Statistics and SPC

Two things sharing a common name can still be different

Donald J. Wheeler

Students typically encounter many obstacles while learning statistics. In 44 years of teaching I have discovered some distinctions that help students overcome these obstacles. This article will remove some sources of confusion concerning the relationship between SPC and other statistical techniques. While the distinctions made here have not been widely taught, I have found that they are essential to the proper use of both SPC and the traditional statistical techniques. These distinctions will allow you to use SPC to improve quality, increase productivity, and gain competitive advantage, while also using traditional statistical techniques appropriately.

THE BACKGROUND

Walter Shewhart's *Economic Control of Quality of Manufactured Product* was published in 1931. In this book he was very careful to make a distinction between the techniques of statistical inference and the approach he was taking for the analysis of observational data. However, in 1935 E. S. Pearson glossed over these distinctions and sought to reformulate Shewhart's ideas in terms of the elements of statistical inference. Among the many complications introduced by Pearson we have the idea that the data have to be normally distributed prior to placing them on a process behavior chart. In his rebuttal to Pearson's book Shewhart wrote: *"we are not concerned with the functional form of the universe* [i.e. the probability model], *but merely with the assumption that a universe exists."* Since this statement is in italics in the original, either Shewhart or his editor, W. Edwards Deming, felt that this statement was very important. While we will consider what these two different approaches mean in practice, we need to begin with some useful distinctions.

TWO TYPES OF DATA

The first difference between SPC and the techniques of statistical inference has to do with the type of data used by each. The techniques of statistical inference were developed to analyze *experimental data*. Such data are generally collected under different conditions with the purpose of determining whether changes in the input variables have an effect upon a response variable.

SPC was created for the analysis of *observational data*. As the name suggests, observational data are a by-product of some routine operation. These data may be deliberately and intentionally collected, but they are collected while the process is operated in an ordinary manner. Observational data simply track the underlying process while the input variables are usually held constant.

Thus, the first difference has to do with the type of data being considered. The difference in the conditions under which these two types of data are obtained immediately reveals two additional ways that experimental data differ from observational data. An experiment will always yield a fixed amount of data. In contrast, in an observational study we can usually obtain

additional data by simply waiting. Moreover, because experimental data are collected under special conditions, they tend to be more expensive than observational data. Figure 1 summarizes these differences between observational data and experimental data.

Observational Data	Experimental Data
One Condition Present	Two or More Conditions Present
Additional Data Available	Fixed Amount of Data
Less Expensive Data	More Expensive Data

Figure 1: Three Differences Between Observational Data and Experimental Data

DIFFERENT EXPECTATIONS

We also have different expectations for these two types of data. We expect our experimental data to represent the differences between the special conditions being studied. Thus, in an *experimental study* we are looking for differences that we have paid good money to create and that we believe are contained within the data. Moreover, we only get one chance to detect these differences when we analyze the data. The expectation that there will be signals within our data, the finite amount of data available, and the fact that we only get one chance to analyze the data, will tend to make us choose a less conservative and more exploratory approach to the analysis. Hence the traditional five percent alpha-level that is common with the one-time analysis in an experimental study.

On the other hand, when conducting an *observational study*, our data will generally be collected under one condition. As a result we will not expect to find any differences within the data. Furthermore, since any differences will usually indicate unplanned changes in the process, we will want to be sure about any differences we find before we take action. Since additional data will usually be available, we can also afford to play a waiting game with observational studies. These two characteristics combine to make us want to use a conservative analysis. And indeed, the limits on a process behavior chart provide a conservative analysis for each new observation added to the chart. Thus we have a very small risk of a false alarm, so that before we take action we will have strong evidence of a process change. Figure 2 summarizes these differences between observational studies and experimental studies.

Observational Studies	Experimental Studies
One Condition Present	Two or More Conditions Present
Additional Data Available	Fixed Amount of Data
Less Expensive Data	More Expensive Data
Should Be No Signals	Should Be Some Signals
Sequential Analysis Procedure Required	All Data Analyzed at One Time
Conservative Analysis Used	Traditional or Exploratory Analysis Used

Figure 2: Differences Between Observational Studies and Experimental Studies

THE STUDY OF LIONS

Another way in which experimental and observational studies differ is in the nature of the variables under consideration. In an experimental study the objective is to discover cause-and-effect relationships. This means that we will manipulate potential causes (the input variables) and observe what happens to certain response variables. (Do any of the causes have an effect upon the response?)

