Chapter 9
Sampling Methods

9.1 Sampling Defined

The researcher, designer, or quality control engineer must learn not only the mechanics of mathematical statistics, but also the concepts that lie behind them. One of the most fundamental is the method of sampling and plotting the results of sample tests. Too often sampling is not properly planned and this weights heavily on the results of tests as well as on findings.

Sampling is often left to a subordinate who draws samples that are neither random nor representative of the desired population, in total disrespect of the fact that sampling is one of the pillars of statistics and great attention should be exercised in its design (Sect. 9.2). The best way to start is by appreciating that no matter what our profession is, we constantly use samples to make inferences about the population from which they have been derived. This is a process involving a trilogy of human knowledge:

$$\text{Sample} \rightarrow \text{StatisticalInference} \rightarrow \text{Population}.$$

- A *population*, or universe, consists of all events, measurements, or other issues of interest in our present work.
- A *sample* is a subgroup of this population drawn under proper rules for specific reasons of study and research.

Closely associated to the notion of a population is its consistence. Ideally, it consists of a homogeneous mass of measurements, events, manufactured items or other assets whose nature, behavior, or some other variable is of interest to us. To be of value in a scientific study,

- Both the population and its sample have to be clearly defined, and
- The sampling method which we use must be carefully chosen.

There are many reasons why we use samples. One of them is that the population may be so scattered that it is impossible to reach and test all of it. Also, the tests we do may be

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_9,

Table 9.1 A bird's eye view of symbols for parameters and statistics

	Population	Sample
Mean	μ	$\bar{x}$
Standard deviation	σ	s

$\bar{x}$, s are statistics; μ, σ are parameters

destructive; or studying a representative sample of it may be much more effective; or it may be difficult to comprehend the significance of large quantities of ungrouped data.

For any one of these reasons, it is desirable to find a few numbers of mathematical expressions which describe the relevant properties of the entire field of data. These numbers are called the *parameters* of the population: We have spoken of μ for population's mean, and σ the population's standard deviation—which we study through their proxies derived from the sample, respectively $\bar{x}$ and s.

$\bar{x}$ and s are the *statistics* of the *sample*. A relatively small number of individual items can be selected, for instance at random, from the population (see Sect. 9.2 on sampling procedures). By analyzing sample data, we derive statistics which describe the properties of the sample. Inferences concerning the population parameters can then be drawn from these sample statistics. Table 9.1 gives a bird's-eye view of parameters and statistics. In this book we use the symbols μ and σ for mean and standard deviation respectively.

Sampling implies procedural rules based on probability theory which, however, are not always observed. Quite often, people make serious mistakes by sampling data that are not independent or by using small samples. Both errors lead to biased samples and false conclusions.

The population from which a sample is drawn, often called the *universe*, can be finite or it can be infinite. In practical applications, the way to bet is that the population will be finite but large, a reason why we may treat it as though it were infinite. By contrast, the samples we draw will be finite.

The size of a population is usually represented by the letter N. This is the number of events, measurements, individuals, or some other factor under study in the universe. The accuracy of a sample statistic always depends upon the size of the sample that has been determined. The size of a sample is usually represented by n.

A complete set of observations upon which a study, analysis, or experiment is based is called a *sample of n*, where n refers to the number of observations. *If* a group of observations involves different sets, *then* the sum of all sets of n observations in the group will be equal to $\sum n_i$, where i stands for each sample.

- The measurements, events, individuals, and so on in the population will form a distribution whose mean is μ and the variance is σ^2.
- The measurements or other factors in the sample will also form a distribution which will have a mean denoted by $\bar{x}$ and a variance s^2.

The sample's $\bar{x}$ and s^2 which we actually measure are expected to give us information about μ and σ^2, whose values are usually not known. Notice however that $\bar{x}$ and s^2 tend to be different from sample to sample. By contrast, for a particular population μ and σ^2 are constant.

The sample's statistics $\bar{x}$ and s^2 have a distribution, respectively known as sampling distribution of the mean and sampling distribution of the variance. By examining the sampling distribution of the mean we can tell how frequently the $\bar{x}$ of samples will fall in the interval we wish to fall. This is very important in statistical quality control (SQC) by variables, as we will see in Chap. 13. In a nutshell the quality control procedure which is typically followed consists of:

- Selecting a sample of the units of a manufactured product from the population.
- Observing and recording a quantitative test value of the appropriate statistic(s), and
- On the basis of the computed value of the statistic, and of the underlying quality rule, accept or reject the statement that the population possesses acceptable quality.

The reader should appreciate that the use of statistical methods including those employed for quality control, does not eliminate chance variation, but avails means for watching over it. By the use of statistical inference, scientific analysis replaces guesswork. However, chance variation is still present, and therefore a sound statistical plan should account for it.

For example, in a manufacturing process, SQC procedures should consider both the producer and the consumer viewpoint. The producer aims for quantity and hopes for the benefits of quality. He demands protection against the rejection of good product. The consumer aims for quality and hopes for the benefits of quantity. He demands protection against the acceptance of poor product. Thus:

- The perspectives of the producer and the consumer are not diametrically opposed.
- What happens is that they do demand protection against different undesirable events.

