8, for periods 3 through 8,. . . , for periods 6 through 8.

Time	Period, t	$T(t)$	$Y(t)$	$E(t)$	$\Delta X(t)$
9:55	1	70	72.1		
10:08	2	70	70.6		
10:14	3	70	71.3		
10:25	4	70	67.1		
10:32	5	70	71.5		
10:38	6	70	70.3		
10:44	7	70	68.4		
10:50	8	70	71.7		

(f) Make a plot of your values from (e) for $t = 2, \ldots, 6$. (At time t plot the average squared error for periods t through 8.) Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance?

40. Refer to the **Paper Dry Weight** case in problems 38 and 39. At 10:50 the speed pump dial was set back to 4.5 (45%) and left there for $5\,\mathrm{min}$ in order to return the system to the benchmark conditions of problem 38. A new coefficient κ_2 in an integral control algorithm was adopted and beginning at 10:55 this new adjustment algorithm was employed for 7 periods with results summarized in the following table.

Time	Period, t	$T(t)$	$Y(t)$	$E(t)$	$\Delta X(t)$
10:55	1	70	72.0	-2	-3.32
11:01	2	70	71.7		
11:13	3	70	71.1		
11:19	4	70	68.8		
11:25	5	70	69.6		
11:31	6	70	71.8		
11:37	7	70	68.2		
11:43	8	70	69.7		

(a) Find the value of the new coefficient κ_2 used by Jobe and Hart. Then fill in the $E(t)$ and $\Delta X(t)$ values in the table for $t = 2, \ldots, 8$.

(b) Plot $Y(t)$ versus t.

(c) Plot $E(t)$ versus t.

(d) Find the average squared error for periods 2 through 8, for periods 3 through 8,. . . , for periods 6 through 8.

(e) Make a plot of your values from (d) for $t = 2, \ldots, 6$. (At time t plot the average squared error for periods t through 8.) Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance?

41. Refer to the **Paper Dry Weight** case of problems 38, 39, and 40. After making the measurement at 11:43 indicated in problem 40, the speed pump dial was again set back to 4.5 and left there for $5\,\text{min}$ (from 11:44 to 11:49). (This was again done to in some sense return the system to the benchmark conditions.) Hart and Jobe decided to include both integral and proportional terms in a new control equation and $\kappa_2 = 1.66$ and $\kappa_1 = .83$ were selected for use in equation (3.52). (The same integral control coefficient was employed, and a proportional coefficient half as large as the integral coefficient was added.) This new adjustment algorithm was used to produce the values in the table below.

Time	**Period,** t	$T(t)$	$Y(t)$	$E(t)$	$\Delta E(t)$	$\Delta X(t)$
11:49	1	70	70.9	$-.9$		
11:54	2	70	70.3	$-.3$	.6	0
11:59	3	70	68.8			
12:06	4	70	70.0			
12:12	5	70	69.6			
12:18	6	70	69.3			
12:24	7	70	68.4			
12:30	8	70	68.4			
12:36	9	70	69.8			

(a) Find the values of $E(t)$, $\Delta E(t)$, and $\Delta X(t)$ for periods 3 through 9.

(b) Plot $Y(t)$ versus t.

(c) Plot $E(t)$ versus t.

(d) Find the average squared error for periods 3 through 9, for periods 4 through 9,. . . , for periods 7 through 9.

(e) Make a plot of your values from (d) for $t = 3, \ldots, 7$. (At time t plot the average squared error for periods t through 9.) Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance?

42. Refer to the **Paper Dry Weight** case of problems 38 through 41. Yet another control algorithm was considered. κ_1 from problem 41 was halved and the coefficient κ_2 was left at 1.66. The pump speed dial was set to 4.5 at 12:37. Thereafter, the new "PI" control algorithm was used to produce the values in the table below.

(a) Find $E(t)$ for all 9 periods and $\Delta E(t)$ and the corresponding $\Delta X(t)$ for periods 2 through 9.