As soon as we start to do an experimental study we quickly discover that there are more input variables than we can reasonably include in the experiment. So we have to make a choice about what input variables to include in the experiment (call these the X variables), and what variables to leave out of the experiment. This latter set of variables can be further divided into those input variables that are held constant during the course of the experiment (the Y input variables), and those input variables that are ignored during the experiment (the Z input variables). This latter group would include the environmental variables that are beyond the control of the experimenter as well as those input variables that are thought to have minimal impact upon the response variable. Thus, there are three categories of input variables for every experiment:

1. those that are studied by the experiment, the X input variables;
2. those that are held constant during the experiment, the Y input variables; and
3. those that are ignored by the experiment, the Z input variables.

To keep the Z variables from undermining the experiment we tend to randomize how they show up in the study. This will shuffle the effects of the Z variables. The idea behind randomization is to get the effects of the Z variables to (hopefully) average out within each treatment over the course of the experiment. (Be sure to keep your fingers crossed!)

As a result of these three classes of input variables all experimental results must be interpreted with two caveats:

1. The results have to be interpreted under the assumption that randomization worked to average out the extraneous effects of the Z variables. (This assumption becomes more reasonable as the experiment contains more observations for each treatment combination.)

2. And the results have to be interpreted in the context of those specific levels that were used for the Y input variables. (If some of the Y input variables interact with some of the X variables then your results might be different had you chosen different levels for the Y variables.)

This is why experienced statisticians are always careful to look for those things that might have gone wrong during an experiment. In the words of George Box, "If you are looking for gold and accidentally find some silver along the way, stop and mine the silver." When things go wrong in an experment it is often a clue that the Y or Z variables are having an effect upon the response variable.

Observational studies collect data in a completely different way. To understand this difference we need to make a distinction between those input variables that are controlled during routine operations (call these the control factors) and all the remaining potential input variables

that are not controlled during routine operation (call these the uncontrolled variables). (In an experimental context, the X and Y variables tend to come from the set of control factors and the Z variables will tend to come from the set of uncontrolled variables, but the correspondence is not exact.) Observational studies tend to observe the response variables while all of the control factors are held constant. So what can we learn from an observational study?

With all of the control factors being held constant the process ought to behave very consistently and predictably. If it does not do so, then it has to be one of the uncontrolled variables that is making its effects known. Unplanned process changes that occur while the control factors are held constant are signs that the set of control factors is incomplete.

While we may not know which of the uncontrolled variables caused an unplanned change, the change itself will focus our attention on the time and place where at least one of the inputs changed. Observational studies may not prove that a cause and effect relationship exists, but they can alert us to the possibility of certain relationships. They are a powerful tool for discovery. Once we have an idea about a possible cause-and-effect relationship we can use simple experiments to examine our idea.

So what about the study of lions? There are many things we can learn about lions at the zoo. They are large. They eat meat. They have an impressive roar. But if you want to become an expert on lions, you are going to have to study lions in the wild. (Hopefully you can avoid being on the menu.) As you watch lions interact with their environment you will discover things that you would never have learned at the zoo. Experimental studies are like studying lions in a zoo. Here you are seeking the answers to specific questions. Observational studies are like studying lions in the wild. Here you are opening yourself up to discovering things never before dreamed of. There is a time an a place for each type of study.

Some statisticians belittle observational studies by claiming that they prove nothing. However, observational studies have been around for a long time. They have their origin in Aristotle's teaching that the time to discover those causes that influence a system is to look at those points where the system changes. While experiments may allow us to obtain definitive answers for specific questions, they are of little use when we do not know what questions to ask. Discovery must plant the seeds for experimentation, and observation is the mother of discovery.

APPLICATIONS OF TECHNIQUES

One-time analysis techniques such as t-tests, the Analysis of Variance, and the Analysis of Means were explicitly created for the analysis of finite sets of experimental data. Sequential analysis techniques such as the Average and Range Chart or the XmR Chart were *expressly created for the sequential analysis of continuing streams of observational data*.

As shown in Figure 3, we may adapt the sequential analysis techniques for use with finite sets of experimental data, but we cannot easily go in the other direction. One-time analysis techniques cannot be meaningfully adapted for use in situations requiring a sequential analysis technique. So, the two types of data require different analysis techniques. They also result in different objectives for the analysis.