In conclusion, the complete set of observations upon which a statistical analysis is based is a *sample of n*, where n refers to the number of observations, measurements, events or other factors. The sample of observations is usually assumed to be representative of a much larger number of possible events or measurements based on observations or experimental conditions. This larger group of potential observations is the *population*.

Measurements both of the population and of the sample are distributed in some way from a minimum value to a maximum value. We refer to the plotting of this distributed data as a *distribution* which has a central tendency and a variance. To characterize this distribution we need to measure its expected value and dispersion which is done through $\bar{x}$ and s for the sample; and μ and σ for the population.

9.2 Principles Underpinning Sampling Plans

"Sampling" as a process is probably as old as mankind, but the concept of using inference based on a sampling plan to eliminate rule of thumb, bias, or some obviously unacceptable assumptions, is relatively new. The goal is an objective estimate in contrast to the biased and subjective approaches that are used so often. Given that it is not feasible, or it is not economically acceptable, to test a whole population we work on a sample.

Say that the officer in charge of procurement has to decide whether a population of springs is acceptable or non-acceptable. One way to screen the lot would be that of testing every single spring in it. *If* the testing was destructive, *then* after it was complete there would not remain any useful springs. If the testing was not destructive, the cost of this 100% quality control procedure might have been the prohibitive factor. So we may be better off by:

- Selecting a sample,
- Examining carefully every unit included in this sample, and
- Basing our decision on the outcome of this examination.

Statistical theory has established criteria for sound sampling procedures. One of them is that the properties of the sample should correspond as closely as possible to those of the population. If the sampling is repeated a number of times, the mean of the samples means should approximate nearer and nearer the population's expected value. In the general case, there are two methods of drawing a sample:

- Random sampling and
- Representative sampling.

The assumption with random sampling is that all possible choices are equally probable. A table of random numbers helps in the selection of items from the population. The keywords in random sampling are "without bias", which conditions the sampling plan. *If* some objects in the population are more likely to be chosen than others, *then* the sample is said to be *biased*. Typically, subjective methods of selection from a population lead to biased samples.

There exist different plans representative of sampling. Random is often the best option. Another more interesting one involves stratification of the population to provide more focused information pertinent to each stratum. For stratification purposes the population is subdivided into several parts, or strata, and the number of observations in the sample is apportioned among these strata.

Stratified sampling can also be the result of a *proportional testing strategy* employing a partition and *allocation scheme* in which test cases fall into subdomains. Some studies have demonstrated that, under certain conditions, stratification is an option at least as effective as random sampling and testing.

The problem with stratification is that the steps leading to it may include bias. In addition, stratification must have a valid reason. It would be wrong if it is dictated by past laboratory practises that have outlived their life cycle. Another shortcoming of stratification is that while the data represent the desired stratum, but are not representative of some critical factor(s) which are not evenly distributed among the strata.

The different sampling plans we will study in this and the following section are only to a small degree alternatives. Each case has a sampling plan which fits it best, and misleading results can occur if samples are not taken correctly.

An approach leading toward a better sampling procedure starts with the realization that "sampling", essentially means *data sampling*, and these data must have a reason for being.

Consider as an example a case in the processing industry. The chief engineer makes the prediction of uniformity regarding certain characteristics of the firm's main chemical product. The key variables described in the specifications are constant mean and constant spread.

The prediction is plotted as a function of time. If the production engineer uses $\bar{x}$ and R SQC charts, where $\bar{x}$ is the mean and R stands for the range, he can readily see if and when his process gets out of control. Say that in the beginning the process is in control, subsequently however:

- An $\bar{x}$ point above the upper control limit (UCL) indicates a shift in the distribution,
- An R point above the UCL indicates a larger spread, which is also an out-of-control situation.

As this brief example documents, SQC charts are interactive tools, easy to establish and their pattern is well understood (see also Chaps. 13 and 14). The challenge is that of choosing and implementing the proper sampling method, which is also a vital part pf the design of:

- Experiments and
- Control conditions.

Take as an example of control conditions that of a factory installing a plan for statistical inspection of manufactured items. Certain specifications for the factory's product are stipulated in the client's contract which also states that prior to delivery the produced goods shall be inspected to assure that only a small portion of the items could fail to meet specifications.

The quality control engineer can use this "small portion" as consumer's risk β and produce an operating characteristics curve which defines sample size after α has been chosen, usually by senior management. He could also use some practical findings; mathematically sound data samples permit us to do a much better job than is possible with 100% inspection.

We can make a smaller number of measurements and estimate the true value from samples by applying *the principle of persistence* of small numbers: *If*, in a group of quantitative phenomena selected without bias, a small proportion of the group deviates sharply from the characteristics of the remainder of the group, *then* this tendency will persist:

- No matter how large the group may be made, and
- Irrespective of the number of samples selected.

It is often said that it is better to take a large number of small samples. This is true in two cases: sequential sampling (Sect. 9.3) and SQC by variables (Chap. 13). Even then there is a limit on how small a sample could be, and there is as well a counterargument. Small samples can have severe aftermath on the accuracy of statistical inference. Estimated approaches based on statistical tools at or near their limits involve:

- A high variance, and
- Significant margins for error.

Examples are statistical estimates with a sample of four or five, and projections simply based on averages without accounting for differences among the members of a population—whether people, manufacturing applications, events, or other observations. A sample size of five is acceptable in SQC by variables in manufacturing because of the orderly way in which successive samples are taken. On the contrary, in experimental studies small samples can lead to unreliable results.