(b) Plot $Y(t)$ versus t.

(c) Plot $E(t)$ versus t.

(d) Find the average squared error for periods 3 through 9, for periods 4 through 9, . . . , for periods 7 through 9.

Time	**Period,** t	$T(t)$	$Y(t)$	$E(t)$	$\Delta E(t)$	$\Delta X(t)$
12:42	1	70	66.2			
12:45	2	70	66.4			
12:51	3	70	67.2			
12:58	4	70	69.4			
1:04	5	70	69.5			
1:10	6	70	69.2			
1:16	7	70	70.1			
1:22	8	70	66.2			
1:29	9	70	71.7			

(e) Make a plot of your values from (d) for $t = 3, \ldots, 7$. (At time t plot the average squared error for periods t through 9.) Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance?

43. Refer to Example 37.

(a) Plot $Y(t)$ versus t.

(b) Plot $E(t)$ versus t.

(c) Find the average squared error for periods 4 through 11, for periods 5 through 11,. . . , for periods 9 through 11.

(d) Make a plot of your values from (c) for $t = 4, \ldots, 9$. (At time t plot the average squared error for periods t through 11.) Does the appearance of this plot give you hope that any transient or "startup" effects have been eliminated before the last few periods and that those periods adequately represent control algorithm performance?

44. Refer to the **Paper Dry Weight** case and specifically the plots in problems 38(d), 39(f), 40(e), 41(e), 42(e), and 43(d). Which control equation seems to be best in terms of producing small average squared error?

45. Rewrite the PID control equation (3.52) so that $\Delta X(t)$ is expressed in terms of a linear combination of $E(t)$, $E(t-1)$, and $E(t-2)$, the current and two previous errors.

46. Fill levels of jelly jars are of interest. Every half hour, three jars are taken from a production line and net contents measured and recorded. The range and average of these three measurements are calculated and plotted on charts. One of these charts is intended to monitor location of the fill distribution and the other is useful in monitoring the spread of the fill distribution.

(a) What is the name for the chart used to monitor location of the fill level distribution?

(b) What is the name for the chart used to monitor spread of the fill level distribution?

(c) What is the name and value of the tabled constant used to make retrospective control limits for process location?

(d) What are the names and values of the two tabled constants used to make retrospective control limits for process spread or short term variability?

(e) In this context, what constitutes a natural subgroup?

(f) Give an expression for the usual estimate of process short-term variability (σ) based on an average of subgroup ranges..

47. Consider again the scenario of problem 46. Suppose that instead of ranges and averages, sample standard deviations and averages are computed and plotted.

(a) What is the name and value of the tabled constant used to make retrospective control limits for process location?

(b) What are the names and values of the two tabled constants used to make retrospective control limits for process spread or variability?

(c) Give an expression for the usual estimate of process short-term variability (σ) based on an average of subgroup standard deviations.

48. Consider again the scenario of problems 46 and 47 and suppose that instead of three jars, 10 jars are sampled every half hour. Redo problems 46 and 47 with this change. For a given set of ranges or standard deviations say which sets of retrospective control limits are wider apart with this new sample size.

49. Consider again the scenario of problems 46 and 47 and suppose that instead of plotting averages to monitor location, the decision is made to plot medians. What multiple of σ (or an estimate of this quantity) would be used to set control limits for medians around some central value in the case that $n = 3$? In the case that $n = 11$?

50. Consider drained weights of the contents of cans of Brand X green beans. Believable values for the process mean and standard deviation of these weights are 21.0oz and 1.0oz respectively. Suppose that in a Brand X canning factory, 8 of these cans are sampled every hour and their net contents determined. Sample means and ranges are then computed and used to monitor stability of the filling process.

(a) What is the name and value of the multiplier of $\sigma = 1.0$ that would be used to establish a center line for sample ranges?