In an experimental study the objective is to establish a particular relationship beyond a reasonable doubt. You want to be able to say: "Do this and you will get that." To this end you will need to estimate some parameters. These parameters may be the parameters for a

probability model, or they may be parameters for a regression equation, but either way, the objective is to obtain some reliable estimates for a set of parameters. To express the uncertainty in these estimates we often use interval estimates. These interval estimates will be dependent upon our choice of an alpha-level and the amount of data available. Together these two quantities will usually determine the appropriate critical value. Thus, the analysis of experimental data is characterized by alpha-levels, critical values, and an emphasis upon having enough data to obtain reliable estimates.
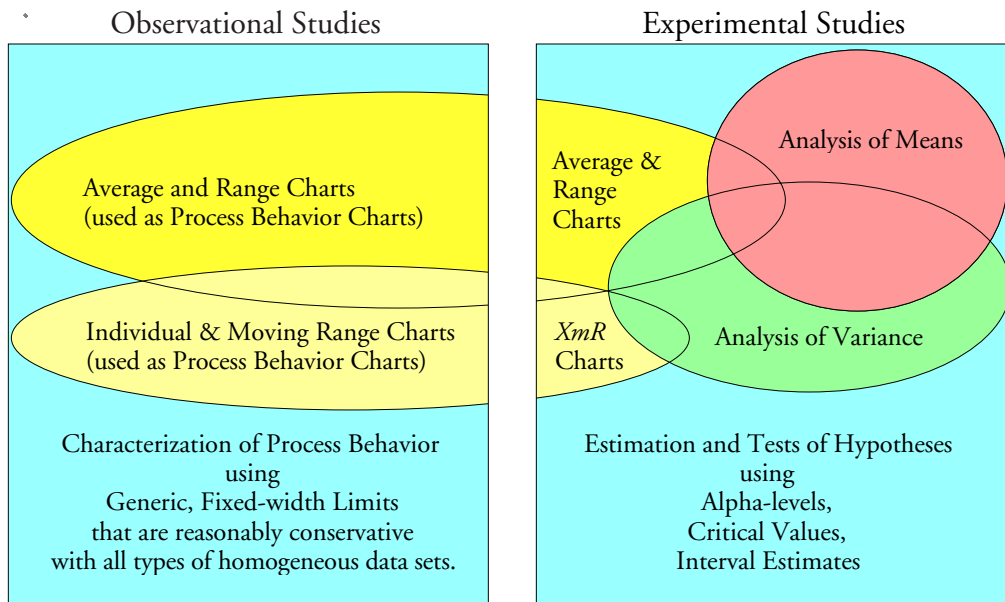
Observational Studies

Experimental Studies

Average and Range Charts
(used as Process Behavior Charts)

Individual & Moving Range Charts
(used as Process Behavior Charts)

Characterization of Process Behavior
using
Generic, Fixed-width Limits
that are reasonably conservative
with all types of homogeneous data sets.

Average &
Range
Charts

Analysis of Means

XmR
Charts

Analysis of Variance

Estimation and Tests of Hypotheses
using
Alpha-levels,
Critical Values,
Interval Estimates

Figure 3: Analysis Techniques for Observational and Experimental Studies

With an observational study we are not evaluating known changes but are instead trolling for unknown and unintentional changes. Here there is no probability model or regression equation with parameters to be estimated. The problem is not one of estimation but rather one of characterization. Process behavior charts seek to characterize the past process behavior as being in one of two rather broad categories. Has the process been operated predictably in the past, or does the process show evidence of unplanned and unexplained changes.

SPC	**Statistical Inference**
Observational Studies	Experimental Studies
Sequential Analysis	One Time Analysis
Characterize Process Behavior	Estimate Parameters for Model
Use Conservative Fixed-Width Generic Limits	Use Alpha-Levels & Critical Values
Discover Unknown Factors	Establish Relationships for Known Factors

Figure 4. SPC vs. Statistical Inference

To make this characterization the process behavior charts of SPC use generic, fixed-width limits. These limits provide a reasonably conservative analysis with virtually every type of homogeneous data set. Thus, no probability model has to be specified. No alpha-level is required. No critical values are needed. With this conservative, one-size-fits-all approach any

signals found are almost certain to be real, and this allows us to reliably characterize the process behavior without going through the rigamarole associated with statistical inferences. To read more about this see my article "*Are You Sure We Don't Need Normally Distributed Data?*" **QDD**, November 1, 2010.

CONSEQUENCES OF THE DIFFERENCES

If the generic limits of SPC show the observational data to be reasonably homogeneous, then we judge the underlying process to be reasonably predictable. At this point it will make sense to talk about the data as if they all came from a single process. If the data all come from a single process then the notion of process parameters makes sense. When the notion of process parameters makes sense, we can extrapolate beyond the data and begin to use our descriptive statistics to estimate these process parameters.