Statistically insignificant samples may be one of the reasons why in social science studies—as contrasted to manufacturing—estimating approaches tend to involve high variance and high margins of error. I recently heard an argument that social scientists in a given study considered a sample of one as being statistically valid, and based on it their conclusions.

Another principle which impacts on sample size, is that of *decreasing variation*. As a large and larger proportion of a group of observations, measurements, events or phenomena is selected from the population by means of successive unbiased samples, the characteristics of each enlarged sample—such as the central value and variance—will differ less and less from the characteristics of the population. This makes it possible to determine the size of the sample in proportion to the whole of data.

A third principle underwriting sampling procedures is *statistical regularity*. If a reasonably large sample is selected without bias from a population, the characteristics of this sample will differ only a little from those of the universe. Some statisticians consider the *principle of large numbers* as the alter ego of that of statistical regularity. *If* an event may happen in only one of two ways and is observed to happen under the same essential conditions a large number of times, *then* the ratio of the number of times that it happens in one way to the total number of trials appears to approach a definite limit.

What is known as a *stability test* of sampling may be used as a rough check upon the adequacy of a sample. This test consists of division of the original sample into two equal samples by means of random sampling; or selection from the original data of a new unbiased sample equal in size to the sample already taken. The next step is noting whether or not the characteristics of the newly selected sample, differ materially from the characteristics of the original sample.

Another principle underpinning sampling plans concerns the fact that the degree of variation permissible between the sample and the whole cannot change arbitrarily since this depends upon the use that is to be made of the sample. The necessary sample size, i.e., the proportion represented by the sample, varies with different problems. The required number in a sample increases as:

- The variation in the individual items increases and
- There is a need for greater accuracy of the results.

Notice, however, that the accuracy of a sample does not increase directly in proportion to the size; it increases with the square root of the size of the sample increases. Hence, to double the accuracy, we must quadruple the size of the sample, and to treble the accuracy we should increase the number of items to nine times the former number.

9.3 Practical Examples with Sampling Plans

Theoretically, in the case of a production process every individual item chosen should be measured and returned to the population before another selection is made. This however is not practical at the production floor, apart from the fact that some samples undergo destructive testing. If the population is large compared with the sample size, very little error will result from the procedure of not returning each individual to the population.

In SQC and other implementation domains, plans permitting more than one number of samples are described as *multiple*. A double sampling plan can be considered as a special case of a multiple sampling plan. A multiple plan involves a finite number of samples, and this contrasts to what is known as a *sequential* sampling plan, which may permit a virtually unlimited number of samples until a quality decision is reached.

Sequential sampling is the rule in SQC because we deal with plural samples in the population domain. Its objective is to reduce the overall number of required observations by making subgroup observations in sequence to each other. However, the best quality control practice is to carefully evaluate the plan to be chosen, otherwise quality control objectives may not be attained.

Usually, sequential sampling plans can be designated having operating characteristics curves (Chap. 10) closely similar to the OC curves of a single sampling plan. The calculation of the operating characteristics curves for sequential sampling plans usually follows the pattern, which is discussed later in this chapter in the context of multiple sampling.

A test known as the *sequential ratio test* is designed to distinguish between alternative hypotheses, based on the likelihood ratio. The latter consists of the independent sample values being measured assuming that the hypothesis of no difference, H_0, is true.

Multiple sampling plans can be double, triple, etc., and they are adopted for two reasons: because they can be more flexible and because (as stated) in the long run they require less inspection. To properly design the multiple sampling plan one should carefully project the series of samples sizes, with associated acceptance and rejection numbers to be used in determining the acceptability of the lot. Tables, for instance, the double sampling plans are usually based on the Dodge–Roming tables, providing information for:

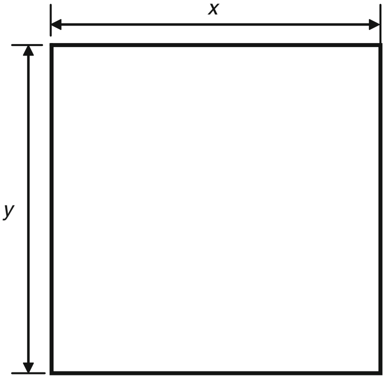

Fig. 9.1 Use of x, y coordinates of a box for sample section

- Sample sizes and
- Accept–reject levels.

For a numerical example on alternative sampling plans, say that the Eastern Electronics Laboratories has been requested to develop alternative SQC plans for inspection: by single sampling, by double sampling, and by triple sampling. According to MIL-STD-105A, developed by the US Military for procurement during WWII, for lot size $N = 300$ the following sample sizes are required by inspection level:

Severity of inspection level	I	II	III
Sample size	15	35	75

Let us suppose that it is decided to tail inspection level II, which requires a sample size $n = 35$ for single sampling procedure. The items in the lots are being delivered in a box and they are arranged so that the inspector can select any unit. Since the inspection plan requires random sampling, a table of random numbers has been used, with two columns of two digits each for each dimension. Hence, selection was made on the basis of x, y coordinates as in Fig. 9.1.