(b) What are the names and values of the multipliers of $\sigma = 1.0$ that would be used to establish upper and lower control limits for sample ranges?

(c) What center line and control limits should be established for sample means?

51. Consider again the situation of problem 50, but suppose that instead of ranges and averages, sample standard deviations and averages are computed and plotted. Answer the questions posed in problem 50 in this case.

52. Consider again the situation of problems 50 and 51 and suppose that instead of 8 cans, only 5 cans are sampled every hour. Redo problems 50 and 51 with this change. Say which sets of control limits are wider apart with this new sample size.

53. **Electronic Card Assemblies**. In a 1995 article in *Quality Engineering*, Ermer and Hurtis discussed applications of control charting to the monitoring of soldering defects on electronic card assemblies. One assembly technology they studied was pin-in-hole (PIH) technology, which uses wave soldering to secure components to printed circuit boards after the leads of components have been inserted through holes drilled in the boards. The most common types of soldering defects encountered using this technology are "shorts" (unwanted electrical continuity between points on an assembly) and "opens" (the absence of desired electrical continuity between points on an assembly).

    Production of a particular card is done in "jobs" consisting of 24 cards. All cards from a job are tested and a count is made of the total number defects found on the job. What type of probability model might plausibly be used to describe the number of defects found on a given job? What type of control chart might you use to monitor the production of soldering defects? Suppose that records on 132 jobs show a total of 2 defects recorded. What retrospective control limits might then be applied to the 132 different counts of defects? Does a job with any defect at all signal a lack of control?

54. **Milling Operation**. A student group studied a milling operation used in the production of a screen fixture mounting. Of primary importance was a "deviation from flatness" measurement. The units of measurement were $.001\,\text{in}$. In the past, deviations from flatness had an average of 2.45 and a standard deviation of 1.40. What do the values of these and the fact that deviation from flatness can not be negative suggest about the plausibility of a normal model for deviation from flatness? For what follows temporarily put aside any misgivings you might rightly have.

    (a) Set up standards given control limits for process location. (Monitoring is to be done on the basis of subgroups of size one.)

    (b) Ten consecutive mountings produced the deviation from flatness values below (units are $.001\,\text{in}$).

$$.5, 4.5, 2.0, 2.0, 3.0, 3.0, 2.0, 4.5, 3.0, 0.0$$

Together with the limits in (a), use these data values to make a control chart for monitoring process aim. Has there been a process change away from the standards? Why or why not?

(c) Find the moving ranges of adjacent observations and the mean of these 10 observations.

(d) Make a retrospective individuals chart using the moving ranges and grand average of the 10 data values. Give the center line and control limits. What do you conclude based on this chart?

55. Refer to the **Milling Operation** case in problem 54.

(a) For purposes of process monitoring only, let a target deviation from flatness be $\mu = 5$, and suppose the process standard deviation is $\sigma = 1.40$, as in problem 54. (In functional terms a 0 deviation from flatness is ideal.) Compute control limits for individuals based on this set of standards. Give the center line and control limits.

(b) Plot the individuals from problem 54(b) using your new limits from (a). Does it appear that there has been a process change from standard conditions? Why or why not?

(c) Discuss the practical meaning of the terms "stability," "shift," and "out-of-control" in light of part (b) and (d) of problem 54 and part (b) above.

56. Refer to the **Lab Carbon Blank** case in problem 21 of Chapter 1 and problem 29 of Chapter 2. Suppose that repeated measurements of the same blank are normally distributed. For convenience, the data are repeated here.

Test Number	1	2	3	4	5	6	7
Measured Carbon	5.18	1.91	6.66	1.12	2.79	3.91	2.87
Test Number	8	9	10	11	12	13	14
Measured Carbon	4.72	3.68	3.54	2.15	2.82	4.38	1.64

(a) Find retrospective control limits and center line for the sequence of measurements. Use "3 sigma" limits.