If the generic limits of SPC show the observational data to be non-homogeneous, then we can reasonably conclude that exceptional variation is present. Because of the conservative nature of the generic limits, the presence of exceptional variation will be strong evidence that the underlying process is changing. When the process is changing the process parameters will be changing. When the process parameters are changing the question of estimating those parameters becomes moot. While the descriptive statistics will still describe the data, we cannot extrapolate beyond the data and use the statistics to estimate the process parameters. Here the key question becomes "What is causing the process changes?"

If we do not address this question regarding the source of the process changes we will have missed an opportunity to improve the process. When we do address this question we will hopefully be making some changes in the process, and our old data will no longer be of interest. Either way, the shape of the histogram and the various descriptive statistics computed from our non-homogeneous data become irrelevant. Such data will merely be a collection of values coming from different processes. When the data are non-homogeneous any discussion of the shape of the histogram, or any attempt to estimate the process characteristics, will be like talking about the constellations in the night sky—a triumph of imagination over substance.

Thus, as shown in Figure 4, SPC differs from statistical inference in many ways. It was created for a different type of data. It is intended for use with a sequential stream of data rather than performing a one-time analysis. It has the objective of characterizing process behavior rather than estimating the parameters for some model or equation. It uses generic limits that are conservative for all types of homogeneous data sets rather than fine-tuning the critical values to achieve some specific alpha-level. And SPC seeks to detect and discover unknown factors that affect an existing process rather than trying to establish that a specific relationship exists between known variables.

As long as the analysis is conservative, when we find a signal we will be justified in taking action. The question is not "What is the alpha-level?" but rather "What caused the unplanned change in the process?" *And this is the objective of SPC, to detect the unknown process changes so that actions can be taken to turn the assignable causes of exceptional variation into control factors that are held constant during production, and thereby to reduce both variation and costs.*

CONFUSION OR CLARITY ?

The differences outlined here are profound. Any failure to understand these differences will inevitably result in confusion. Yet there is a built-in trap that almost guarantees that confusion will reign. It has to do with the nature of statistical training.

Statisticians are trained in the techniques of statistical inference. We see the world in terms of experiments, alpha-levels, critical values, and the estimation of parameters. Eventually this mind-set becomes so ingrained that it becomes difficult to escape. When those who operate with this mind-set are presented with Shewhart's approach they will usually seek to remold Shewhart's approach to fit in with the elements of statistical inference. Pearson did this in 1935. Many others, including this author, have done it since.

In the first edition of my second book I included a chapter in which I tried to reconcile SPC and statistical inference. As I taught my classes I noticed that this chapter tended to confuse my students, and so for years I skipped over this chapter in class. Later, as I was working on a second edition of that book, Dr. Deming sent me his annotated copy. The only comment he had made on that chapter was a single word on the first page—"Omit." As I reread that chapter I saw how I had fallen into the same trap that ensnared Pearson and others, and so in the revision I did, indeed, omit that chapter.

Today many of my colleagues continue to fall into the same trap by trying to "fill in the gaps" between Shewhart's approach and the techniques of statistical inference. Hallmarks of these attempts to "fill in the gaps" are statements like the following.

- "The data have to be normally distributed."
- "You need to transform your data."
- "You need to define a reference distribution before you can compute appropriate limits that will have the correct alpha-level."
- "Your process has to be in-control before you can place your data on a process behavior chart."
- "You can't put autocorrelated data on a process behavior chart."
- "The process behavior chart works because of the central limit theorem."
- "You have to have subgroups of size five in order for the average chart to work."
- "You have to have 30 subgroups before you can compute your limits."
- "You need to remove the outliers from the data before you compute your limits."

All of the statements above are wrong. Dr. Deming categorized statements like these as "such nonsense." As I have tried to show here, any attempt to "fill in the gaps" is based upon a failure to understand how SPC is different from statistical inference. The two use different approaches to different types of data and have different objectives.

There is a time and a place for the techniques of statistical inference. There is a time and a place for the use of SPC. To use both effectively you need to understand how they differ. You also need to avoid those who try to merge the two into one "unified" approach. Alpha-levels, critical values, distributional assumptions, tests for lack-of-fit, and the like all belong to the world of experimental studies and statistical inference.

From the very beginning the elements of statistical inference had no place in SPC. Today they still have no place in SPC. And they shall never have any place in SPC, world without end, amen. Understanding the differences between statistics and SPC can bring clarity. Seeking to unify statistics and SPC will simply create confusion.