Say that finally the *acceptable quality level* (AQL) equal to 4 has been selected (see Table 9.2). *If* a single sampling plan is used, *then* with inspection severity II the inspection department should take a sample of 35 devices and accept the lot if three or less items are defect, or reject the lot if four or more are defects (Table 9.3). With a double sampling plan the first sample would be equal to 25. For two or less defects the lot would be accepted; for five or more it would be rejected. For three or four defects the inspector proceeds with a second sample of 50. In different terms, if the sample has three or four defects the inspector cannot reach an immediate decision; he should take and test another sample before accepting or rejecting the lot.

The same procedure is used with the multiple sampling plan in Fig. 9.2. This example comes from the financial industry; a loan's application known as

Table 9.2 Outgoing quality level single sampling

	Tightened inspection	Normal inspection	Reduced inspection
Outgoing quality level (AQL)	0	4	6
Accept	0	3	5
Reject	1	4	6

Table 9.3 Outgoing quality level double sampling

	Sample size:	A[a] 4	R[b]	A 6.5	R	A 10	R
First	25	≤2	5	≤3	7	5	11
Second	50	≤4	5	≤6	7	10	11

[a] Accept
[b] Reject

Risk-Adjusted Return on Capital (RAROC) which uses operating characteristics curves for judging creditworthiness. RAROC was developed in the late 1980s by Dr. Carmine Vona who was a nuclear engineer by training, and the boss of information technology at Bankers Trust of New York.

Usually, loans are given on an accept–reject basis, at one shot reflecting the client's creditworthiness. Vona thought that this is the wrong approach because it is possible to take an insurance on the risk assumed by the client. That is exactly what RAROC does with multiple sampling. Every time the test moves to the right the interest rate offered to the client is increased by the corresponding amount of insurance for the greater counterparty risk being assumed by the bank.

Vona's concept follows the principle of classification of defects in sampling inspection in a way emulating the testing procedure by the Eastern Electronics Laboratories. The sophistication of the method can be increased if the more critical defects are considered separately from, and are usually given lower AQL values than, less critical defects. Quite often in the industry the quality control department proposes a double classification system which weights the criticalness of defects.

As an example, say that the SQC-classification system in question defines the criticality of defects in a manner distinguishing "major", "midway", and "minor" criteria—hence three thresholds. A code letter is assigned to each inspection point by engineering and cannot be changed by the inspector. The classes are defined in terms of assemblies whose failure would result in:

- *Major*. Serious injury to personnel or loss of the manufactured unit—for instance, in an aircraft the weakened root cord structure
- *Midway*. Failure of the final product to function as intended, but without the loss of personnel or of the unit, as for example defective hydraulic line connection
- *Minor*. Interference with subsequent assembly or repair operations, or reduced quality of the end product, as for instance massive rivets

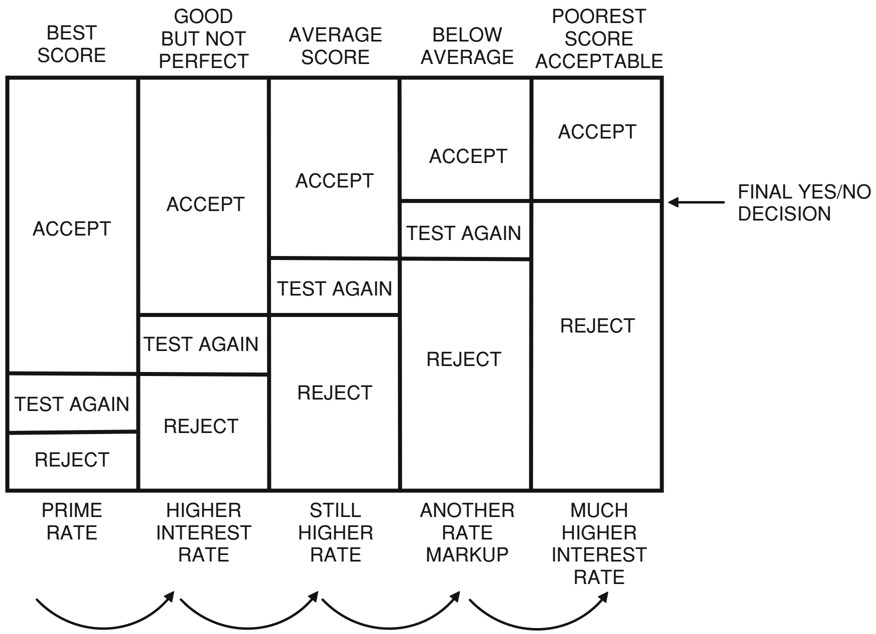

Fig. 9.2 A sequential sampling plan permits to avoid inflexible yes/no decisions on loans, by taking a reinsurance for higher risk

When a defect is found in any class, it is assigned a code number indicating its seriousness; i.e., the probability that this defect will eventually cause failure of the part. This number is usually determined by the inspector by means of criteria such as:

- A discrepancy which will prevent an assembly or installation from operating properly or will cause intermediate failure.
- A discrepancy that may not cause a malfunction or possible failure, but is a deviation from specifications and acceptable workmanship standards.

A sound system to use in connection with quality control is *chargebacks*, also known as *demerits*. In my practice, I have found it top class in keeping inspection personnel on its toes because laxity is penalized postmortem. Here is, in a few words, how it works.