(b) Plot the individual responses versus time and compare them to the limits found in (a). Do you detect any measurement process instability? Why or why not?

(c) Give an estimated mean measured carbon content. Give an estimated standard deviation of measured carbon content, $\overline{MR}/1.128$.

57. Refer to the **Lab Carbon Blank** case of problem 56. Suppose the nominal or "real" carbon content is 1.0.

(a) Find control limits and center line to apply to the data of problems 56. Use 3 sigma limits and $\overline{MR}$ in place of a real average range ($\overline{R}$) in the formula for retrospective limits. Make use of the nominal value 1.0 in place of $\overline{\overline{x}}$.

(b) Plot the x values from problem 56 and compare them to your limits from (a), i.e., make an individuals chart.

(c) What dilemma is revealed by your chart in (b) above and problem 56(b)? Discuss this using phrases such as "consistency of location," "shift in the mean," "off-target process," and "unstable process."

58. Refer to the **Lab Carbon Blank** case in problems 56 and 57. It is unknown whether the carbon measurements were made by the same person or by as many as 14 different people. What configuration of operators would be most effective in isolating instrument changes in the measurement of carbon content? Defend your answer in terms of the concept of "sources of variability."

59. Consider the analysis of a series of samples some time after their collection.

(a) What might be learned from an analysis based on standards given control charts?

(b) What might be learned from an analysis instead using retrospective limits on the control charts?

60. Refer to the **Paper Dry Weight** case in problem 38. Recall that the target for dry weight of $20\,\mathrm{lb}$ bond paper is $70\,\mathrm{g/m^2}$. The pump speed controlling the flow of liquid pulp mixture onto the conveyor-roller mechanism was held fixed at 4.5 (45% of maximum flow) in the production of the data in problem 38. Assume that under stable process conditions dry weights are normally distributed. The dry weight of a sample was recorded for each of 12 consecutive samples, approximately equally spaced in time.

(a) Find control limits to apply to the data of problem 38. Use the nominal dry weight of 70 as a target value and employ $\overline{MR}/1.128$ as an estimate of σ. Does the process appear to be stable? Why or why not?

(b) 100 measurements correspond to how many subgroups in the context of problem 38?

(c) Suppose that the limits of (a) are applied to the future monitoring of individuals. About what ARL is produced if σ is as in part (a), but μ increases from its standard value by $3.5\,\mathrm{g/m^2}$? Assume the process is stable and dry weights are normally distributed.

(d) If completely retrospective control limits were used ($\overline{x}$ was used in place of the target value for dry weight) would your conclusion in (a) change? Why or why not?

61. **Transmission Housings.** Apple, Hammerand, Nelson and Seow analyzed data taken from a set of 35 transmission housings. In addition to the side cover hole diameter considered in problem 4 of Section 2 of this chapter, they also examined upper bore hole diameters on the transmission housings. For y the hole diamter

in inches, the values below concern $x = (y - 3.5000) \times 10^4$, diamters stated in ten thousandths of an inch above $3.5000\,\text{in}$. Specification limits for the upper bore hole diameter were $3.502 \pm .002\,\text{in}$.(Below, 19 represents $y = 3.5019$, 28 represents $y = 3.5028$, etc.)

Measured Diameter	19	28	25	22	18	20	20	14	20	12	16	16
Transmission Housing	1	2	3	4	5	6	7	8	9	10	11	12
Measured Diameter	22	22	22	21	23	21	20	18	18	18	12	11
Transmission Housing	13	14	15	16	17	18	19	20	21	22	23	24
Measured Diameter	13	12	16	12	10	20	21	15	28	26	24	
Transmission Housing	25	26	27	28	29	30	31	32	33	34	35	

(a) What is the subgroup size?

(b) Give appropriate retrospective lower and upper control limits and center line for monitoring the hole diameters.

(c) What does the chart using on your limits from (b) indicate about the stability of the upper bore hole diameter production process? Why?