In the course of production a defect may be found by device level, subassembly, system inspection, or post-inspection personnel. If a defect is found which should have been cleared previously by another department, it becomes a chargeback to that department—or, if a fine grid is used, to the person responsible for laxity.

If a production worker observes a defect prior to inspection, he points it out so that inspection sends it to be reworked, if possible, to specifications. If this is not possible, the defect is recorded in the production pickup section of the inspection book, with the classification code letter and number. If, however, a defective item passes through to, say, the assembly operations and is subsequently flashed out, *then* there is a chargeback.

Inspection-found defects are coded and recorded on regular pickup sheets. Defects missed by inspection and detected later will be treated as demerits. Each SQC checkmark may contain several defects, such as missing rivets, oversize holes, and so on. The various classes of defects should be considered in separate categories and given different AQL values. Alternatively, rather than creating a large number of categories, these defects can be weighted according to code letter and number as demerits.

Quality control tables give the demerits assigned to each combination of code letter and AQL requires. *Chargebacks* are weighted *double* in consideration of the increased risk of not finding the defect prior to delivery, while those checks of defects made by the production line are given in the tabulated value. When all the above factors have been considered, a numerical total of demerits is available as a measure of quality of the lot (or of a major unit).

This reference to a quality control plan comes from a real-life application. Among the advantages it provided have been that conditions causing the process to go out of control were investigated sooner; adverse conditions affecting a small number of units were readily apparent; and differences between shifts showed up as out of control points in a quality control chart which tracked the sample's spread (*R* chart for range of measurements, Chap. 13).

Some inference rules were established to facilitate accept–reject decisions because of out of control points on SQC charts. Attention was paid as to whether these had to be attributed to a lower quality being produced, poorer inspection techniques, or an assignment of code numbers in borderline cases influenced by the inspector.

9.4 Discovery Sampling[1]: A Case Study

Discovery sampling is a step forward in acceptance sampling. Its value lies in the simplicity with which it yields a prescribed measurable result, by introducing the concept that products from a group of machines and workers have a process average that may be used as a measure of quality by taking into account three factors:

- Some lots do not actually contain defectives.
- Lots that do contain defectives are likely to contain a small percent of them.
- The average percent of defectives is defined at the level of the bank of products between production floors (intermediate stocks).

[1] Originally developed by Lockheed Aircraft, discovery sampling has been a little known but powerful SQC tool useful to all sorts of enterprises—both big and small.

In procedural terms discovery sampling starts with inspection of 10 items from a lot. The lot is accepted if no defectives are found. Above zero percent there is a graduation in decisions. In a practical implementation, for a given lot size the average percent defective in stock was less than 0.25%, and the average product quality was better than 99.75%. Given the AQL specifics of this application, the lot was accepted.

The notion underpinning this procedure is that a lot that is 100% conforming to specifications will be accepted in any case, while a lot that is totally unacceptable will be discovered if only one part is inspected. The entire sampling risk is confined to that category of lots in which both good and defective items occur; such a lot is defined as *partially defective*. The fraction of partially defective lots presented for acceptance is:

$$A = \frac{\text{Partially Defective Lots}}{\text{Good Lots} + \text{Partially Defective Lots}}$$

In the implementation in reference for individual departments "*A*" was found rarely to exceed 20%; it was less than 10% in most cases. The value of "*A*" was reasonably stable at less than a value prescribed by quality control. This indicated a constant cause system.

Because partially defective lots are at the core of discovery sampling, their fraction should be verified before installation of that method, as well as at intervals thereafter because quality wise the production pattern may change. A verification is essentially based on measurement of the process average and dispersion trends.

Attention should be paid to the distribution of partially defective lots within the group. A study of 20,000 lots made in different inspection areas by a major American manufacturing firm documented that they had the same or very similar distribution curve. The *average percent defective* in stock was found to be less than:

$$100 \cdot \frac{A}{4n + 2}$$

where

A = the fraction of partially defective lots presented for acceptance

n = sample size or the number of items inspected per lot.

Content with the results of discovery sampling, the management formalized its procedure through the organization into the following steps: Start with assigning intervals for p, the fraction defective. For instance:

0–5
6–10
11–15

The fraction defective p is computed for each partially defective lot by dividing the defectives found by the total lot quantity. When all of the partially defective lots are distributed within the aforementioned intervals, the quality control procedure computes the decimal fraction for the lots found within each interval. The probability of acceptance for a lot that contains defectives when the sample size is 10 and no defectives are allowed in the sample, is:

$$P_a = (1 - p)^n$$

where n is the sample size. Assigned intervals for p, quantity of lots found within the internal and computed statistics should be entered in a 6-column table as follows:

1 Assigned intervals for p	2 Quantity of lots found within the interval	3 Fraction of lots occurring within the interval	4 Probability of acceptance with a sample of 10, no defectives allowed	5 Mid-point of the interval	6 Fraction of defectives in stock for partially defective lots

The fraction of defectives in stock for partially defective lots for each interval is the product of Columns 3–5. This product is entered in Column 6. The meaning of fraction defectives in stock partially for defective lots is that "so many lots" (typically a small to very small fraction) occurred for each interval, and accepted by the plan.

The sum of Column 6 from assigned interval for p from interval 0–5 to interval 96–100, is the total contribution for defectives from all of the lots in all of the intervals of the partially defective group. The product of the sum of Column 6 and the percentage of partially defective lots 100 A, gives the average percent defective (APD) in stock.

If a frequency density function or operating characteristics curve are desired, each value in Column 3 must be multiplied by the number of intervals—in the present example 20—and plotted over the range for each interval. This gives a frequency density histogram. A smooth curve representing the histogram is a frequency density function for material presented for acceptance.

The example which we have seen has been discovery sampling by attributes; an inspection part is either good or bad with no measure of the degree to which this is true (Chap. 14). If this degree can be measured, then it is preferable to use an SQC plan by variables (Chap. 13).

Say that five parts were selected at random from a lot are arranged in order of size. The dimension of the part is specified as 1 cm with tolerance of ±0.010. These five parts are within tolerances characterized by the following distribution.

	Dimension
Upper limit of tolerance	1.010
	1.006
	1.002
Ordered sample	1.001
	0.998
	0.993
Lower limit of tolerance	0.990

If either the largest or smallest value exceeded the tolerance the lot should be held for disposition. Otherwise, discovery sampling discards the highest and lowest measurements—in this case 1.0006 and 0.993—and calculates the difference of the remaining extreme measurements.

	1.002
	−0.998
Difference	0.004

Then, it subtracts this difference from the smallest remaining measurement:

	0.998
	−0.004
Difference	0.994

If this number is greater than the lower tolerance, the lot is accepted as being above the minimum. The next step is to add the same difference to the largest remaining measurement.

	1.002
	+0.004
Difference	1.006

If this number is less than the upper tolerance, the lot is accepted as being below the upper tolerance. If either of the above numbers exceed the corresponding engineering tolerance, the lot is held for disposition.

In essence, discovery sampling by variables excludes the extremes of the sample and works with the extremes of the middle measurements (in this example three out of five). If the sample contained ten parts, then the upper and lower two are discarded—essentially the two largest and the two smallest measurements.

The rest of the discovery sampling procedure works in a similar manner to that of the example presented to the reader.

9.5 Sampling Errors

In 2000, as the dot.com bubble had burst, many Internet companies went against the wall taking along with them some of the established names in information technology and in business systems. One of them was Xerox. In September 2000, its cash on hand was a razor-thin $154 million for a total debt of $17 billion, including a $7 billion credit line which was projected to be exhausted by the end of that year.

A new president took over and one of the first things he did was order up a review of the economics of the existing Xerox product line. He was presented with charts showing that Xerox was “world class” in terms of manufacturing and development costs. But the company’s profit and loss (P&L) statements told a different story and his response was: How do you know?

It turned out that Xerox staffers had relied on a sample of 1994 market data, so limited as to exclude most of Xerox’s Japanese competitors. The new CEO ordered them back to the drawing board. Weeks later, he finally was presented with evidence that Xerox had failed to maintain its hard-won parity with the Japanese.

The proper sample of competitors and challenges they presented made it clear that Xerox was at a large and material cost disadvantage against the Japanese across the copier market. The staffers had been very imprudent by using a sample of obsolete statistics. Besides that, the sample of competitors they had taken was too narrow leaving out their most formidable challengers. The result was:

- Loss of market share and
- A critical condition in the company’s finances.

At about the same time, a study by the Information Technology Association of America (ITAA) suggested that companies seeking to hire information specialists in the next few years will be faced with a severe shortage. ITAA put the deficit at 191,000 professionals. Such a large number of a deficit was derived from a study of responses from:

- 149 out of 1,000 IT companies in a sample, and
- 122 out of 1,000 in another sample of non-IT companies.

In both cases the 1,000 companies were the sample targets, but that number was never reached. Those responses that were received indicated that there was an average projected vacancy level of 33 positions per IT company employer and between four and five per non-IT company employer. The guestimate of 191,000 IT vacancies seems to have been made by multiplying these figures [1].

It does not require great ingenuity to appreciate that the samples were substandard. The survey data was inadequate and the sampling methodology itself was wanting. The flaw in the sampling methodology has been that such small number of responses make it quite likely that the survey had a significant bias.

- The sample was small and
- There has been a very low level of response.

Both factors led to statistical bias. In addition, there was a failure to statistically validate the estimates being made by a follow-up survey. The estimate of vacancies is a tricky subject. A specific problem is that data from employers is only a small part of the information necessary to determine supply–demand imbalances and therefore level of projected vacancies.

This is by no means an exceptional case. Many sampling inspections and surveys fail to account for the number of applications, as well as the number of individuals interviewed. Yet, both the number of offers made and of salary levels offered (accepted or rejected) are important data for an accurate estimate of vacancies. A good statistical supplement would have been:

- The number of hires,
- The number of reductions in workforce, and
- The people who leave voluntarily.

None of this information was provided in the survey and therefore in the report. For instance, vacancies can balloon if employers offer lower salaries than those prevailing in the industry. In this case they should not be taken as indicators of real demand for a given profession. The lesson to be learned from this and similar cases is that not only the size of samples but also their composition, and the questions being asked, are very important in a statistically valid analysis leading to an estimate or projection.

Another example of defective sampling methodology which came to my attention in the technical auditing of a financial institution was that of poorly documented credit ratings. Without a well-documented history of defaults, including many instances of successful and unsuccessful debt service, it is not possible to be sure that indicators typically used by credit officers will pick up future problems.

Similar cases exist in quality control. This is a problem connected to focusing on one's aims prior to using sampling approaches, and it is indeed inherent in the analysis of all types of phenomena for which only small samples exist. What I just stated about an erroneous credit sampling and screening methodology has been confirmed by the careful study of past instances of default. Defaults have engulfed entities with relatively:

- Low as well as high debts,
- Good as well as poor management,
- Long-established product lines as well as more recently launched innovative instruments.

A stratified sampling of low debt levels, poor management practises, and classical type loans will provide incomplete and most likely misleading evidence for management decision. A counterparty's credit behavior should not only be unearthed by focusing on high debts and poor governance but as well—and most importantly—by concentrating on permutations of risk factors.

Another example where sampling methods have been wanting is that of unreliable statistical samples in opinion polls. Take politics as a first example. In the US from 1992 to 2010 the number of presidential polls has more than quadrupled. Theoretically, that means a lot of more information available to the public. But the proliferation of rapid polls — known as down and dirty—makes much of that information unreliable.

According to expert pollsters, as a minimum opinion surveys should be taken over a period of at least 3 days and include at least 1,000 voters. By contrast, there has been plenty of single-day polls based on samples as small as 250 people. There are also one-night polls which are totally irresponsible and verging on product liability—as a veteran pollster has it.

In conclusion, like any other activity which is worth doing sampling has its rules. These rules are not always observed and what happens is a misuse of the term *sampling* which can end up in liability. The reason behind such false sampling plans is human error. Therefore, it is not enough to hear that "this result (or conclusion) has been based on sampling." It is also necessary to know how the sampling was done.

9.6 Product Innovation Requires More Sophisticated Sampling Plans

While the science of statistics moves relatively slowly, innovation in man-made products is characterized by rapid advances which in turn require more sophisticated sampling plans. Old concepts are still precious as evidence on the transition in the implementation of statistical methods and tools, but they are not a guide to new applications which require plenty of imagination as well as detail in the way a sampling plan works.

There are plenty of examples on product liability due to sampling errors, and not only in connection to political opinion polls. Detroit, for instance, has made major blunders in projecting the type of motor vehicle American and international clients will require. This wrong-way strategy has been based on the narrow perspective of opinions by clients biased toward SUVs.

One might have thought that Detroit's Big Three employ the most effective methods a technology of prognostication can provide. Practically, this is far from being the case as documented by the fact that the formerly mightiest vehicle manufacturer in the world, General Motors, brought itself to its knees and Chrysler (the smaller of the Big Three) went along.

While the insistence to continue designing and manufacturing big gas-consuming automobiles when the price per gallon hit $4, was a top management blunder, other failures were directly debited to the engineering of Detroit firms. An article in Automotive Design put it in these terms: “To Europeans, US domestic products were deemed to be relatively crudely engineered, of poor build quality, using low-grade materials and of questionable design—never mind indifferent dynamics and thirsty engines” [2].

It took the ordeal of going in-and-out of bankruptcy for GM and Chrysler to change their product management’s perspective and their method for sampling potential customers to unearth needed changes in design in a way that engineering can be ahead of the curve. After GM’s bankruptcy, its engineering division has been restructured to take advantage of regional centers of excellence. A new top management decided the company had to design:

- Small and mini cars in South Korea,
- Compact models in Europe,
- Global mid-size truck platforms in Brazil, and
- Mid-sized platforms and trucks mainly in North America.

Among themselves these centers aim to cater for all markets around the world, by providing a global engineering competence. So far so good, but for such laconic classification of regional competence, is it enough to define the mission given to design engineers? I doubt it, because it is too general and therefore ineffectual.

The big car–small car debate is more than 30 years old; it is not a newcomer in automotive competition. Packing the car with a great deal of *in-car* electronics[2] is a dozen years old design feature. Therefore, it is no more an indicator of future competitive advantages. The new generation of competitive advantages work under the impact of novel customer inputs motivated, among other reasons, by:

- Death statistics due to car accidents and
- How automotive manufacturers could contribute to the reduction in fatal accidents.

While the weakest link in car driving is the man behind the wheel, the equipment’s features too play a role. This is an engineering challenge closely connected to *product assurance* (Chap. 1) and to *service assurance* (see Chap. 4). However, as we will see in the following paragraphs the widely held notion that the reasons behind car accidents have a universal bearing is wrong.

- Major countries tend to have an individual pattern of background reasons for auto accidents, and
- Samples taken indiscriminately from the global population will be biased, because the global population of drivers is by no means homogeneous.

---

[2] *In-car* electronics has more or less reached a saturation point, with new offers distracting rather than helping the driver.

Let us start with statistics. In mid-2011 it was estimated that worldwide in that year there were 5.5 million car crashes. Precisely because the global population of drivers is not homogeneous, the designer's challenge is that, more or less, every one of these care accidents will be different. "The only time you get two crashes the same is in a crash lab", wrote Thomas M. Kowalick in Automotive Engineering "and that's not real-world data. If you could take just one day's worth of crashes in America —which is 20,000 tow-aways—that would be more data than you collect in one year in the lab" [3].

An upcoming device in the automotive industry, and a "plus" in competition, is the *event data recorder* (EDR)[3]—a so-called 'black box'—which by all evidence will play an important role in establishing causes and culpability in crash events. The intent of such information is to help in revealing:

- Driver and component failures,
- Accident patterns, and
- Injury risks by type of accident.

It could also be instrumental in detecting vehicle defects, directly contributing to future vehicle development and design. In addition, because the EDR is a car component its information will help in making more accurate statistics regarding accidents with motor vehicles which vary widely from country-to-country and over time. When security in auto transport escapes the control of authorities and accidents mount, these statistics convey the message of a crisis.

The wider spread of EDR is a long delayed innovation in private car features and it comes on the heels of another innovation—that of automatic breaking by car electronics to avoid a crash. Another similar advance is that of an automatic pilot which can park the car in a narrow space at a signal from a smart phone. In both cases:

- A system of microprocessors activates the car's engine, gearbox, steering and brakes and
- Sensors alert the car's command about the risk of bumping into other cars or people.

At least theoretically, this increases the security of the car's passengers and of the surroundings.

Pilotless cars, such as the Volkswagen Sharan are still laboratory models. Many people at the mid-September 2011 Frankfurt Motor Show were asking not only how the cars of the future will be powered, but what kind of security control will feature and how they could ease the driver's job. That is an issue where behavioral patterns are important, and such patterns vary from one country to another (more on this later).

[3] Back in 1996, in the US, Ford and GM began installing the early predecessors of EDRs in their cars.

“Where does the car end and the phone begin?” asked Chris Anderson, the editor of *Wired* magazine, at a brainstorming session organized by Audi, the carmaker [4]. By all likelihood, a future car will be more like a computer on wheels, networked with the surrounding infrastructure and other vehicles. This looks like science fiction, but it is at least an innovation—which cannot be said of the 110-year-old electric car.

It is appropriate to recognize that like many other man-made devices—for instance airplanes—motor vehicles are both an opportunity and a risk. In Chinese, the first syllable of the word *weiji*, for crisis, is *wei* and means danger; by contrast the meaning of the second syllable: *ji* is opportunity. The challenge is that of designing a better mousetrap, which has always been a guiding light in business.

In the case of auto accidents the “better mousetrap” is electronic devices specifically developed to prewarn the driver of an impending accident, which is within the each of current technology while Sharan is still in the lab. Within this perspective the mission of the design engineer is particularly influenced by two factors: product assurance and functionality. An optional balance is doable provided that there is:

- A wealth of statistics *personalized* at the drivers level, and
- Sensors sensitive enough to differentiate one type of accident from another.

As already mentioned in preceding paragraphs accidents happen in a variety of ways, and they have no unique pattern around the globe. After classifying them into seven main categories (plus a minor class of “other kinds”) Automotive Design has provided some most interesting auto accident statistics from America, Germany, and Japan [2].

- Collision with fixed object or vehicle leaving the carriageway, has a frequency of 46% in the US, 32% in Germany, and only 17% in Japan
- At 34% accident with a pedestrian is the highest in Japan, followed by Germany at just 14% and the US at 12%
- At a 21% likelihood, collision with a vehicle which turns into or crosses a road is also the highest in Japan, but the US at 18% and Germany at 15% are not far behind
- Germany leads in parallel vehicle accidents with 21% of all crashes; Japan and the US, respectively stand at 11 and 10%, have half the German score
- There is a nearly equal probability of accident with a vehicle moving ahead. It stands at 6% of all cases in Germany, 5% in the US, and 5% in Japan
- The likelihood of an accident with a vehicle moving laterally in the same direction is 3% in Germany, 2% in the US, and 2% in Japan
- Worst of all, the head-on collision with an oncoming vehicle is 2% in Japan, 1% in Germany, and 1% in the US—this still being a big percentage for this type of accident

Developed for all major markets of motor vehicles such statistics can be instrumental in promoting product assurance in tomorrow’s competitive markets. They can be used to advantage from car design to the development of sophisticated devices which respond to specific safety requirements country-by-country and eventually driver-by-driver. Since accident patterns are not universal a "good for everybody" design makes no sense.

The mission of the designer is that of analyzing accident patterns in various regions and in different countries, understanding and formalizing the foremost safety needs and incorporating service quality features to be activated/deactivated on driver's sign-on and sign-off. This has been done for nearly two decades with fuzzy chips which:

- Learn the car owner's driving profile and
- Optimize the use of fuel in accordance with this driving pattern.

In conclusion, significantly improving product assurance through higher security is a quantum leap in competition. Big achievements are made when engineering designers are given the goal to put muscle to well-defined breakthroughs, seeing them through in an evolutionary way from an ideal to a model and from there to a product appealing to the customer.

Such a goal, however, must be focused and the chosen course documented through studies based on correct sampling of the user population. Years ago at UCLA, my professor of quality control mentioned an old saying about errors with statistics. It went like this:

> There's a great text on errors in estimation.
> Once you trip on it, entails
> Twenty-nine distinct damnations.
> One sure, if another fails.

References

1. Electronic Design, 2 Sept 1997
2. www.automotivedesign.eu.com, Sept 2010
3. www.automotivedesign.eu.com, July/Aug 2011
4. *The Economist*, 17 Sept 2011